

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 by teacher preparation programs seeking accreditation. In response to these new CAEP specifications, the researcher at the College of Education (COE), University of Hawai'i at Mānoa, revised the COE program completer 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.

Considerations for Writing Survey Items

Is the question biased? (*Bad example:* "Our program is consistently ranked in the top 50 in the country. On a scale of 1-10, how would you rate our program?" *Better:* "On a scale of 1-10, how would you rate our program?")

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

Is it a compound question? (*Bad example:* "How prepared are you for teaching reading and writing?" *Better:* "1. How prepared are you for teaching reading?" and "2. How prepared are you for teaching writing?")

Does the question include a double negative? (*Bad example:* Should students not be required to participate in field experiences?) Yes No *Better:* Eliminate the word "not" from the question

Are the answer choices mutually exclusive and exhaustive? (*Bad example:* How many years of teaching experience do you have?) 0-3 3-5 5-10 10+ *Better:* 0-3 4-6 7-9 10+

Try to use consistent response formats. Use the same survey scale (i.e. response options) on items in a section

Do you allow respondents the option to pass on answering? It's best to let respondents skip questions, especially if it requires potentially sensitive information

Recommended Reading: Johnson, R. L. & Morgan, G. B. (2016). *Survey scales: A guide to development, analysis, and reporting.* New York, NY: Guilford Press.

Research Questions

(1) Does a 5-point bipolar scale ('Strongly Agree' to 'Strongly Disagree') or a 4-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 *a* (two-parameter) graded response model (GRM) or the equal *a* (one-parameter) GRM model a better fitting measurement model for each of the instruments data sets?

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

Research Method

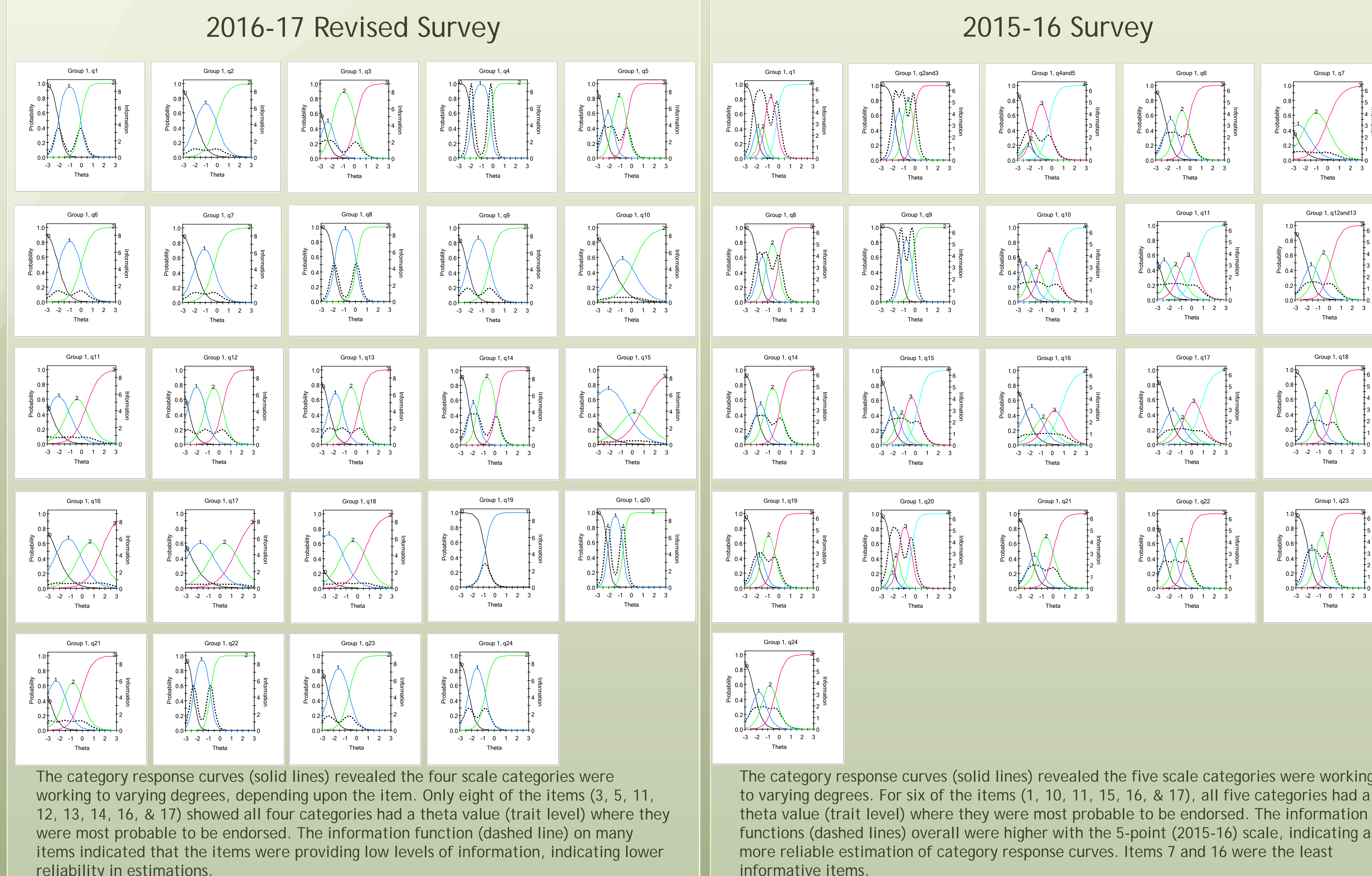
Both the old (2015-16) survey and the revised (2016-17) survey were administered as web-based surveys to all teacher candidates who completed their teacher education program in the semester of the survey's administration. For the 2015-16 academic year survey, 155 of the 229 program completers (68%) completed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

- Univariate, reliability, & unidimensionality analysis with SPSS v22
- Unidimensional item response theory (IRT) utilizing Samejima's Graded Response Model (GRM) with free *a* and equal *a* parameters with IRTPRO v4
- Item level analysis to inspect item functioning

Results

Both the 2015-16 and 2016-17 survey scales indicated strong reliability (Cronbach's alpha: 0.951 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 2015-16 and 2016-17, 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 \leq 0.8). Therefore, the free *a* model was selected as the best fitting model. For the 2016-17 survey, the RMSEA values for both models indicated goodness of fit (RMSEA \leq 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 -2LL, AIC, and BIC values in comparison with those of the equal *a* model.

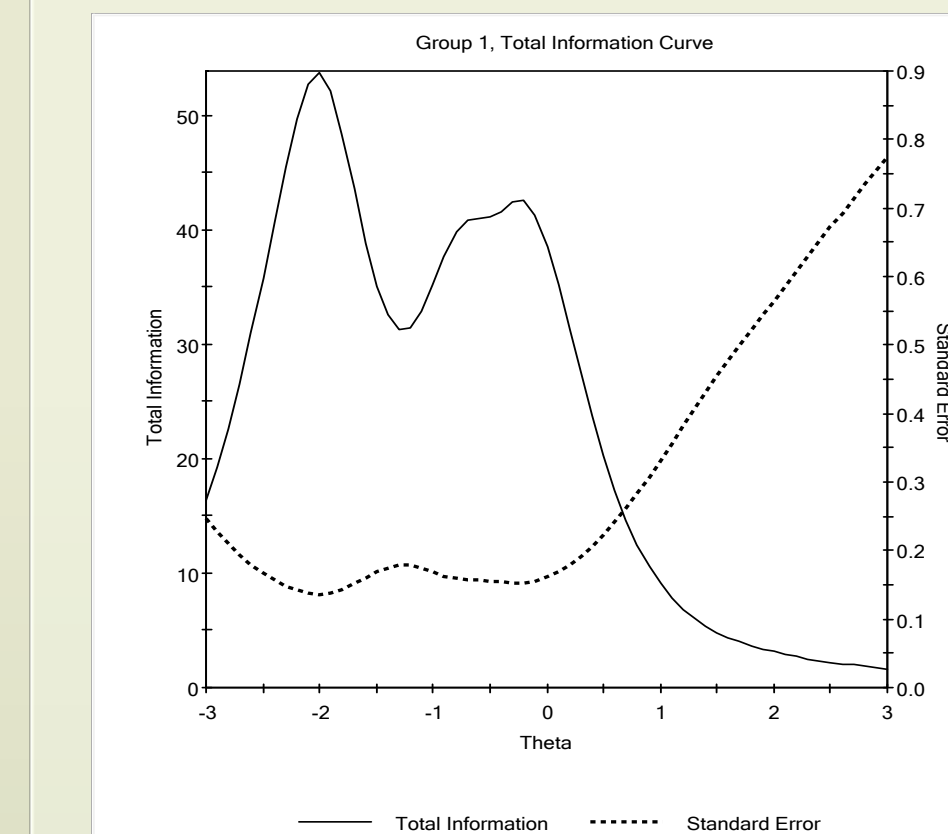


The category response curves (solid lines) revealed the four scale categories were working to varying degrees, depending upon the item. Only eight of the items (3, 5, 11, 12, 13, 14, 16, & 17) showed all four categories had a theta value (trait level) where they were most probable to be endorsed. The information function (dashed line) on many items indicated that the items were providing low levels of information, indicating lower reliability in estimations.

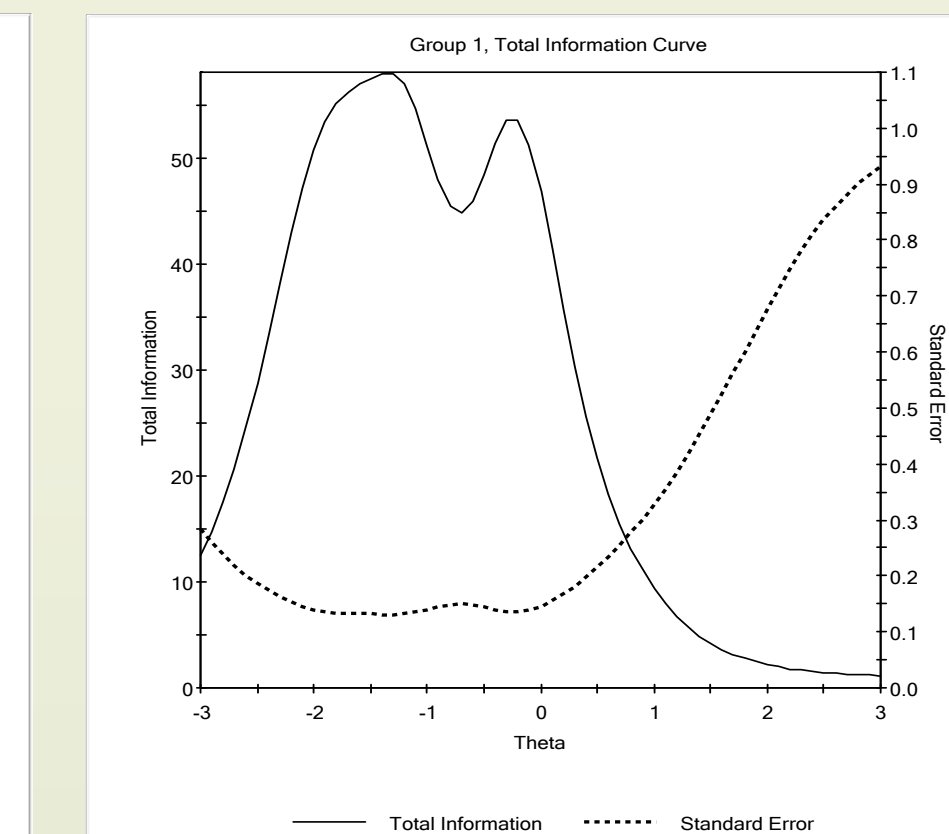
The category response curves (solid lines) revealed the five scale categories were working to varying degrees. For six of the items (1, 10, 11, 15, 16, & 17), all five categories had a theta value (trait level) where they were most probable to be endorsed. The information functions (dashed lines) overall were higher with the 5-point (2015-16) scale, indicating a more reliable estimation of category response curves. Items 7 and 16 were the least informative items.

2016-17 Revised Survey Items		2015-16 Survey Items	
Unipolar 4-Point Likert Scale: Completely prepared (4), Mostly prepared (3), Somewhat prepared (2), & Not at all prepared (1)		Bipolar 5-Point Likert Scale: Strongly agree (5), Agree (4), Neutral (3), Disagree (2), & Strongly disagree (1)	
Item #	Item	Item #	Item
	Stem: At the end of your COE teacher education program, to what extent are you prepared to...		Stem: My teacher education program helped me...
1	Create developmentally/age appropriate learning experiences?	1	Create developmentally/age appropriate learning experiences based on my understanding of how learners grow and develop.
2	Adapt to learner diversity and individual differences?	2/3	Adapt to learner diversity and individual differences to ensure inclusive learning environments that enable each learner to meet high standards.
3	Ensure inclusive learning environments that enable each learner to meet high standards?		
4	Create learning environments that support individual learning?	4/5	Create positive learning environments that support individual and collaborative learning.
5	Create learning environments that support collaborative learning?		
6	Demonstrate knowledge of the content in your field?	6	Demonstrate knowledge of content in my field.
7	Incorporate student standards, such as the Common Core Standards, into your teaching practice?	7	Incorporate student standards adopted by the Department of Education, such as the Common Core Standards, into my teaching practice.
8	Plan instruction that supports every student in meeting learning goals?	8	Plan instruction that supports every student in meeting learning goals.
9	Use a variety of instructional strategies to engage all learners?	9	Use a variety of instructional strategies to engage all learners.
10	Integrate technology effectively into curricula and instruction?	10	Integrate technology effectively into curricula and instruction.
11	Teach reading, including working with students who have reading difficulties?	11	Increase my ability to teach reading, including working with students who have reading difficulties.
12	Incorporate reading strategies across your curriculum?	12/13	Incorporate reading and writing strategies across my curriculum.
13	Incorporate writing strategies across your curriculum?		
14	Use assessment strategies appropriate to your students' needs?	14	Use assessment strategies appropriate to my students' needs.
15	Teach student who have disabilities?	15	Increase my ability to work effectively with students who have disabilities.
16	Teach student who do not speak English as their first language?	16	Increase my ability to work effectively with students who do not speak English as their first language.
17	Teach student who are gifted and talented?	17	Increase my ability to work effectively with students who are gifted and talented.
18	Teach student who are from different cultures?	18	Increase my ability to work effectively with students from different cultures.
19	Demonstrate professionalism as a new teacher?	19	Demonstrate professionalism as a new teacher.
20	Take responsibility for student learning and success?	20	Develop a sense of responsibility for student learning and success.
21	Work with parents and families to better support student learning?	21	Foster effective communication and collaboration in the learning environment with students, families, colleagues, and community members.
22	Engage in professional reflection to become a stronger teacher?	22	Become a stronger teacher through professional reflection and discussion with others.
23	Understand your specific strengths as a new teacher?	23	Understand my specific strengths as a new teacher.
24	Target areas of need for your own professional growth?	24	Target areas of need for my own professional growth.

2016-17 Survey



2015-16 Survey



The graded response modeling total information curve, as well as the item information functions, indicated that both the revised 4-point scale (2016-17) and the 5-point scale (2015-16) were best measuring the construct of preparedness to teach for program completers whose theta value (trait level) was between -3 to +0.75. The surveys appear to not be as informative of a measure for program completers who reported the highest levels of preparedness to teach, and thus have theta values greater than +0.75.

Additional Item Analysis

2015-16 Survey: The least discriminating items were 7 (*a* = 1.59) and 16 (*a* = 1.82). The most discriminating items were 9 (*a* = 5.06) and 2/3 (*a* = 4.80). The most difficult item to endorse was 9 whose *b* values started at *b*₁ = -1.24, followed by item 2/3 whose *b* values start at *b*₁ = -1.74. The least difficult item to endorse at *b*₁ was 7 (*b*₁ = -3.24). Item 16 had low probability, factor loading, and discrimination values.

2016-17 Survey: The least discriminating items were 15 (*a* = 1.22) and 17 (*a* = 1.33). The most discriminating items were 4 (*a* = 5.99) and 20 (*a* = 5.52). The most difficult item to endorse was 10 whose *b* values started at *b*₁ = -1.71, followed by item 8 whose *b* values start at *b*₁ = -1.85. The least difficult item to endorse at *b*₁ was item 15 (*b*₁ = -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, a next step will be to investigate if a 5-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 CAEP specification to use even numbered scales may not provide better measurement. *Note: In January 2017, CAEP 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 survey.

How to improve the instrument: Add items that measure preparedness for program completers with higher theta 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 item as results indicated that it was not functioning well on both versions of the survey.

Jessica Miranda
wellsjes@hawaii.edu

