

**Smarter Balanced Cut Score Validation
Final Report
December 12, 2017**

Executive Summary

On November 1-3, 2017, a group of 11 English Language Arts/literacy educators and 10 mathematics educators from Smarter Balanced states met to consider the current cut scores for grades 9 and 10. Tables ES-1 and ES-2 show the final cut scores resulting from those reviews. Cut scores that have been modified are highlighted in yellow. The original versions of those cut scores are shown below each table.

**Table ES-1
ELA/Literacy Final Cut Scores and Impact**

	Cut Scores			% At or Above Cut Score			% In Group			
	Level 2	Level 3	Level 4	Level 2	Level 3	Level 4	Level 1	Level 2	Level 3	Level 4
Grade 8	2487	2567	2668	72%	41%	9%	28%	31%	32%	9%
Grade 9	2489	2571	2672	67%	38%	9%	33%	29%	29%	9%
Grade 10	2491	2577	2678	76%	47%	13%	24%	29%	34%	13%
Grade 11	2493	2583	2682	72%	41%	11%	28%	31%	30%	11%

Original Cut Scores:

- Grade 9 Level 2 – 2488 (increased by 1 point)
- Grade 9 Level 4 – 2670 (increased by 2 points)
- Grade 10 Level 4 – 2677 (increased by 1 point)

**Table ES-2
Mathematics Final Cut Scores and Impact**

	Cut Scores			% At or Above Cut Score			% In Group			
	Level 2	Level 3	Level 4	Level 2	Level 3	Level 4	Level 1	Level 2	Level 3	Level 4
Grade 8	2504	2586	2653	62%	32%	13%	38%	30%	19%	13%
Grade 9	2517	2601	2676	56%	27%	9%	44%	29%	18%	9%
Grade 10	2533	2614	2697	61%	33%	12%	39%	28%	21%	12%
Grade 11	2543	2628	2718	60%	33%	11%	40%	27%	21%	11%

Original Cut Scores:

- Grade 9 Level 2 – 2515 (increased by 2 points)
- Grade 9 Level 3 – 2599 (increased by 2 points)
- Grade 10 Level 2 – 2529 (increased by 4 points)

Impact

For both ELA/literacy and mathematics, impacts of changes were very small and in some instances 0. For ELA/literacy, the three changes resulted in 1% fewer students in the grade 9 Level 2 and

above with 0% change in the number of students at Level 3 or above (due to rounding). For mathematics, the change in the grade 9 cut scores at Levels 2 and 3 resulted in 1% fewer students in Level 2 or above and 1% fewer students in Level 3 or above. Moving the grade 10 Level 2 cut score from 2529 to 2533 decreased the number of students in Level 2 or above by 1%, but the percentages of students at Level 3 or above was unchanged.

Evaluation of Final Cut Scores

After cut scores had been finalized, participants had an opportunity to voice their support for them as well as for the process by which the final cut scores had been derived. In general, the English Language Arts/literacy group gave high marks to both the cut scores and the process, with 92 percent affirming the validity of the final cut scores. The mathematics group, while generally endorsing the process, was far less convinced of the validity of the final cut scores. Only 12 percent agreed or strongly agreed that the final cut scores were reasonable, including those they themselves had set. Their chief concern was a lack of clearly defined expectations of ninth and tenth graders.

Introduction and Background

Measurement Incorporated (MI) contracted with the Smarter Balanced Assessment Consortium (Smarter Balanced) to conduct validation of cut scores for grades 9 and 10 as part of a larger contract to create item maps and playlists for 96 interim assessment blocks (IABs).

Smarter Balanced established four achievement levels for its summative assessments, with Level 3 serving as the benchmark for state and federal reporting. In 2014, MI worked with Smarter Balanced to conduct standard setting for grades 3-8 and 11. In that engagement, MI used a Bookmark procedure (Cizek & Bunch, 2007, Chapter 10), employing ordered item booklets (OIBs) and item maps that panelists used to identify thresholds for Levels 2, 3, and 4.

As states expressed interest in using the high school tests for grades 9 and 10, Smarter Balanced facilitated the validation of cut scores for those grades as well. The American Institutes of Research (AIR) conducted an analysis of cut scores set in 2014 and interpolated cut scores for grades 9 and 10 (AIR, 2015). Those results are presented in Tables 1 and 2. The scale cut scores for grades 9 and 10 are **bolded** in these tables. It was these cut scores – for Levels 2-4 for grades 9 and 10 for English Language Arts/literacy (ELA/literacy) and mathematics – that MI undertook to validate in 2017.

Table 1
Cut Scores for ELA/Literacy Grades 8-11

Grade	Level 2	Level 3	Level 4
8	2487	2567	2668
9	2488	2571	2670
10	2491	2577	2677
11	2493	2583	2682

Table 2
Cut Scores for Mathematics Grades 8-11

Grade	Level 2	Level 3	Level 4
8	2504	2586	2653
9	2515	2599	2676
10	2529	2614	2697
11	2543	2628	2718

Methodology

Cut score validation occurred on November 1-3, 2017 with 11 ELA/literacy and 10 mathematics participants from Smarter Balanced states, nominated by their states. Orientation began on the afternoon of November 1, and work concluded at noon on November 3. The work plan began with these givens:

1. The purpose of the grade 9 and 10 cut scores is to provide on-track information to schools, districts, and states. Cut scores for grade 11 were intended to reflect readiness for college. The distinction between college-ready and on-track is an important one. For grades 3-8, cut scores were meant to reflect student work that put them on track to perform satisfactorily at the next grade and, if remaining on track until grade 11, being ready for college by the end of high school.
2. Grade 8 and grade 11 cut scores cannot be changed.
3. Grade 9 cut scores cannot be below grade 8 cut scores or higher than grade 10 cut scores.
4. Grade 10 cut scores cannot be lower than grade 9 cut scores or higher than grade 11 cut scores.
5. The range of cut score possibilities for ELA/literacy is quite small for Level 2 and only moderately larger for Levels 3 and 4.

Additionally, participants would need to understand what the current cut scores are, how they were derived, and how they impact students in grades 9 and 10. Beyond that, they would need to know what they would be expected to accomplish and how they would be expected to accomplish it. The work was thus broken down into the following tasks, approved by Smarter Balanced staff in early October 2017.

- Orient participants to the goals and tasks
- Review practice tests and performance tasks
- Review the grade 11 achievement level descriptors (ALDs)
- Review the 2014 high school tests, page by page
- Verify or move grade 9 and 10 cut scores
- Draft interpretive language to go with grade 9 and 10 cut scores
- Evaluate the process

The overall agenda is presented in Table 3. Merged cells in Table 3 indicate joint meetings; otherwise the ELA/literacy and mathematics groups met separately. Details are provided in the sections that follow.

**Table 3
Agenda**

Date	Time	ELA/Literacy	Mathematics
11/1/2017	1:00 P.M.	Welcome participants and provide a general orientation	
	1:30	Review practice tests, performance tasks, and other illustrative item sets	Review practice tests, performance tasks, and other illustrative item sets
	3:15	Review the grade 11 ALDs	Review the grade 11 ALDs
	3:55	Wrap-up	Wrap-up
	4:00	Adjourn	Adjourn
11/2/2017	8:30 A.M.	Orient participants to the tasks of validating cut scores and drafting interpretive language	
	9:30	Review OIB page by page	Review OIB page by page
	Noon	Lunch	
	1:00 P.M.	Validate/adjust cut scores for grades 9 and 10	Validate/adjust cut scores for grades 9 and 10
	3:55	Wrap-up	Wrap-up
	4:00	Adjourn	Adjourn
11/3/2017	8:30 A.M.	Review activities of previous day with Smarter Balanced leadership and facilitators	
	9:15	Draft interpretive language for cut scores	Draft interpretive language for cut scores
	11:15	Evaluate Process	Evaluate Process
	11:45	Wrap-up	Wrap-up
	Noon	Adjourn	Adjourn

Orient Participants to the Goals and Task

Work began with an orientation session on Wednesday afternoon and a follow-up session on Thursday morning. The Thursday afternoon session focused on the overall goals for the meeting and a breakdown of the various work sessions. The longer Thursday morning session focused on the specific tasks of ordered item booklet review and ground rules for modifying a cut score. The orientation consisted of a PowerPoint presentation (included under separate cover) that covered these topics:

- The 2014 standard setting process

- The subsequent creation of cut scores for grades 9 and 10
- Limitations imposed by the proximity of cut scores for grades 8-11
- The creation and use of ALDs
- The relationship between ALDs and cut scores

MI staff provided a brief summary of the 2014 standard setting and its relevance to the current task. They then presented an overview of the AIR study and the starting points for the cut-score validation task. MI staff then turned their attention to the development and use of threshold ALDs.

Review Practice Tests and Performance Tasks

After the Wednesday afternoon orientation, the ELA/literacy and mathematics groups separated to begin review of their respective tests. They used the Smarter Balanced practice tests and performance tasks to demonstrate technology enhanced items and other item types to gain a better understanding of the types of items students encounter as well as the relative difficulty of those items. Facilitators projected items on a screen and lead discussions of the item characteristics.

Review the Grade 11 ALDs

Once participants had examined the tests, they turned their attention to the ALDs. Because there are currently no grade 9 or grade 10 ALDs, participants used the grade 11 ALDs to understand how the grade 11 cut scores came to be in terms of the difficulty of the items on the pages in the ordered item booklet that correspond to key phrases in the ALDs. As participants reviewed the grade 11 ALDs, they created tables of key phrases to use later in reviewing the tests.

Review the 2014 High School Tests, Page by Page

The facilitators presented the same ordered item booklets (OIBs) that panelists used in October 2014 to set cut scores for grade 11, with minor exceptions. NAEP, PISA, and other non-Smarter Balanced items were removed, along with items in the Mathematics OIB that had been deemed to be outside the scope of grade 9 and 10 coursework. The mathematics committee later reviewed the items that been deemed out of scope. The purpose of this exercise was to answer the following standard-setting questions for each item:

- What knowledge or skill must a student have in order to answer this item correctly or obtain this score on this constructed-response item?
- What makes this item more difficult than previous ones?

In the 2014 standard setting, panelists reviewed every item in the OIB answering these two questions before identifying any thresholds. Only after reviewing the entire OIB did panelists go back and bookmark pages to set thresholds for Levels 2, 3, and 4. Participants followed this approach in 2017 for grade 9 and 10 cut score validation as well.

Verify or Move Grade 9 and 10 Cut Scores

Given the relatively limited amount of maneuvering room, particularly for ELA/literacy, participants discussed the viability of current grade 9 and 10 cut scores and options for moving them. The facilitators used the OIB to locate the grade 9 Level 2 cut score and began the discussion there. Participants were also able to see the item on the OIB page that would come closest to the grade 8 level 2 cut score as well as the current grade 10 Level 2 cut score. Given that the grade 8 cut (2487 for ELA/literacy or 2504 for mathematics) was considered inviolate, the range of possible alternative cut scores for grade 9 Level 2 was fairly small. However, that did not mean that discussion would be curtailed. Participants had the ALDs for grade 11 and access to other support materials that were available to panelists in 2014.

The final cut scores set in 2014 do not always correspond exactly to the pages in the OIBs. Instead, some fall between pages. Participants were able to see the cognitive requirements of items in the vicinity of each cut score for each grade by examining the items on either side of the final 2014 cut score. This review gave participants an understanding of the content basis of the cut scores as well as the relatively small distances between common cut scores across grades. For example, at Level 2, the grade 9 ELA/literacy cut score was only one scale score point higher than the grade 8 cut score, and the grade 10 ELA/literacy cut score was only three scale score points higher than the grade 9 cut score.

Because of the dynamic nature of the grade 9 and grade 10 cut scores, participants discussed both for a given level simultaneously. Specifically, they focused on the implications for the grade 10 cut score for any potential change in a grade 9 cut score and *vice versa*. At the end of the discussion, however, the motion to change a cut score was for a single grade, with grade 9 motions being addressed first and grade 10 second. Thus, the sequence of motions was grade 9 Level 2, grade 10 Level 2, grade 9 Level 3, grade 10 Level 3, grade 9 Level 4, and grade 10 Level 4.

It was possible to make a motion to accept the current cut score or some specific alternative. Given that a motion to change a cut score would effectively override previous work, a 2/3 majority vote was required, as was the case in vertical articulation in 2014. It was not necessary to entertain motions for all six cut scores. If any existing cut scores seemed totally appropriate, participants had an opportunity at the end of the session to introduce a motion to accept all remaining cut scores.

The process began on Thursday afternoon. The ELA/literacy group finished by the end of the day, and the mathematics group continued into Friday morning.

Draft Interpretive Language for Grade 9 and 10 Cut Scores

This task was added to the schedule to be completed if all other tasks could be completed by Thursday afternoon. The intent of this task was that with validated cut scores in hand and a clear understanding of what those cut scores mean in terms of item content expectations, participants would suggest interpretive language to accompany the cut scores. Using the tables they generated in the review of the grade 11 ALDs, participants were to suggest additional words or phrases to be used to describe the levels for grades 9 and 10. Because ELA/literacy content standards treat grades 9 and 10 as a single unit, the plan for the ELA/literacy group was to create a single table. High school mathematics content, on the other hand, is more course oriented, so the plan for that group was to create two tables, one for grade 9 and one for grade 10.

Evaluate the Process

At the end of the process, MI facilitators distributed evaluation forms for participants to complete.

Results

Cut scores. Tables 4 and 5 present the final cut scores for ELA/L and Mathematics for grades 9 and 10. Changes from the original 2015 cut scores are **highlighted**, and the original cut scores for those levels are listed below each table.

Table 4
ELA/Literacy Final Cut Scores and Impact

	Cut Scores			% At or Above Cut Score			% In Group			
	Level 2	Level 3	Level 4	Level 2	Level 3	Level 4	Level 1	Level 2	Level 3	Level 4
Grade 8	2487	2567	2668	72%	41%	9%	28%	31%	32%	9%
Grade 9	2489	2571	2672	67%	38%	9%	33%	29%	29%	9%
Grade 10	2491	2577	2678	76%	47%	13%	24%	29%	34%	13%
Grade 11	2493	2583	2682	72%	41%	11%	28%	31%	30%	11%

Original Cut Scores:

Grade 9 Level 2 – 2488 (increased by 1 point)

Grade 9 Level 4 – 2670 (increased by 2 points)

Grade 10 Level 4 – 2677 (increased by 1 point)

Table 5
Mathematics Final Cut Scores and Impact

	Cut Scores			% At or Above Cut Score			% In Group			
	Level 2	Level 3	Level 4	Level 2	Level 3	Level 4	Level 1	Level 2	Level 3	Level 4
Grade 8	2504	2586	2653	62%	32%	13%	38%	30%	19%	13%
Grade 9	2517	2601	2676	56%	27%	9%	44%	29%	18%	9%
Grade 10	2533	2614	2697	61%	33%	12%	39%	28%	21%	12%
Grade 11	2543	2628	2718	60%	33%	11%	40%	27%	21%	11%

Original Cut Scores:

Grade 9 Level 2 – 2515 (increased by 2 points)

Grade 9 Level 3 – 2599 (increased by 2 points)

Grade 10 Level 2 – 2529 (increased by 4 points)

Motions and votes. All cut score changes were made through a democratic process requiring a motion, second, discussion, and vote. To pass, a motion required a 2/3 majority, as had been the case in vertical articulation in 2014. Tables 6 and 7 show the results of all motions. There were several failed motions in the mathematics group. The votes are listed as Yes-No-Abstain.

Table 6
ELA/Literacy Motions and Votes

Motion	Date/Time	Second	Vote	Result/Action
Adopt 2489 as the Level 2 cut score for grade 9 and to adopt 2491 as the level 2 cut score for grade 10.	11/2/17 3:05 p.m.	Yes	11-0-0	Motion carried
Adopt 2571 as the Level 3 cut score for grade 9 and to adopt 2577 as the level 3 cut score for grade 10.	11/2/17 3:30 p.m.	Yes	10-1-0	Motion carried
Adopt 2672 as the Level 4 cut score for grade 9 and to adopt 2678 as the level 4 cut score for grade 10.	11/2/17 3:58 p.m.	Yes	11-0-0	Motion carried
Adopt all cut scores as final.	11/2/17 4:02 p.m.	Yes	11-0-0	Motion carried

Table 7
Mathematics Motions and Votes

Motion	Date/Time	Second	Vote	Result/ Action
Change Grade 9 Level 2 cut score from 2515 to 2517	11/2/17 2:40 p.m.	Yes	2-6-2	Motion failed
Consider graphical representation as a lens to decide cut scores	11/2/17 3:30 p.m.	No		Motion failed for lack of second
Change Grade 10 Level 2 cut score from 2520 to 2533	11/2/17 3:37 p.m.	Yes	7-2-1	Motion carried
Change Grade 9 Level 2 cut score from 2515 to 2517	11/2/17 3:41 p.m.	Yes	7-3-0	Motion carried
Change Grade 9 Level 3 cut score from 2599 to 2601	11/2/17 3:49 p.m.	Yes	7-3-0	Motion carried
Change Grade 10 Level 3 cut score from 2614 to 2611	11/2/17 3:53 p.m.	Yes	3-6-1	Motion failed
Change Grade 9 Level 4 cut score from to 2680	11/3/17 8:42 a.m.	Yes	5-3-2	Motion Failed
Change Grade 9 Level 4 cut score from 2676 to 2674	11/3/17 8:50 a.m.	Yes	2-7-1	Motion failed
Change Grade 9 Level 4 cut score from 2676 to 2682	11/3/17 8:55 a.m.	Yes	6-4-0	Motion failed
Accept all remaining cut scores	11/3/17 9:25 a.m.	Yes	5-0-5	Motion failed
Accept Grade 10 Level 3 cut score of 2614	11/3/17 9:45 a.m.	Yes	8-0-2	Motion carried

The final Mathematics vote to accept the grade 10 Level 3 cut score was necessary because a previous vote to adopt all remaining Mathematics cut scores failed. Five participants had voted in favor of the motion, and none voted against it, but five abstained, denying the group a 2/3 majority.

Interpretive language for grade 9 and 10 cut scores. The ELA/literacy participants began this activity early Friday morning. The facilitators explained the purpose and intended process of the activity and solicited input from the panelists with respect to both process and desired outcomes. After a rather lengthy discussion, the ELA/literacy group discontinued the activity. The mathematics group

continued reviewing cut scores on Friday morning. When they finished (around 10 A.M.) and were given an opportunity to begin this activity, several stated that there was not enough time to complete the task in a quality manner. There was a brief discussion about the nature of the task. One member of the group who had been on the committee that drafted the grade 11 ALDs pointed out that it had taken a week or more to complete the task. Although the facilitator explained that the group was not being asked to draft grade 9 or 10 ALDs, the group stood firm in their unwillingness to complete this task. The mathematics group facilitator checked with the ELA/Thus, the Mathematics group did not complete this task either.

Evaluations. Tables 8 and 9 present the results of evaluations of the process and outcomes of the ELA/literacy and mathematics cut score validation activities (total of 11 for ELA/literacy and 10 for mathematics). The entries are defined as follows: SD = Strongly Disagree; D = Disagree; ? = Uncertain; A = Agree; SA = Strongly Agree. Percentages may not add to 100 due to rounding. Comments are included in Appendix A. ELA/literacy participants made few comments. Some mathematics participants entered multiple comments.

Table 8
Evaluation Results for ELA/L (N=11)

Statement	SD	D	?	A	SA
1. The background information about the original grade 9 and 10 cut scores was clear and informative.	0	0	2	4	4
2. The discussions of achievement level descriptors was useful.	0	1	3	2	4
3. The facilitator clearly explained the process for reviewing and adjusting cut scores.	0	1	1	4	5
4. The facilitator clearly explained the implications of adjusting a cut score in terms of impact on students.	0	1	2	5	3
5. I was able to participate in the discussion of cut score adjustment/validation.	0	0	1	2	8
6. The process was fair.	0	1	0	3	7
7. The process was orderly.	0	1	2	3	5
8. The final cut score for grade 9 Level 2 is reasonable.	0	0	0	7	4
9. The final cut score for grade 9 Level 3 is reasonable.	0	0	0	7	4
10. The final cut score for grade 9 Level 4 is reasonable.	0	0	0	7	4
11. The final cut score for grade 10 Level 2 is reasonable.	0	0	0	7	4
12. The final cut score for grade 10 Level 3 is reasonable.	0	0	0	7	4
13. The final cut score for grade 10 Level 4 is reasonable.	0	0	0	7	4
14. The suggestions made by my group for interpretive language cut scores were reasonable.	0	0	2	4	4
Total	0	5	13	69	64
%	0	3	9	46	42

Table 9
Evaluation Results for Mathematics (N=10)

Statement	SD	D	?	A	SA
1. The background information about the original grade 9 and 10 cut scores was clear and informative.	0	1	3	5	1
2. The discussions of achievement level descriptors was useful.	0	2	3	3	2
3. The facilitator clearly explained the process for reviewing and adjusting cut scores.	0	2	0	5	3
4. The facilitator clearly explained the implications of adjusting a cut score in terms of impact on students.	0	5	0	3	2
5. I was able to participate in the discussion of cut score adjustment/validation.	0	0	0	4	6
6. The process was fair.	0	1	4	5	0
7. The process was orderly.	0	1	1	6	1
8. The final cut score for grade 9 Level 2 is reasonable.	0	3	6	1	0
9. The final cut score for grade 9 Level 3 is reasonable.	0	3	6	1	0
10. The final cut score for grade 9 Level 4 is reasonable.	0	2	8	0	0
11. The final cut score for grade 10 Level 2 is reasonable.	0	3	5	2	0
12. The final cut score for grade 10 Level 3 is reasonable.	0	3	5	2	0
13. The final cut score for grade 10 Level 4 is reasonable.	0	2	7	1	0
14. The suggestions made by my group for interpretive language cut scores were reasonable.	0	0	7	2	1
Total	0	28	55	40	16
%	0	20	39	29	11

SD = Strongly Disagree; D = Disagree; ? = Uncertain; A = Agree; SA = Strongly Agree

In general, the English Language Arts/literacy group gave high marks to both the cut scores and the process, with 92 percent affirming the validity of the final cut scores. The mathematics group, while generally endorsing the process, was far less convinced of the validity of the final cut scores. Only 12 percent agreed or strongly agreed that the final cut scores were reasonable, including those they themselves had set. Their chief concern was a lack of clearly defined expectations of ninth and tenth graders.

Discussion

The grade 9 and 10 cut scores derived by AIR in 2015 appear to be reasonable for the purposes they were created to serve in that they place expectations for students in grades 9 and 10 meaningfully between those for students in grades 8 and 11. Moreover, the participants in the cut score validation task readily perceived the logic of those cut scores and had no serious objections to any of them. In all instances in which they recommended changes to a cut score, the overall impact was one percent or less in terms of which students were classified at which achievement level.

The differences in outcomes for the mathematics group, relative to the ELA/literacy group, stemmed from a difference in perspective of the nature of the validation task itself. The mathematics group fully understood the process of deriving one number (in this case, a cut score) through linear or curvilinear interpolation between two other numbers. Indeed, the group found the procedure quite reasonable. The issue was one of perception of their responsibility. The mathematics group viewed its task as one of validating a mathematically derived cut score through content review. While members appreciated the opportunity to view the items on which their decisions were to be based, they were reluctant to render firm decisions in the absence of approved achievement level descriptors for grades 9 and 10 from which they could derive content-based decisions about the cut scores.

With regard to the task of drafting interpretive language for grade 9 and 10 cut scores, it may be constructive to take a closer look at the discussions in both rooms. Indeed, those discussions shed light on the cut score validation activity as well.

When the ELA/literacy group took up the task, one of the panelists immediately suggested that they examine the grade 8 and 11 ALDs to determine and evaluate differences. Another panelist produced the ELA/literacy ALDs for grades 6-8 and 11 and began to read from them, pointing out the similarities among all sets of ALDs for virtually all targets and claims. The group decided that the only real distinguishing factor from grade to grade was the complexity of text(s) with which students were confronted. One panelist used the term “recursive” to describe the skills and knowledge found in the ALDs, stating that the very same skills were involved in the development of literacy skills over the years and that such is reflected in the ALDs.

Although it might appear from this discussion that the panelists were balking at the task, they were actually engaged in a very lively, informative discussion regarding literacy skills, student development, text complexity, college readiness, and ALDs. The panelists came to consensus that they did not have the necessary information to meaningfully accomplish the activity. They said they would need, at the very least, a clearer notion of what texts and questions were administered to grades 9 and 10 students and what expectations (ALDs) teachers might have for their students’ performance given such test material.

One of the MI facilitators notified Smarter Balanced personnel, Dr. Patricia Reiss, that the participants did not think the activity could be completed meaningfully. Dr. Reiss listened to the panelists’ discussion points and decided that the activity could be curtailed.

A telling anecdote from the morning was when one of the MI facilitators asked educators why they thought the distinction between grades 9 and 10 skills was impossible, they stated that, they would like to see “high school” ALDs, not ALDs by high school grade level. They did not think grade-specific ALDs would be meaningful or useful.

When the mathematics group began the task mid Friday morning, one of the participants who had participated in the creation of the initial ALDs noted that the time allotted for the task was not nearly enough, that her group had spent at least a week just developing draft language for the initial threshold ALDs. The rest of the group agreed. After conferring with the ELA/L facilitator and Smarter Balanced staff, the Mathematics facilitator agreed to halt the activity.

The participants in both groups took their tasks very seriously, but the mathematics group in particular struggled somewhat with the notion of validating statistically derived cut scores through content review when no content expectations were available to guide them. This issue, combined with the fact that at grades 9 and 10 there is the added factor of traditional vs. integrated Mathematics curriculum, kept the group from reaching consensus with regard to cut scores. Several times during the discussion of one cut score or another (particularly for grade 9), someone would ask whether the cut score they were being asked to validate was for students enrolled in traditional or integrated mathematics curricula. Many of the abstentions were by participants who could not get past that divide. For them, the issue seemed to be within-state differences in the way the integrated approach is implemented rather than across-state differences in approach.

References

- American Institutes for Research (2015). *Establishing Cut-Scores for Common Grades 9 and 10 ELA/L and Mathematics*. Washington, DC: Author.
- Cizek, G. J., & Bunch, M. B. (2007). *Standard Setting: A Guide to Establishing and Evaluating Performance Standards on Tests*. Thousand Oaks, CA: Sage.

Appendix A
Evaluation Comments

English Language Arts/literacy

Facilitators good Process a bit opaque - small group/individual work mixed in would've been more efficient We spent too much time on process questions

Mathematics

The process was difficult without ALD's for 9th and 10th grades. It seemed like a lot of guess work, which many group members were not comfortable with. Many people got frustrated with the process and abstained from voting, so we really could not move forward.
The process for determining the 9th and 10th grade cut scores needs more work. First, the content shown was not just 9th grade content. So to ask a group to make recommendations with no ALD or just an ordering at least of how 9th graders performed (OIB) does not make sense. However, the problems in the OIB presented were appropriate for 10th graders because no matter which pathway (Integrated or Traditional) a student takes, I think the cut scores could be better determined from the given information. I think making any suggestions for 9th grade based on the information given is not appropriate at this point. I would rather have teachers/educators come together and agree on performance outcomes for Grade 9 (irrelevant of pathway) at the very least. We have the greater responsibility to make sure recommendations are an accurate reflection no matter what pathway a student is in in high school. While one can argue that ultimately it's the state that decides, there is a reliance on SBAC providing the most accurate cut scores as possible, otherwise, I fear that the buck just gets passed off to someone else who may not have the expertise or the political neutrality to make decisions.
I think we would have accomplished the same amount of work in one day or maybe a day and a half. We are used to working long days. We could have started at 8, had a 1/2 hour long and worked til 5, easily.
Based on the 2014 data we had to review, it was not apparent that the cut scores were reasonable. We came to the best consensus based on the data and information we were given.
While we were reviewing cut scores for a high school test, there has to be agreement on where content should fall - where within each grade level - only then can a group of educators determine specific grade level achievement level descriptors. I am excited for the future work that will re-visit and validate the cut scores across the grade levels.
It seemed to me that the true consensus was that what we were asked to do was inappropriate. If we accepted all of the constraints, we agreed on a cut score, but most in the room did not agree with those constraints.
We did not have an agreed upon picture of a 9th or 10th grade student. We did not agree if we were following a traditional or integrated pathway. We did not feel that the ordered item booklet was an accurate reflection of today's HS students.
I felt that we vacillated between different criteria for our decisions (OIB, content at grade level, equity issues, constraints of 8th and 11th grade cut scores). The facilitators were wonderful and respectful in a very difficult situation

<p>The scale scores from the field test assigned to the OIB reflect how "difficult" items were at that time based on a variety of factors, including exposure to new ideas from CCSS at the ITS level. The scale scores at times contradict the threshold and range ALD for HS math.</p>
<p>This is problem when we then look to adjust the scale scores based on what to expect from 9th and 10th grade students using the items assigned scale score with no specific ALDs for 9th and 10th grade. I strongly suggest that the scale scores are re-evaluated and a new vertical articulation is done using some guidance from created 9th and 10th ALDs.</p>
<p>This would have been easier to look at OIB and propose cut scores in smaller groups first and then have the whole group. More questions at varying levels were needed. This process definitely makes me want to get more involved in this work to understand better and to bring current HS teacher voice into discussions, decisions, etc.</p>
<p>Group size might be problematic as one voice could dominate/sink process. Communication before event could have been better. I was more OK with levels Day 2; Day 3 it seemed people lost sight of purpose a bit. I worry they voted for things Day 2 and then changed mind and took it out on level 4 Day 3. Thank you for putting up with us.</p>
<p>Would have appreciated opportunity to do practice tests, looking at ALD (Wed afternoon), and looking at OIB (Thurs morning) in small groups or pairs instead of altogether. Feeling lack of confidence in establishment of cut scores - doesn't feel related to content or standards - just to interpolation of data. Questioning why this should be sent to states for use.</p>
<p>Facilitators were incredibly professional, well-organized, and neutral in a somewhat tense environment at times. Logistics were not very well-communicated - meeting dates changed, no info provided in advance about breakfast or lunch on Wednesday, etc. Found # and flow of "observers" less than helpful.</p>
<p>I sincerely commend Michael and Lisa for their tenacity, patience, calmness and professionalism in handling our group. It was "tough" three days discussions but was worthwhile to learn perspectives from the different states. I value the discussions and process.</p>
<p>I suggest the following for future work: 1.) Ask the participants to bring their copies of standards and ALDs (I'm kinesthetic and I would have grounded my argument based from evidence!) 2.) Invite a group of HS Ts to work on Grades 9 and 10 ALDs (perhaps start with our group!)</p>
<p>I really enjoyed the conversation with everyone in the room. The facilitators were gracious in the face of some push back. I think having a test blueprint will be important for these cut scores to be used well by teachers. I will be interested to see how year 1 students score on the comprehensive interim if it is used as an intermediate final exam for students who are not yet ready to take the summative test.</p>