APPENDIX 7-P: Review of the Standard Setting Report

Prepared by Measured Progress

On the Mathematics and ELA Standard Setting Process for

Grades 3 – 8 and 11

Barbara S. Plake, Ph.D.

Distinguished University Professor, Emeritus

University of Nebraska-Lincoln

November 17, 2015

Review of the Standard Setting Report

Prepared by Measured Progress

On the Mathematics and ELA Standard Setting Process for

Grades 3 – 8 and 11

Following the completion of the standard setting process to establish the cut scores for mathematics and ELA for grades 3 – 8 and 11 on the National Center and State Collaborative Assessments for students with significant cognitive disabilities, Measured Progress prepared a Technical Manual to document the process and results. The purpose of this report is to provide a review of that technical report to evaluate the degree to which the report meets industry standards for such technical reports for documenting standard setting procedures and provides sufficient validity evidence to support the intended interpretations and uses for the resultant cut scores.

Expectations for what should comprise a technical documentation and validity evidence from a standard setting study is articulated by Hambleton, Pitoniak, and Coppella in the chapter, Essential Steps in Setting Performance Standards on Educational Tests and Strategies for Assessing the Reliability of the Results (Hambleton, Pitoniak, & Coppella, 2012). According to Hambleton, Pitoniak, and Coppela, the following information should be provided in the technical documentation from a standard setting process:

- How the standard setting method was chosen
- Information on panelist recruitment and qualifications
- Agenda for the study

- Performance category descriptors and a description of process through which they were constructed
- Detailed description of the method implemented
- Types of feedback provided
- Facilitator scripts
- Summary of panelists ratings and how they changed across rounds
- Recommended performance standards after each round
- Standard error of measurement
- Standard error of judgment
- Summary of impact data
- Evaluation of validity information
- Recommendation of policy body for final cut scores

In order to review the quality of the report prepared by Measured Progress on the standard setting process on the NCSC's mathematics and ELA assessments for Grades 3 – 8 and 11, an analysis was conducted to compare the documentation provided in that report to these expectations for the content of technical documentation using each of the categories identified by Hambleton, Pitoniak, and Coppella.

<u>How the standard setting method was chosen</u>

The technical manual, on page 3, provides the following rationale for why the Bookmark Standard Setting Method was chosen:

> The standard setting process used was the bookmark procedure (see, e.g., Cizek & Bunch, 2007; Lewis et al., 1996; Mitzel et al., 2000). The main reason for choosing this method was that the assessment consists primarily of multiple-choice items but also includes some constructed-response items, and the bookmark procedure is appropriate for use with assessments that contain primarily or exclusively multiple choice items, scaled using item response theory (IRT) (Cizek & Bunch, 2007).

Information on panelist recruitment and qualifications

The technical manual contains a lengthy description on panelists' recruitment and the target goals for participation (Section 2.7). The intent was to have each of the participating states represented on the panels, and with a broad range of educator responsibilities. It was not possible to achieve gender or race balance across panels, but that was due to the lack of gender and race diversity in the lists of recommended panelists by the partner states.

Agenda for the study

The agenda for the study is contained in Appendix C.

Performance category descriptors and a description of process through which they were constructed

An entire chapter of the technical manual is devoted to the description of the process through which the Performance Level Descriptors were constructed. The final Performance Level Descriptors for ELA and Mathematics comprise Appendix B.

Detailed description of the method implemented

Chapter 3 of the technical report, Tasks Completed During the Standard Setting Meeting, provides a comprehensive description of how the method was implemented. Included in the process was a large group orientation followed by assembling in breakout groups by content and grade level, experience with the test, review of the PLD, completing an item map for the panelists to comment on the items as they appeared in the ordered item booklet (noting what competencies the item required and why it was more difficult than the previous item in the booklet), completing a description of the competencies of students on the borderline between adjacent performance categories,

participating in a practice exercise, and then completing a readiness evaluation. Following these preliminary tasks, panelists participated in three rounds of ratings, and completed evaluations.

Chapter 4 (Tasks Completed After the Standard Setting Meeting) describes the cross-grade articulation and policy adjustment meetings.

Types of feedback provided to panelists

Different types of feedback were provided to the panelists as they progressed from Round 1, to Round 2, and finally to Round 3. Following Round 1, panelists were given feedback about the individual cut scores for themselves and fellow panel members at their table. They were also told the median cut scores for the panel. Following Round 2, in addition to the feedback similar to what they received after Round 1, they also received impact data before making their Round 3 ratings. This information is documented in sections 3.10 – 3.14 in the technical documentation.

Facilitator scripts

Facilitator scripts are contained in Appendix I.

Summaries of panelists' ratings and how they changed across rounds

Following each round of ratings, the technical report presents, for each content area and grade panel, the median cut scores for Level 1, 2, 3, and 4. This information is contained in Sections 3.10 – 3.14 of the Technical Manual.

Recommended performance standard after each round

The recommended performance standards after each round are shown in the summaries of panelists' ratings in Sections 3.10 – 3.14 of the Technical Manual.

Standard error of measurement

This information is not contained in the Technical Manual. However, this index applies to the interpretation of the cut scores, not to their technical quality. Therefore, I do not view this as a shortcoming of the technical documentation.

Standard error of judgment

Instead of reporting this index, an index of the cohesion of the panelists, called the Mean Absolute Deviation, is presented for each round of rating for each content area and grade panel. This provides information that is relevant to the standard error of judgment.

Summary of impact data

This information is provided following Round 3 (page 27). It is also provided when any adjustment was made during the cross-grade articulation (pages 32 – 33) and policy meetings (pages 36).

Evaluation of validity information

Sometimes validity evidence is provided from a standard setting process that captures evidence from a variety of sources: procedural validity, internal validity, external validity. For the NCSC standard setting, the Technical Manual provides validity evidence to support procedural validity (through the documentation of the process and through the evaluations that gathered information on the panelists' readiness to make their ratings, and on various aspects of the standard setting process) and for internal validity (through the reporting of internal consistency of the ratings through the Mean Absolute Deviation of the ratings). There is strong evidence contained in the technical documentation to support both procedural and internal validity. Evidence for external validity is harder to obtain, especially this early in the program. However, the use of the

cross-grade articulation and policy adjustments provide a source of external evidence because the criterion used for making these adjustments was one of “reasonableness”.

Recommendation to the policy body

Partner states served as the policy body in approving the cut scores.

Conclusion

The technical documentation contained in the Standard Setting Report Prepared by Measured Progress on the Mathematics and ELA Standard Setting Process for Grades 3 – 8 and 11 meets industry standards for reporting on a standard setting study.

References

Hambleton, R.K., Pitoniak, M.J., & Coppella, J.M. (2012). Essential steps in setting performance standards on educational tests and strategies for assessing the reliability of the results. In G. J. Cizek (Ed.), Setting Performance Standards: Foundations, Methods, and Innovations, Second Edition, pp. 47-76 New York: Routledge.