Synopsis of Validity Evidence for the Cutscores Derived from the Grades 3 – 8 and 11 Standard Setting for NCSC ELA and Mathematics Assessments

Report Prepared by Barbara S. Plake, Ph.D.
August 18, 2015

Introduction

NCSC commissioned a standard setting activity prepared and conducted by Measured Progress. This standard setting activity was comprised four major components: 1) pre-administration planning, 2) administration of the standard setting panels, 3) vertical articulation of the results from the standard setting panels, and 4) state articulation of the results from the vertical articulation. The purpose of this report is to provide an overall impression of the evidence to support the validity of the final cutscores for the NCSC ELA and Mathematics assessments in grades 3 – 8 and 11. A more fulsome report of the validity evidence is forthcoming once more information is provided about the psychometric integrity and panelists' evaluation results are provided.

Pre-Administration Planning

Measured Progress shared several planning documents prior to the standard setting meeting. This included a document entitled Preliminary Recommendation for Standard Setting: Standard Setting Procedures and Schedule of Tasks.

Validity Evidence from Pre-Administrative Planning: Overall, I found the recommended procedures compliant with current standard setting practices. Further, via email I had the opportunity to review the orientation slides and also found them to be consistent with current standard setting practices.

Administration of the standard setting panels

The standard setting panelists met in Indianapolis August 10 – 12 to establish their recommended cutscores. Panelists met in grade and content groups, with panelists in grades 3 – 4 meeting as one panel, likewise panelists in grades 5-6 and 7 – 8 comprised separate panels. Grade 11 panels only focused on one assessments whereas all of the other panels consider 2 adjacent grade assessments for their content area. Each grade and content panel was lead by a trained facilitator who worked from a script to provide consistent directions to the panelists. The standard setting began with a general session for all panelists where an overview of the NCSC assessment system was provided followed by a general overview of the activities the panelists were to engage in over the Bookmark standard setting process.

The panelists provided three rounds of ratings once they had engaged in a indepth training process which included a) taking the test, b) reviewing the PLDs for the relevant assessment, c) filling out an item map detailing the KSAs for each item in the Order-item

Booklet (OIB), d) discussing these item maps, e) developing borderline performance level descriptors for levels 2, 3 and 4, and e)participating in a practice exercise for setting the 2/3 cut in a short practice OIB. Prior to embarking on setting their round 1 cuts, panelists filled out a training evaluation, which was reviewed by the panel facilitator to ensure all panelists indicated their preparedness to move into the operational rounds and to answer any outstanding questions or concerns by the panelists.

Between rounds, differing levels of feedback were provided. Following round 1, panelists were informed of the cutscores for the individual panelists followed by table level discussions, focusing on the location of the cutscores by the individual panelists. These discussions were at the table level; each panel had 2 tables and each table had a predesignated and trained table leader. Following their round 2 cutscores, in additional to the results of the individual panelists, panels were given impact data. Following round 2 feedback, the discussions of the results, again focusing on the locations of the cutscores within the OIB, centered on the full groups, not with the separate tables within the group. Following the panelists' round 3 ratings, final results for each panel for each test were provided. Finally, panelists filled out process evaluations for the standard setting process for the assessment. This process was followed for each of the two tests assigned to each panels (except for grade 11 which only focused on one assessment), but the panelists did not complete the practice bookmark placements for the second assessment in their set.

Validity Evidence for the Panelists Cutscores

These steps are consistent with current practice for the Bookmark Standard Setting Method. For the most part, these steps were successfully implemented and when minor issues emerged they were handled immediately and appropriately. There is no evidence to suggest that there is any reason to question the validity of the resultant cutscores from the cutscores produced by these panels.

<u>Cross-grade articulation</u>

Table leaders for all of the panels met to conduct the cross grade vertical articulation of the cutscores. Panels met separately for Mathematics and ELA. Panelists were shown the final impact data for all grades within content area. Panelists were asked to identify cutscores that resulted in impact data that did not look consistent or looked out of line. The panelists began their considerations at the 2/3 cutscore (that differentiated non-proficient from proficient levels of performance). If panelists felt a cutscore needed to be revisited, the panelists then looked at the borderline performance descriptors for that grade and cutscores and reviewed the OIB in the vicinity of the cutscore in question. Panelists made some changes, but in some cases, even though they felt a change might be in order, retained their original cutscore value because they felt a change could not be supported based on the borderline performance descriptors.

Validity Evidence from the Cross-grade Articulation

Because the decisions made by the cross-grade articulation panel were made with respect to the borderline performance descriptive in relation to the test items, the validity of the process was maintained.

State Articulation of the Results from the Vertical Articulation

Following the cross grade vertical articulation by the table leaders for ELA and Mathematics, state representative from 13 states/territories met, either in person or telephonically, to review the results from the cross grade vertical articulations. These results were considered separately for ELA and Mathematics. The state representatives first considered the results from Mathematics and made a few adjustments with the intent of making the overall results look more coherent. Next they considered the results from the cross grade vertical articulation for ELA. Because of time pressures, a decision was made to prepare a document that showed potential adjustments for the 2/3 cutscore for Grade 5 ELA that would be emailed confidentially to state representative for comment, vote, and potential follow-up phone call to discuss and vote on final decisions about potential changes to that cutscore.

Validity Evidence from State Articulation of Results from the Vertical Articulation

The participants in the meeting Thursday, August 13th and on the August 14th conference call were given the opportunity to contribute to, and vote for, any potential changes/adjustments to the cutscores. All changes were made with thoughtful consideration of the borderline performance descriptors or with a goal to provide coherent results. Respect for the panelists cutscores and expertise was maintained throughout the process.

Conclusion:

Because all steps in the process adhered to current best practices in standard setting, there appears to be good reason to consider the resultant cutscores to be valid and appropriate values for setting performance levels on NCSC's ELA and Mathematics Assessments for grades 3 -8 and 11.