

CONVENING AN ARTICULATION PANEL AFTER A STANDARD SETTING MEETING: A HOW-TO GUIDE

Marianne Perie
Center for Assessment
November 2, 2006

The final technical phase in many large-scale assessment standard setting projects, when multiple grades are tested, is the articulation meeting. The purpose of the articulation meeting is to bring policymakers and other stakeholders together to review the results of the standard-setting workshops in the context of the state assessment program goals. The articulation panelists review the results of the cut scores as compared to the results in adjacent grades or across different subjects in the same grade taking into consideration the desirability of consistency across subjects in terms of both the percentage of points required to reach each level and the percentage of students reaching each level of performance.

The facilitator will provide them information on how the cut score was determined, where it was placed, and the projected consequences for the students. The meeting could be accomplished in about half a day. The following sections list each task that should be completed prior to and during the articulation workshop. For the tasks to be completed during the articulation workshop, the approximate time required is specified. A sample agenda and recording form are included at the end of this paper.

Choose Panel

We recommend the articulation panel be comprised of a subset of the original standard-setting panelists supplemented by state or local policymakers, including district superintendents and curriculum supervisors. The articulation panel could be smaller than the standard-setting people, possibly 6-10 people. For instance, we could invite 5 state and district policy experts, perhaps including up to two members of the state Department of Education, as well as three members of the original standard-setting committee. The original standard-setting panelists will be asked to provide their perspective on the process and the thinking that went into determining the initial cut scores.

Create Table Shells

Between the standard-setting meeting and the articulation meeting, several tables and figures should be prepared.

Table 1 provides information on the range of raw score cut scores from one round to the next.

Table 1: Sample Table Shell of Raw Score Cut Scores by Round

Proficient	Round 1	Round 2	Round 3
Low			
High			
Average			
Advanced	Round 1	Round 2	Round 3
Low			
High			

Median

Table 2 provides information about the student subgroups and the percentage of students who would be categorized in each performance level given the cut scores recommended by the standard-setting panel. A table like table 2 should be prepared for the assessment being evaluated as well as for all assessments that this assessment is being compared to. For instance, if the articulation panel is evaluating grade 4 reading and grade 6 reading, a table should be prepared for grades 3 and 5 reading in addition to grades 4 and 6 reading. The State might want to consider providing tables for writing or mathematics at similar grade levels as well.

Table 2: Sample Table Shell of Percent Distribution of Students Using Standard-Setting Recommended Cut Scores

Demographic Group	Number of Students	Mean raw score	% Basic	% Proficient	% Advanced
All Students					
Special Education					
Limited English Proficient					
Gender					
Female					
Male					
Race/ethnicity					
American Indian					
Asian					
Black					
Hispanic					
White					
Economic Status					
Economically Disadvantaged					
Non-Economically Disadvantaged					
Migrant Status					
Migrant					
Non-migrant					
Proficient cut = X; Advanced cut = Y					

It is often helpful to provide percent distribution data, impact data, for scores on either side of the cut score to show the articulation panelists the consequence of moving the cut score. This can be done in several ways. One way is to simply provide the cut score +/- 2 points (or however many the policymakers want to consider) as in Table 3. (Note that half-scores should be used only if they are viable score points.) The range around the cut score also can be determined using the standard error of measurement (SEM) or the standard error of judgment (SEJ), as in Table 4. All appropriate subgroups should be included in tables 3 and 4.

If SEMs and SEJs are used it is important to clearly explain these concepts to the articulation panelists. Typically, we say that no test is perfect and no standard setting procedure is perfect. Every test has error of measurement and every standard setting has what could be termed "error of judgment." Measurement error occurs because no instrument measures a student's level of knowledge and skills precisely. Judgment error occurs when not every standard setting

panelist agrees on a cut point. The state Department of Education and their State Board may desire to reduce the number of examinees who fall below the panel recommended cut scores due to one of these sources of error in order to reduce the numbers of “false negatives” and thus will decide to lower the cut score(s). On the other hand, they may desire to reduce the number of examinees who attain a score above the recommended cut score due to the error at each level in order to reduce the number of “false positives” and thus raise the cut score(s).

Table 3. Sample Table Shell of Results of the Standard-Setting Workshop: Proficient

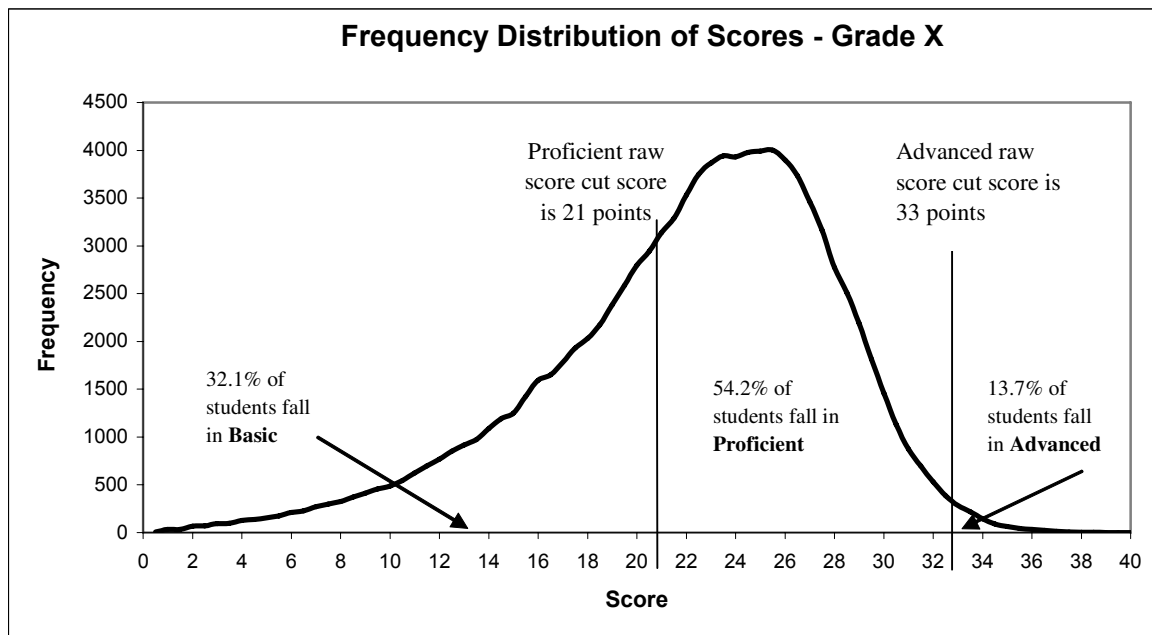
		Percentage of Students At or Above Each Score				
	Percent Raw of total score score	Total	Asian	Black	Hispanic	White LEP SWD
	23					
	22.5					
	22					
	21.5					
Recommended Proficient cut score ►	21					
	20.5					
	20					
	19.5					
	19					

Table 4: Sample Table Shell of Results of Standard Setting Workshop: Proficient

		Percentage of Students At or Above Each Score				
	Percent Raw of total Score score	Total	Asian	Black	Hispanic	White LEP SWD
Recommended cut score + 2 SEMs ►	26					
	25					
Recommended cut score + 2 SEJs ►	24					
Recommended cut score + 1 SEM ►	23					
Recommended cut score + 1 SEJ ►	22					
Recommended Proficient cut score ►	21					
Recommended cut score - 1 SEJ ►	20					
Recommended cut score - 1 SEM ►	19					
Recommended cut score - 2 SEJs ►	18					
	17					
Recommended cut score - 2 SEMs ►	16					

In addition, to help solidify the percentage distribution of scores, it is helpful to create a graphic of the distribution of student scores such as in Figure 1 and drawing in the cut scores to show the percentage of students scoring at each level.

Figure 1: Sample Distribution of Students with Standard-Setting Recommended Cut Scores



Smoothing the Cut Scores

As a final preparatory step, the workshop leader might consider providing the articulation panelists with information on where the cut scores would be placed if they were smoothed statistically across all grade levels. Because there is more than one way to smooth the cut scores, we recommend first asking school and district-level personnel what pattern they expect to see. Would they expect to see the same percentage of students scoring at proficient or higher at each grade level? Increasing percentages across grades? Decreasing percentages? This information can be used to drive the smoothing. Then, in addition to the information shown in Tables 3 and 4, the smoothed cut score can also be noted in the far left hand column. The articulation panelists will need to understand how the smoothing was done, but then they can take this information into consideration when making their recommendations.

Workshop

Welcome and Introductions (15 minutes)

Once all the panelists have registered and been seated, a representative from the state department of education should welcome them to the standard setting workshop, and introduce the key staff at the workshop. The panelists should then introduce themselves. After the introductions, panelists should be instructed to complete the nondisclosure forms.

Summarize Standard-Setting Procedure (1 hour)

The facilitator should first provide the articulation panelists with a description of the standard-setting method, so articulation panelists understand how the cut scores were determined. As mentioned earlier, we recommend inviting about three participants from each standard-setting panel to help with this summary and to provide their insight of the important issues discussed in determining the standard. An important component of this training is to share the performance level descriptors (PLDs) with the articulation panelists and discuss how the borderline student

was defined. It is important to ensure that the articulation panelists have a complete understanding of the standard-setting process before proceeding.

Present Impact Data (30 minutes)

Next, the articulation panelists should be shown tables of results, including impact data for the recommended cut scores. If they request it, they could also see impact data for the cut scores +/- a range raw score points that could be determined by a pre-set number of points (as shown in table 3), or within 2 SEMs or 2 SEJs (as shown in table 4). Impact data will be provided not only for the state but for each subgroup included in the state report cards (as shown in tables 2, 3, and 4). The impact data tables should be provided for the focal assessments and the comparison assessments. The facilitator should ask the panelists to consider the importance of consistency across subjects. It is often useful to have this discussion regarding expectations prior to seeing any comparative information. That is, ask the articulation panelists if they expect to see the same percentage of students scoring at Proficient in every grade level, decreasing percentages, increasing percentages, or a random pattern. Then, show them the impact data tables, including the smoothed data if it has been calculated, and have them discuss the current trends compared to their expectations.

Discuss Final Cut Scores (45 minutes)

After they have been trained on the methodology and have seen the tables and figures, the articulation panelists should be asked to discuss the results. The discussion ought to center on whether the percentage of students reaching Proficient and Advanced are viable percentages in that state. They should be given a subset of materials used in the standard-setting workshop, which could include ordered-item booklets, test booklets, or sample student work. They should also have the PLDs, and any information about how the panelists described the borderline students. If requested, panelists will have a table similar to Table 3 or 4 that shows the distribution around the recommended range.

Remember that the primary task of the articulation panelists will be to determine the reasonableness of the standards from both a statistical and content perspective and to recommend adopting either the standard-setting cut score(s) or another cut score within the given range. Depending on the methodology used in the standard-setting phase, it may make sense to ask the articulation panelists to look at the content implications for changing the cut score. For example, the facilitator can show them where the new bookmark would be placed or provide a sample of a student's body of work at a particular score. The panelists should be sure that the interpretation of any new cut scores continues to match the PLD for that level.

Make Final Recommendations (15 minutes)

When the articulation panelists have come to a decision on what cut scores they each want to recommend, each panelist will complete a recording sheet providing their final recommendation for the cut scores for each test. In addition, they will be asked to complete an evaluation form, documenting their understanding of the process and their reasoning for recommending any cut score other than the one recommended by standard-setting panelists. Recording their reasoning will provide the workshop leaders with additional validity evidence of the necessity of any changes made to the standard-setting recommendations.

Post Workshop

It is important to understand that even the articulation panel recommendations are just recommendations. The cut scores must be adopted by the state policymakers, usually the State Board of Education or a Commissioner. These state policymakers may choose to override the recommendations and adopt the smoothed cut scores or to make other adjustments. However, conducting an articulation workshop provides additional weight to the standard-setting recommendations as it helps to validate the process by taking normative data into consideration and reviewing the cut scores across the whole assessment program. A standard-setting panel provides the content expertise, while an articulation panel provides the policy expertise. Together, they provide the best recommendations possible from human judgment.

SAMPLE ARTICULATION WORKSHOP AGENDA

Date

9:30 – 9:45 a.m. Welcome and Introductions

9:45 – 10:45 a.m. Summarize Standard-setting Procedure

10:45 – 12:00 noon Present Impact Data and Discuss Cut Scores

12:00 – 12:15 p.m. Make Final Recommendations

12:15 p.m. Lunch

Sample Recording Form for an Articulation Workshop (Date)

How clear was each of the following descriptions or materials?

	Very clear	Somewhat clear	Somewhat unclear	Very unclear
The purpose of this meeting				
My role in this meeting				
The explanation of the procedures used in the earlier standard setting meeting				
That the cut scores resulting from this meeting will be used to inform the State Education Department's final recommendation to the State Board of Education.				
The information in the tables of cut scores by round and by table				
The information in the tables of consequence data				

Please record your best recommendation for the cut scores for each assessment.

Subject	Raw Score Cut Score for Proficient	Raw Score Cut Score for Advanced
Grade 3 Reading		
Grade 4 Reading		
Grade 5 Reading		
Grade 3 Math		
Grade 4 Math		
Grade 5 Math		
Etc...		

For any cut score that you change, what influenced your recommendation?

Name _____ Signature _____