Issues and considerations regarding Standard Setting – from The MARCES team

Standard Setting	Issues and Considerations

What is standard setting?
• Standard setting is the process of setting cut scores on a test, which indicates whether a student has achieved an established level of proficiency (Zieky & Perie, 2004). Based on performance on a test, students are classified into one of multiple proficiency levels such as basic, proficient, or advanced.
 	
General steps in standard setting
• Hambleton and Pitoniak (2010) summarized nine steps for setting performance standards:
  1. Select a standard setting method
  2. Choose a panel and design
  3. Prepare descriptions of performance categories
  4. Train panelists to use the method
  5. Collect ratings from panelists
  6. Provide feedback and facilitate discussion among panelists
  7. Compile panelist ratings and obtain performance standards
  8. Conduct panelist evaluation
  9. Compile validity evidence and prepare technical documentation.
	• Qualification of judges involved in standard setting process
	• Concept of borderline performance
	• Likelihood of errors of classification when using cut scores
	• Reliability of the classifications made on the basis of cut scores
	• Alignment of cut scores across grades for vertical comparison across grades
	• Importance of normative information

Standard setting methods
• Historically standard setting methods are classified into two categories: *test-centered methods* versus *examinee-centered methods*.
  • *Test-centered methods* require panelists to make judgment about assessment tasks
  • *Examinee-centered methods* require panelists to focus on examinees' test scores.
	• Standard-setting methods for constructed response items or performance tasks have not been well developed and fully researched while standard-setting for multiple-choice items are well developed and generally clear.
	• Standard-setting methods for computer-based tests, particularly computer adaptive tests have not been well developed and fully researched.

- Hambleton et al. (2000) used four categories and identify the corresponding standard setting methods.
1. Methods of reviewing test items and scoring rubrics:
   1.1. Angoff method
   1.2. Extended Angoff method
   1.3. Ebel method
   1.4. Nedelsky method
   1.5. Jaeger Method
   1.6. Bookmark method
2. Methods of reviewing examinees
   2.1. borderline group method
   2.2. contrasting groups method
3. Methods of reviewing examinee work:
   3.1. paper selection method
   3.2. body of work method
4. Methods of reviewing score profiles
   4.1. Judgment policy capturing method
   4.2. dominant profile method
   4.3. item cluster method

For detailed procedure for each method, please see
Zieky & Perie (2006). A primer on setting cut scores on tests of educational achievement. Princeton, NJ: Educational Testing Service.
http://www.ets.org/Media/Research/pdf/Cut_Scores_Primer.pdf

Hambleton, R. K., & Pitoniak, M. J. (2006). Setting performance standards. In Brennan, R. L. (4th Ed.), *Educational Measurement.* Westport, CT: Praeger Publisher.

PARCC and Smarter Balanced Standard-Setting for New Common Core State Standards

Current Status of Standard Setting: PARCC	**Current Status of Standard Setting:** Smarter Balanced (SB)
• PARCC consortium drafted grade-level performance level descriptions (PLDs) in ELA/Literacy and mathematics and released them to public in summer and fall, 2012. They are still in the process of revising performance level descriptions based on feedback. • PARCC assessment has five levels of performance and the name of each level has not been decided yet. To see PARCC PLDs, go to http://www.parcconline.org/ela-plds for ELA/Literacy and go to http://www.parcconline.org/math-plds for mathematics) • PARCC will conduct standard setting for the summative assessments in ELA and mathematics after the first full year of implementation (Summer, 2015) (http://www.parcconline.org/sites/parcc/files/PARCCTimelineforPolicymakers_5-26-13.pdf).	• SB drafted grade-level and content-level achievement level descriptions (ALDs) in October in 2012. The draft was approved by Governing State vote on March, 2013. • SB has four levels of performance. (To see SB ALDs, go to http://www.smarterbalanced.org/achievement-level-descriptors-and-college-readiness/) • Intended for different audiences, Smarter Balanced has four types of ALDs: (1) *Policy ALDs* for policymakers, used for test development and conceptualization of assessment. (2) *Range ALD* for test developers, used for item-writing guidance (3) *Threshold ALD* for standard-setting panelists, used for setting cut-scores and standard-setting guidance. (4) *Reporting ALD* for general audience (parents, students, teachers, etc.), used for test score interpretation. • After field-testing in spring 2014, SB will conduct standard-setting for the adaptive summative assessments in grades 3–8 and 11 in ELA and mathematics.

Test Delivery: PARCC	**Test Delivery:** Smarter Balanced (SB)
• *Fixed-form linear computer-based test* will be implemented. All students respond to the same set of test items that will be preselected. Test items are selected independent of students' previous answer choices.	• *Computer Adaptive Test (CAT)* will be implemented. Test items are selected adaptively based on students' previous responses. Test items given will depend on how students respond to questions.
Standard-setting method: PARCC	**Standard-setting method:** Smarter Balanced
• Specific standard-setting method has not been decided yet. • PARCC has reported that they will adopt a holistic approach such as Body of work method after field testing. • The process of setting cut score will consist of three rounds of structured discussion and individual scoring (see Appendix A(3)-(B) in PARCC application proposal). (1) A committee will talk about their review of the data. (2) Individual committee member will recommend the cut score with a rationale based on the data. (3) Each Committee member discusses his or her recommended cut score with other committee members and decide on a final recommendation and rationale for establishing the cut score.	• Specific standard-setting method has not been decided yet. • SB reported using item-based approach such as Bookmark method or a holistic approach such as Body of Work or combination of two approaches for standard setting after field testing. • Judgment from K-12 teachers and higher education faculty as well as multiple sources of empirical data (PISA, TIMSS, NAEP, SAT and ACT) will be used to help guide in setting benchmarks.
Consortia issues: PARCC	**Consortia issues:** Smarter Balanced
• Consortium standards versus. State's standards	• Consortium standards versus. State's standards • Selection of appropriate standard setting method in CAT environment • Standard setting method for the CAT will raise new challenges (e.g., how a test form is created for standard setting).