

Planning and Conducting Standard Setting

**Michael B. Bunch
April 17, 2020**

Table of Contents

Overview	5
Part 1: Plan.....	6
Primary goal – Get Approved Cut Scores.....	6
Who?.....	6
What?.....	6
When?.....	7
Where?.....	8
How?	8
Why?	8
Plan Outline.....	8
Purpose/background of this standard setting	8
Basic methodology.....	9
Description of computer programs	9
Timeline	10
On-site agenda.....	10
Round-by-round description of activities	10
Analysis procedures	11
Sample feedback.....	12
Follow-up activities	13
References	14
Appendices.....	14
Performance Level Descriptors.....	14
Articulation of PLDs	15
Special Requirements for Online Standard Setting	16
General.....	16
Cybersecurity	16
Clarity of directions.....	16
Fidelity of process	16
Reliability of software, hardware, and connectivity	18
Clarity of feedback	19
Webinar design and execution	20
Technical support.....	20
Plan Approval	21
Identify the plan approval process	22
Follow the process	24
Final Steps	25
Contingency plans.	25

Part 2: Preparation.....	26
Primary goal – Get Approved Cut Scores.....	26
People	26
Your staff.....	26
Panelists	27
Site staff	31
Others	32
Materials and Equipment	32
PowerPoint presentations	33
Facilitator scripts.....	35
Test materials for panelists to use.....	35
Forms for panelists to use.....	35
Data entry forms and processes	36
Data analysis and output processes	36
Miscellaneous signs, badges, and related materials	37
Logistics.....	38
Travel and lodging.....	38
Shipping.....	40
Paper	40
Online.....	41
Management Activities	41
Plan revisions	41
Communications	41
Part 3: Implementation.....	42
Primary goal – Get Approved Cut Scores.....	42
Preliminaries	42
Opening Day.....	42
Training Sessions.....	44
Presentation of external data	45
Evaluation of training.....	46
Rounds	46
Round 1	46
Between Rounds 1 and 2	47
Round 2	47
Between Rounds 2 and 3	47
Round 3	48
Between Round 3 and vertical articulation	48
Vertical articulation	48

Management Activities	51
Making rounds	51
Daily debriefing.....	52
Care and feeding.....	53
Coordination	53
Part 4: Follow-Up	54
Final Report.....	54
Components of the final report	54
Review and Approval	55
Other Activities	57
Develop or update a score scale.....	57
Revise the final report	57
Forward the final cut scores	57
Review/revise PLDs.....	58
Settle up.....	58
Miscellaneous	58
References	60
Links to Key Standard Setting Reports.....	61
Appendix A: Sample Panelist Selection Letters and Forms	63
Appendix B: Scripts and Work Forms	73
Appendix C: Readiness and Evaluation Forms.....	102
Appendix D: Job Descriptions	110
Appendix E: Additional Files Under Separate Cover.....	117

List of Tables and Figures

Tables

Number	Title	Page
1.1	The Usual Suspects	21
2.1	Sample Panelist Target Matrix for One Subject	29
2.2	Composition of VAC Based on Table 2.1	31
2.3	Paper Products to be Shipped	40
3.1	Vertical Articulation Motions and Actions	50

Figures

Number	Title	Page
1.1	Sample room arrangement for one standard setting panel	11
1.2	Sample feedback from an online body of work standard setting (Round 1)	13
1.3	Sample paper-based bookmark form	17
1.4	Sample item map with one bookmark entered	18
1.5	Sample hardware specifications	19
1.6	Hypothetical application of the Hofstee method to a high school science test	23
2.1	Sample vertical articulation interactive spreadsheet	34
2.2	Sample bookmark tally sheet in Excel	36
2.3	Sample feedback based on tally sheet in Figure 2.2	37
3.1	Room assignment slide from opening plenary session	44
3.2	Display of external data in an item map	46
3.3	Sample VAC seating chart	49

Overview

Standard setting is part science, part art, and part craft. This manual addresses the art and craft parts almost exclusively. If you are reading this manual, you are possibly one of the following three people, with primary emphasis throughout on Person #1:

1. The person who will actually plan and conduct standard setting
2. The person who will contract with and oversee person #1
3. The person who will observe standard setting and make a report to person #2

For person #1, this manual describes what you need to do before, during, and after standard setting. For person #2, this manual tells you what you can and should expect of person #1. For person #3, this manual tells you what you should be looking for as you observe person #1 and prepare your report to person #2.

This manual is presented in four parts:

1. Plan – the official document that describes exactly what will be done in advance of, during, and after standard setting;
2. Preparation – the myriad details to attend to in order to get ready for standard setting;
3. Implementation – step-by-step instructions on how to conduct any standard setting activity, using any method;
4. Follow-Up – the things to be done after standard setting in order to get cut scores approved and implemented.

There are also several appendices and links to external documents:

- Appendix A contains sample letters and forms that may be adapted to a particular standard setting activity.
- Appendix B contains generic scripts for various standard setting activities that may be adapted to a particular standard setting activity.
- Appendix C contains readiness and evaluation forms we have used with great success in recent years.
- Appendix D includes job descriptions of people involved in standard setting
- Appendix E lists several additional resources.
 - *Worksheets* – one Word document and three Excel workbooks that will be useful in preparing a plan and being prepared for almost any contingency
 - *PowerPoint Presentations* – six presentations for a variety of purposes
 - *Other Files* – a combination of papers I have presented at NCME and miscellaneous forms I have developed, and Excel workbooks I have created to expedite analyses and presentation of data

In addition, there are [links](#) to standard setting reports prepared by others as well as to [datasets](#) associated with *Standard Setting: A Guide to Establishing and Evaluating Performance Standards on Tests* (Cizek & Bunch, 2007), located on the Sage.pub website.

Part 1: Plan

This part of the manual addresses the plan for standard setting: drafting it and getting it approved. Planning will be an iterative process that goes right up to the start of standard setting, and sometimes beyond.

Primary goal – Get Approved Cut Scores

The purpose of standard setting, whether in a K-12, certification/ licensure, or other setting, is to obtain cut scores that can be approved and used. Thus, the purpose of the plan – a written document that spells out exactly what you will do and when and how you will do it – is to make sure you reach that goal. [Worksheet 1](#) is designed to help you do that. It addresses these basic questions.

Who? The first part of your job is to find out who will actually make the final decision to approve, disapprove, or modify the cut scores. That is not always as easy as it may sound. In licensure and certification, where the power structure is smaller, it's usually the governing board. But who on the board, exactly, makes that decision? Is it the chair, a committee, or the whole board? In a K-12 setting, particularly for a state or consortium, there are so many different actors with so many different roles, that you will need a playbill to keep them all straight. Before you start drafting a plan, make a list of all the stakeholders and the roles they will play in adopting the cut scores. Each group in this list will have specific responsibilities and will be able to help in specific ways. What do you want each of them to do? How will you secure their help and interact with them?

After the decision makers come the people who will actually be involved in standard setting activities: your staff, panelists, agency staff, site staff, and others. Each will play an important role. Name those people and define their roles.

See the **Who** portion of *Worksheet 1* for details.

What? For a single test with a single cut score, this will be easy. For everything else, it will be a little tougher. Consider a single test with three subtests (e.g., Facilities Manager with subtests for Supervision, Customer Relations, and Financial Management). Is each section to be pass/fail, or will there be other performance levels? How will the subtests be combined to determine who gets certified and who does not? Will a high score on one component make up for a low score on another (compensatory standards) or will the examinee have to pass all three components to be certified (conjunctive standards)? If there are multiple exams (e.g., the Certified Public Accountant exam), must the candidate pass all exams at once, or can individual passing scores be banked? For a typical statewide set of exams in two or more subjects at multiple adjacent grade levels, is there any concern for alignment of cut scores across grades or subjects?

See the **What** portion of *Worksheet 1* for details.

When? While the answer to this question may seem obvious, it usually isn't. The purpose of getting cut scores is to apply them to score reports. Therefore, we have to start with the first date score reports have to be delivered. Work backwards from that date to the date you have to generate score reports. Work backwards from that date to the date you need approval from whoever approves the cut scores in order to complete score reporting. Now look at the schedule of the approver(s). For a state board of education, for example, board meetings are usually set a year in advance, with meeting dates published online. When is the last board meeting before you need to produce score reports? That's the day you will need to present and get approval of the cut scores. Also, is this a one-meeting or two-meeting board (see **How** below)? It makes a difference. Who needs to weigh in after standard setting but before the board decision? How much lead time do they need? All those things take time.

The list goes on. The bottom line is that you need to identify all these people, the roles they play, and when they play them, and construct a timeline that allows all of them to play their roles before you have to report scores. The best time to start working on that timeline is the day you decide to create a test that will be used to separate people into two or more categories; i.e., one that will require cut scores. That is the time to ask:

- When will this test be operationally administered for the first time?
- When will scores for the first operational administration of this test be delivered?
- Will those score reports need to include performance/achievement level designations?
- Who will make decisions about use of the cut scores?
- When will they make those decisions?

Work backwards from there.

Upon arriving at the beginning of your backwards journey, you can start filling in all the logistical elements of the journey forward:

- Meeting site (see **Where**)
- Preparation of materials
- Travel arrangements (self, staff, participants, others)
- Accommodations, meals, security
- Shipping of materials and equipment

See the **When** portion of *Worksheet 1* for details. *Note: Fixing the location and date of standard setting and composition of the pool of panelists needs to happen immediately. Other matters (e.g., methodology, exact agenda) can be negotiated over a longer period of time. In fact, you should settle location, meeting dates, and panelist pool composition before you start writing a plan.*

Where? This is not a trivial concern. Particularly for standard setting activities involving far-flung participants, the site for standard setting is extremely important and needs to be settled at the very outset. Online standard setting is gaining popularity in part because it minimizes the complexity of **Where**. Multi-state, face-to-face standard setting maximizes it. Moreover, the **Where** issue often complicates the **Who**, **When**, and even **How** aspects of standard setting. For example, if you conduct standard setting at a hotel in Las Vegas, some potential panelists simply won't come (or may not be allowed to participate). If the only hotel large enough for your event can't accommodate you before August, you won't have cut scores in time for the July board meeting.

See the **Where** portion of *Worksheet 1* for details.

How? What is the process? What will you actually do? What will you do first, second, last. How will you do it? What specific standard setting procedure will you use? What materials will you use, and how will you use them? How will you train standard setting panelists (and your staff if you will not be doing everything yourself)? How will you get cut scores?

See the **How** portion of *Worksheet 1* for details as well as **Plan Outline** below.

Why? Perhaps this should be the very first question. Why are you conducting standard setting, other than the fact someone told you you had to do it? Is there a legal or administrative mandate? If so, what exactly is it? Does that mandate spell out how many levels there should be, who should be tested and classified, and how results will be used, beyond classification of examinees? For example, will results be used to evaluate teachers or programs? Will they be released to the public? Spend some time with the sponsors of the test and other key stakeholders to find out what they hope to accomplish by conducting standard setting.

Plan Outline

Now you are ready to draft a standard setting plan. I have found the outline below to be very helpful. The details will vary from plan to plan, but these basic elements seem to be in every plan I've ever written.

Purpose/background of this standard setting. Why are you conducting standard setting? For statewide or districtwide assessments, there is usually guiding language in legislation or board resolution specifying why you need to set cut scores. Review that language and incorporate it either *verbatim* or by reference in the introduction to your plan. This section should clearly spell out the population affected, the number and nature of performance levels, and the history of actions leading up to the decision to have this test for these subjects for these grades with these performance levels.

For certification and licensure testing, purpose can be fairly easily addressed by reference to a board decision. As with K-12 assessment, quote or make specific reference to that decision in the introduction to the plan.

Basic methodology. Once you have explained why you are doing standard setting, it is time to describe how; i.e., the procedures you will use, the tests you will focus on, the number and characteristics of the panelists who will be involved, the people who will lead the activity, and the statistics you will use. This will need to be the most detailed section of the plan. Make sure you describe your methodology exactly as you plan to implement it, and then implement it exactly as you described it. By that, I mean exactly as the final plan – reviewed, modified, and accepted by external reviewers and clients – described it.

This is where you describe in detail the specific standard setting procedure you will follow – Angoff (and its plethora of modifications), benchmark, Beuk, body of work, bookmark, briefing book (now called evidence-centered), Hofstee, item descriptor match, or some other procedure. Keep in mind that the plan will be reviewed and ultimately approved by a variety of individuals, some of whom may not be as knowledgeable about standard setting as you would hope. Tell why you think the procedure you have chosen is the most appropriate for this type of test and the ultimate uses of the cut scores you will derive. Provide a little background on the procedure, and work through any particularly tricky aspects of its psychometrics. It is not always necessary to derive every equation or spell out every analysis, particularly for the better-known procedures.

Describe in detail what task panelists will complete and how their work will be transformed into cut scores. How many rounds of standard setting will there be? How will each round differ from the one just before it? What will happen between rounds? How will you verify the accuracy of the cut scores you derive from the information panelists provide? Will there be vertical articulation?

Be sure to describe decision rules (e.g., mean vs. median, RP50 vs. RP67, rounding up or down) and their rationales. For example, if you propose to use RP67 rather than some other response probability, explain why you chose that RP value and perhaps even why you rejected other values. For rounding, tell when you will round; i.e., at the panelist level, the table level, the roomwide level, or some other level. An average of averages of numbers rounded more than once may not be everyone's idea of a precise cut score.

Description of computer programs. It is one thing to describe the statistics you plan to use; it is another thing entirely to specify how you will compute those statistics. Different programs yield slightly different results. If you are using a particular commercial program, specify the version and release date. If you are using proprietary software, specify exactly how and for which analyses you will use it. Be prepared to demonstrate your software prior to standard setting. Prepare dummy data to run during the demonstration.

Your plan should show sample output (textual, tabular, and graphical) of any software you plan to use and describe how you will use that output during standard setting. If there are handoffs from one software package to another, describe those handoffs and the safeguards and quality assurance procedures you will have in place during standard setting to make sure the results you report to panelists during standard setting and stakeholders after standard setting are accurate.

Timeline. Between the time you present the plan and the time you actually conduct standard setting, several events must take place. You have panelists to recruit, materials to prepare, travel and meeting site arrangements to make, and a host of other tasks to complete. When will you complete them, and what are the dependencies among the various tasks? As noted above, a backward approach works well when there are dependencies. Start with the first day of standard setting. What do you need to do to get yourself and everyone else there? Now what do you need to do to make those things happen? And so on back to today's date. Make sure you include necessary follow-up activities.

A critical event, often overlooked in preparation of timelines, is the time needed for approval of the plan. Make sure you include that. If you don't know, talk to the one person ultimately in charge of plan approval to get not just an estimate but a commitment to a deadline for approving the plan. Delays in the approval process can throw everything off. Make it a high priority to get a commitment on plan approval.

On-site agenda. The on-site agenda should be a fairly detailed table showing the key activities of each day, with start and stop times and a brief description of the activity. Because some activities will be for the entire group and some will be for smaller groups, the agenda should clearly indicate which activities are of which type. A description of the details of the various activities should immediately follow the agenda.

Round-by-round description of activities. Each standard setting will probably start with an opening plenary session. What are the objectives of this activity? Who will speak (not necessarily the name but the title or titles, such as Assessment Director, Chief Facilitator, etc.)?

Shortly after the opening plenary session, there will be a series of training sessions. What will be the focus of each one? Who will lead those sessions, and what information will they provide? Describe each training session in enough detail for a reviewer to get a clear understanding of what you hope to accomplish and how you plan to accomplish it. Later in the process (not with the submission of the original plan for approval), you will need to provide much more detailed descriptions of the training sessions, including scripts, PowerPoint presentations, and panelist materials for review and approval, so keep that deliverable in mind as you prepare the plan.

After describing all training sessions, describe what will happen in Round 1 – how panelists will be grouped, where they will go, what their objective(s) will be, what kinds of tasks they will complete, what kinds of instructions they will receive, and how their activities will be

monitored. It may even be helpful to include diagrams of room arrangements such as shown in Figure 1.1.

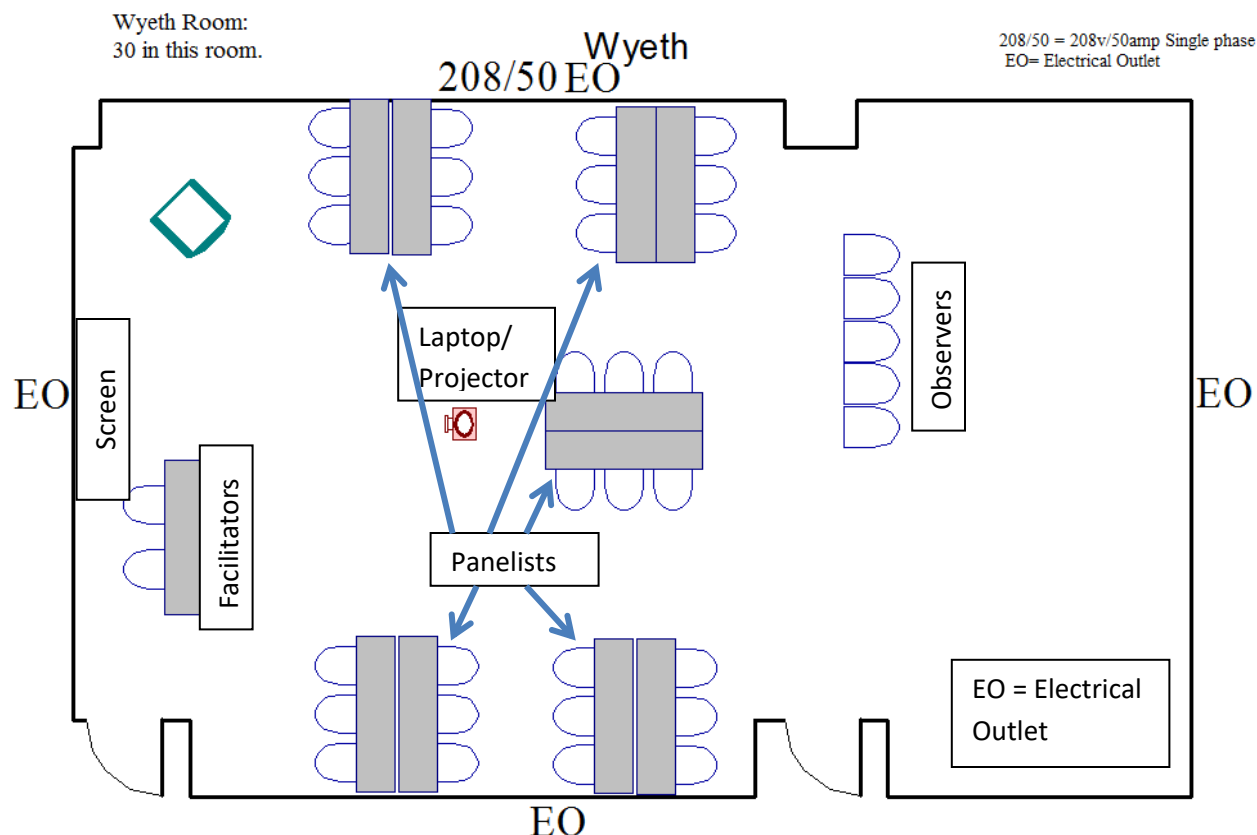


Figure 1.1. Sample room arrangement for one standard setting panel

Describe how you will end Round 1, what you will collect from panelists, and what kinds of instructions you will give to panelists as they finish. Describe how you will collect their Round 1 responses and how you will analyze them.

Repeat this process for each round. For the final round, tell how you will conclude the overall activity, whether there will be any follow-up, and what kind of information you will share with panelists.

Analysis procedures. Describe in detail how cut scores will be derived. You have already described the procedure and computer programs you will use to complete all analyses. Now you can lay out each analysis procedure step by step: How information flows from panelist to you, how you aggregate the information for a group of panelists, how you transform that information into cut scores and other feedback.

With the advent of college and career readiness for academic assessments, evidence that the cut scores derived from standard setting are indicative of future success has become an important part of the process. For high school tests, evidence of future success may include

scores on college entrance exams or other tests of content beyond that on the test for which cut scores are being set. For elementary and middle school tests, evidence of future success may include scores on high school exams. The standard setting plan should indicate what external evidence, if any, will be included in the process and how it will be used. Suggestions as to the type of evidence and how it might be used are given below.

- *Comparing impact data* – Results for the National Assessment of Educational Progress (NAEP) have historically been less flattering than most state assessment results; percentages of students at or above Proficient, for example, are typically lower than on state exams. As a reality check on a new test that is supposedly more rigorous than the one it replaces, consider comparing fourth and eighth grade impact for NAEP with penultimate-round impact for the new test or with impact from the previous test. Discuss in advance with the sponsoring agency what their expectations are in this regard.
- *Benchmarking* – Gary Phillips (2012) has described a method of statistically linking the test for which cut scores are to be derived with other tests such as NAEP, PIRLS (the Progress in International Reading Literacy Study), or TIMSS (Trends in International Mathematics and Science Study). Establishing the linkages can be fairly complicated and is open to criticism. Before assuming this is the way to go, discuss the matter thoroughly with the sponsoring agency, and work out a detailed plan for obtaining and using the necessary benchmarking data.
- *Embedding items* – For one test, we secured permission to embed NAEP items in the fourth and eighth grade field tests, and we scaled those items to the new test scale. The NAEP items chosen had scale values at or near the NAEP Proficient level. By rescaling those items to the new test, we were able to determine the point on the new test scale that would approximate Proficient on the NAEP scale. Embedding requires close cooperation with the owners of other tests and timely receipt of the necessary data.
- *Cut score validation* – Instead of comparing scores on the tests for which cut scores are being derived to those on other tests in the current year, consider tracking test takers over two or three years (e.g., high school to college, fourth grade to sixth grade), and comparing performance in the current year to performance in subsequent years. How many of the students who scored at the Proficient level or above in fourth grade are doing Proficient level work in sixth grade? If the answer is encouraging, the cut score for Proficient was probably on target. If not, perhaps it needs to be modified.

Sample feedback. It is important to describe what you will tell panelists after each round to prepare them for the next round. Sometimes it is helpful to show dummy output in the plan to give a clearer indication of what you plan to share with panelists. However, it is equally important to explain why you plan to present feedback in this manner and tell how you will present it and encourage panelists to use it. Figure 1.2 shows fabricated feedback for Round 1 of an online modified body of work standard setting activity.

Grade 7-8 Narrative

Grade	Median		Min		Max		Range	
	Level 2	Level 3	Level 2	Level 3	Level 2	Level 3	Level 2	Level 3
7	25.0	28.0	23	28	26	31	3	3
8	27.0	31.0	25	29	28	32	3	3

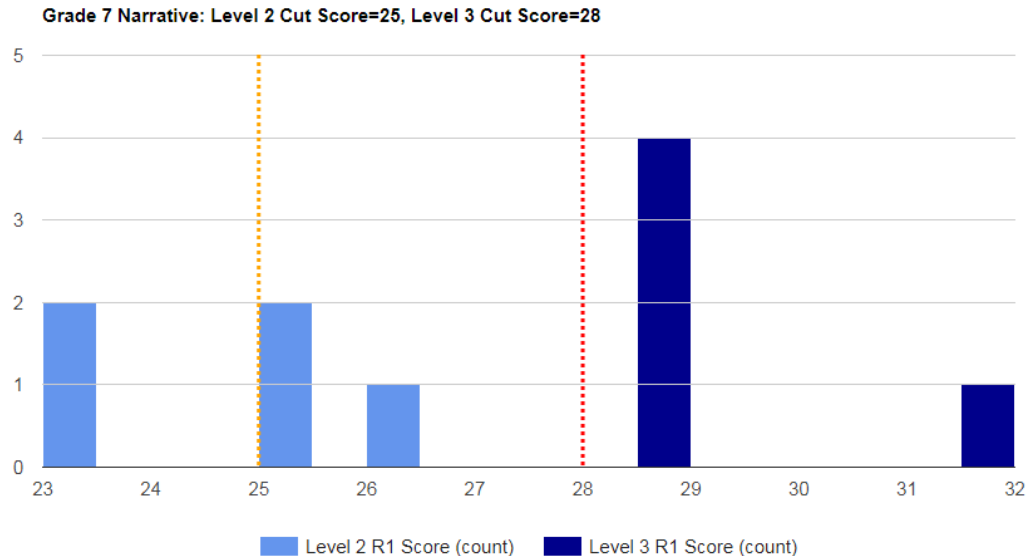


Figure 1.2. Sample feedback from a modified online body of work standard setting (Round 1)

The upper portion of Figure 1.2 presents Round 1 feedback in tabular form. This report for the grade 7-8 panel provides the median cut score as well as the minimum and maximum cut scores entered by panelists. Panelists also receive their own computed cut scores for each level. The purpose of this table is to allow panelists to see where their individual cut scores fell in the overall range of cuts. This information is meant to stimulate a discussion of why different panelists set their cut scores at different places. The bottom portion of Figure 1.2 provides this distribution information graphically so that each panelist can see not only where his or her cuts were, relative to the group median, but also how many others agreed with them.

It is important to note that tables and graphs like these are essentially the same whether the activity is done face to face or online. For online presentation of this information, however, it is crucial that the facilitator present this information and ask pointed questions of individual panelists to make sure they understand. In a face-to-face setting, the facilitator would have the advantage of cues from panelists' body language.

Follow-up activities. Will there be vertical articulation? If so, how will you conduct it? What will you do between the time standard setting ends and cut scores are approved. Will you write a

formal report, present results to the sponsoring agency, help agency staff prepare for that presentation? Spell out in some detail what you plan to do.

References. There will be two major sources of references: test program documentation and methodology documentation. For the introductory section of the plan, it is often helpful to provide reference to any document that describes the testing program, any relevant legislation or board resolution that establishes the testing program, and any existing technical report or manual pertaining to the testing program. For the methodology section, cite one or two good descriptions of the procedure you will use plus other sources supporting each decision rule (e.g., rounding, RP values, etc.). If citing a textbook in the methodology section, be sure to cite the specific chapter (e.g., Cizek & Bunch, 2007, Chapter 10: The bookmark method).

Appendices. In the initial plan, submitted months prior to standard setting, it is unlikely that you will have final versions of training materials or data output. Those will come in the final version of the plan a few weeks before standard setting. You should, however, be able to include a few sample forms, such as the readiness and evaluation forms, you intend to use. Like all other aspects of the plan, these forms will be subject to review and modification by the client and external reviewers. It is a good idea to get these in front of reviewers as early as possible.

Performance Level Descriptors

For decades, we conducted standard setting without performance level descriptors (PLDs). These are sometimes called Achievement Level Descriptors (ALDs). [I will use PLD throughout this manual, but everything here applies equally to ALDs.] William Angoff (1971), for example, simply directed us to consider the “minimally acceptable person,” and a generation of standard setters spent a good part of their standard setting time defining that person. PLDs came into play in the 1990s as we moved from Angoff methods to more sophisticated methods and multiple cut scores to define multiple performance levels.

There are four distinct types of PLD, each for a specific purpose: policy, range, target, and reporting. Different standard setting methodologies require different kinds of target PLDs. For example, traditional body of work requires target PLDs that describe the levels in global terms. Bookmark and modified body of work require target PLDs that define the threshold of each level. The item descriptor match (IDM) method requires panelists to match each item to a specific target PLD, which could be either a global or threshold description. The key feature of a PLD used for the IDM method is its specificity. Those instances in which the IDM method is used frequently employ PLDs created specifically with this method in mind.

Whichever standard setting method you select, you will need to make sure there are PLDs that will support your plan. For new testing programs, there may only be general policy PLDs (basically, the label and a very brief description). If that is the case, your first task will be to work with the sponsoring agency to craft the PLDs you will use in standard setting. That task

may include the creation of range PLDs to guide item development, but range PLDs are beyond the scope of this manual.

If target PLDs have been developed, make sure you are thoroughly familiar with them and that the facilitators who will be responsible for one or more tests are also familiar with the ones they will be using. Schedule a meeting with the sponsoring agency's content specialists to review the PLDs and develop a clear understanding of their interpretation of the PLDs. Ultimately, it will be the sponsoring agency's interpretation of the PLDs you will have to get across to panelists.

If PLDs have not been developed specifically for standard setting, include a PLD drafting and review process in your overall plan. In some instances, the sponsoring agency will develop the PLDs; in other instances, the agency will expect you to develop them. Find out which approach you need to take, and plan accordingly.

If you are responsible for creating the PLDs, schedule a one-day workshop with content matter experts for that purpose. In advance of that workshop, work with the sponsoring agency's content experts to create frameworks, starting with the policy PLDs and, if they exist, range PLDs. Create shells that you will complete at the one-day workshop. Do not assume that a group of content experts will be able to create a whole set of PLDs from scratch in one day. Work you will do in advance of that meeting, with the sponsoring agency's and your own content experts will not only move the process along more efficiently, it will guarantee a final product that can be approved by the sponsoring agency.

Since PLDs define performance levels within the framework of the test rather than in general terms, it is very important that anyone working on the PLDs be very familiar with the test. It is not necessary that PLD developers see every test item that might appear on a test. It is sufficient that they become familiar with the test blueprint and several sample items for each cell. Of course, if they have completed all the necessary security agreements and will need to see all test items at some other point in the process, then the content experts should definitely have access to all items.

Ideally, you should aim to have a good draft of the PLDs a month or so prior to standard setting. That will give you time to shepherd them through the internal and external review process and make final refinements well in advance of standard setting. Once you have obtained final approval of the PLDs you will use in standard setting, make sure your facilitators understand them well enough to explain them to panelists. That includes going over fine points that were modified during the review process.

Articulation of PLDs. For K-12 assessments, the PLDs will likely be used for standard setting for multiple contiguous grades with vertical articulation. Before the PLDs are submitted for final approval, a cross-grade committee of content experts should review the full set to make sure they are vertically aligned or articulated. Does the language for grade 7 Proficient fit nicely between the language for grade 6 Proficient and Grade 8 Proficient? Is there a steady increase in expectations from grade to grade for a given performance level? Do the PLDs follow the learning progressions? If the answer to any of these questions is No, go back over the PLDs and

adjust language at one or more grades for one or more levels until there is a consistent upward progression from grade to grade for each performance level. For more detail on this matter, consult Egan, Schneider, & Ferarra ([2012, p. 93](#)).

Special Requirements for Online Standard Setting

General. In terms of planning and preparing for standard setting, there is not much difference between face-to-face and online events. The main difference is that nearly everything that would have been printed will be uploaded into a program that links and presents tests, rating forms, item maps, and support documents on a computer screen or other computing device.

Note: Find out what kinds of devices examinees use to take the test. Don't use anything with a screen smaller than a standard iPad's (no iPad minis or smart phones).

Cybersecurity. With paper-based operations, you only have to keep secure materials under lock and key, sign them in and out by security code, and count everything at the end of every session. Online standard setting presents a new layer of security concerns. Presentation of standard setting materials, particularly test items and examinee responses, must be as secure as they were during operational test administration. The same cybersecurity protocols should be observed. No one should be able to gain access to the standard setting program or materials. If you are conducting online standard setting, the plan should contain a detailed explanation of how you will guarantee that. The same applies if you will be meeting face to face but using software to present materials to panelists. Those applications are further broken down into sessions with a completely closed, self-contained system (as, for example NAEP standard setting with BOWTIE was conducted several years ago) and open, web-based activities. In both instances, you will have both on-site and online security issues to address. Spell out in the plan how you will address them.

Clarity of directions. If the sessions are to be truly remote, clarity of directions becomes more of an issue than in a face-to-face session where the facilitator can repeat, clarify, or restart explanations. No matter how clear you think your directions are, have several people use them without any further help from you, and see how well they follow them. Then ask them to explain to you what they did and why. You may find some gaps in your directions. Find out what was confusing or incomplete, fix it, and try again until you get the results you need.

Fidelity of process. Until online standard setting becomes the rule rather than the exception, it is likely to be necessary to demonstrate that the results you get would not have been appreciably different if you had conducted standard setting face to face with paper. Obviously, a test item on screen will not look exactly the same as it does in a paper test booklet. There is nothing you can do about that. If the test is administered online, try to make the test booklet you present at standard setting look as much like the operational test as possible.

The crucial aspects, however, are the forms panelists will use to render their judgments, and the key is level of complexity. Is the task presented to the panelist in an online standard setting activity equivalent to that of the task presented in a paper-based standard setting? The only

way to know for sure is to field test the online setup. Let's consider a bookmark procedure with a 72-page ordered item booklet and a bookmark form similar to the one shown in Figure 1.3.

Standard Setting – Bookmark Form

Grade (Circle one) 3 4 5 6 7 8 Panelist Number _____

Bookmarks (Enter Page Number.)

Round	Level 2	Level 3	Level 4
1			
2			
3			

Figure 1.3. Sample paper-based bookmark form

For this activity, panelists would work through an ordered item booklet (OIB) and enter three page numbers in Round 1, three more in Round 2, and three more in Round 3. In Rounds 2 and 3, panelists would see what they had entered in the previous round and have the option of repeating that page number or entering a new one for each cut. The cognitive task is to apply PLDs to each item in the OIB and make three decisions – one for the threshold for Level 2, one for Level 3, and one for Level 4.

In the equivalent activity online, panelists would examine the same OIB (although presented in a different medium), apply the PLDs (which will look identical to the paper-based ones, at least as far as computer screen technology is able to make them look identical), and enter three marks on an item map that might look something like Figure 1.4.

Online Standard Setting

Welcome [Sample User!](#)

Home Instructions **Item Map**

Round: 2 Math / Grade: 6 Round 1 Round 2 Submitted: 9/1/2014 10:00 AM

Page	RP67	Scaled Score	CCSS	DOK	Reviewed	Comment	Data Field 1	Data Field 2	Cut
1	-.477	213	MA.6.EE.7.1	1	<input checked="" type="checkbox"/>				
2	-.471	221	MA.7.NS.3.1	2	<input checked="" type="checkbox"/>				
3	-.463	229	MA.6.EE.7.2	1	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>			
4	-.452	235	MA.6.EE.6.1	1	<input checked="" type="checkbox"/>				
5	-.438	241					View		2
6	-.420	257	MA.6.EE.6.1	2					
7	-.416	268	MA.6.EE.6.1	1					
8	-.404	277	MA.6.EE.6.1	1					
9	-.392	283					View		

Submit

Figure 1.4. Sample item map with one bookmark entered

Here, a panelist has gotten to page 5 of the online OIB and decided that would be a good place to mark the beginning of Level 2. This panelist examined the same PLDs and answered the same standard setting questions as the panelist completing the paper bookmark form shown in Figure 1.3. For the paper form (Figure 1.3), a facilitator would collect the forms, deliver them to a data entry person who would enter them into the data analysis program, and the data analyst would process the data to produce the Round 1 cut score. In the online application, the page number indicated by the number **2** in figure 1.4 would be automatically recorded as a bookmark for Level 2 and combined with all the other entries for this level, to produce an aggregate cut score for Level 2 for that panel. At each stage of the process, the differences between paper-based and online applications are minimized or eliminated.

Reliability of software, hardware, and connectivity. Word to the wise: Don't launch a system you haven't thoroughly tested. That includes load testing, although it's not likely that you will have hundreds of thousands of people hitting your system at once. Still, it is good to know what happens when 40 or 50 panelists hit your system at once. The key test is whether or not the software faithfully accepts panelist input, transforms it into accurate cut scores and feedback, and presents it in a way that makes sense.

What about hardware? For an on-site, online meeting, you have control over the hardware, particularly if you provide it for everyone. Asking everyone to bring their own hardware to an on-site, online meeting is an invitation to disaster, so **don't do it**. It is more expense and trouble to provide all the hardware yourself, but it is well worth it. Figure 1.5 shows the specifications for Chromebooks we recently used for an on-site, online standard setting. This level of specificity is becoming a requirement. Even if it is not, it's best to know what you have and that it works before you go live.

Screen Size	14 inches
Screen Resolution	1366x768
Processor	1.6 GHz Intel Celeron D
RAM	16 GB SDRAM DDR3
Memory Speed	1600 MHz
Hard Drive	16 GB 16GB eMMC flash memory on board
Graphics Coprocessor	Intel HD Graphics 400
Chipset Brand	Intel
Card Description	Intel-HD400/IGP
Wireless Type	Bluetooth
Number of USB 2.0 Ports	3
Average Battery Life (in hours)	10 hours
Operating System	Chrome OS
Weight	3.31 pounds
Dimensions	13.1 x 9.1 x 0.9 inches
Processor Brand (Count)	Intel (2)
Computer Memory Type	DDR3 SDRAM
Flash Memory Size	16 GB
Hard Drive Interface	Solid State
Power Source	AC & Battery
Batteries	1 Lithium Polymer battery

Figure 1.5. Sample hardware specifications

If you are conducting standard setting completely remotely, make sure to warn online panelists about computing devices you know your operating system does not support. Get as many devices as you can, and try running the software on them. If it runs some of the time but not all of the time on a particular device, ban that device specifically in instructions to online panelists. Better yet, if you know of three or four that work well every time, recommend them in introductory instructions. Invite online panelists to contact you about any concerns they may have about their devices' compatibility with your program before going live.

Don't overlook connectivity and bandwidth, even if your total group size is small. Don't take a hotel banquet manager's word for them either. Get in writing a description of the connectivity and bandwidth you will have at your meeting location if you are conducting on-site, online standard setting. Better yet, visit the site ahead of time and test the connectivity. For totally remote standard setting, there's not much you can do except to make sure that any problems that may arise don't arise from your end; i.e., make sure you have enough servers in operation and that they communicate with one another and channel users properly. And make sure you have adequate technical support throughout the standard setting period.

Clarity of feedback. When you field test directions for clarity, do the same for inter-round feedback. So far, we have done online standard setting only when we could schedule webinars between rounds to present feedback from the previous round and give instructions for the next round. However, when we conducted standard setting for Smarter Balanced in 2014, a total of

2,660 people logged in and entered a single cut score for one test. Since those online panelists did not enter a Round 2 cut score, there was no need to give them feedback on Round 1.

Webinar design and execution. Design inter-round webinars just as you would inter-round discussions for face-to-face standard setting, and observe these basic rules:

1. *Keep panel size fairly small.* A dozen or so panelists is just about ideal for an inter-round discussion. The facilitator should present and explain the feedback and then open the floor for discussion.
2. *Control the discussion.* Even with a small group, if everyone has an open microphone, the webinar can spiral downward very quickly. Most webinar software has a chat function. That works very well for straightforward questions. The facilitator presents, an assistant monitors the chat function and prompts the facilitator, and the facilitator answers the question for the whole panel. During this time, all panelists should have their devices set on mute. The facilitator can mute all or mute and unmute individual panelists who raise their hands (another function of most webinar software) and allow them to speak. If you have several panels, assign a facilitator and assistant to each one and provide thorough training and a [script](#) to guide them through the feedback sessions. With limited staff, it may be possible to stagger the inter-round discussions so that one facilitator can lead two or more groups.
3. *Provide advance instructions to panelists.* Everyone should know how and when to log in and present their credentials to enter the webinar on time. They should also know some basic things they should and should not do. For example, no one should try to participate in your webinar from a coffee shop, restaurant, or other public place, and a moving car is entirely out of the question. Children and pets should be in another part of the house, and the phone should be muted. Other basic webinar etiquette applies. However, for many of your panelists, this may be their very first webinar, and some of them may not even be aware that their laptops have webcams. Unless they want the rest of the panel to see them in their bathrobes, panelists should be advised of this possibility and encouraged to dress accordingly. Explain all rules in an advance communication, and briefly reiterate them at the beginning of the webinar.

Technical support. If you are conducting standard setting, chances are you are not a member of your organization's information technology technical support team. If you do not have one, go find one. Technical support will make sure the software works every time, on time. They will make sure everyone can log in and will fix whatever problems anyone has if they can't log in. They will fix a host of problems you won't even think about.

If you are conducting on-site computer-based standard setting, take one or more technical support staff with you. They will verify the site's connectivity and bandwidth and solve a host of problems. For example, at a recent on-site online event, we had reasonable connectivity and bandwidth, but in two of the breakout rooms some of the panelists were getting very slow response times. The technical support staff went in and discovered dead spots in the room. Once we moved the panelists to different parts of the room, everything was fine. Would you

have thought of that? Take tech support with you. When you are verifying the bandwidth and connectivity of the site and want to go to verify them, take tech support with you.

Plan Approval

As noted in the previous section, plan approval is likely to be an iterative process involving several people. Find out exactly who those people are before you draft the plan and develop a strategy for presenting the plan to each one. Following these steps should help.

Identify the plan approvers. Use *Worksheet 1* to identify everyone who will be involved in the approval process. Table 1.1 lists the usual suspects. The final category (External Reviewers) is a rather broad one and could include almost anyone, from parents and teachers (K-12 Setting) to current practitioners (Certification & Licensure Setting). Table 1.1 is merely a representative sample; there will likely be more. Track them all down, and find out what they will want to know about the standard setting activity and uses of the cut scores.

Table 1.1
The Usual Suspects

K-12 Setting	Certification & Licensure Setting
State or Local Board of Education	Board of Directors/Trustees
Superintendent/Commissioner	Executive Director
Assistant Superintendent/Commissioner	Assistant Director
Assessment Director	Director of Testing/Psychometrics
Assessment Department Staff	Psychometrician(s)
Technical Advisory Committee	External Reviewers
External Reviewers	

Note that you may not have access to all of them. For example, if you are working with a state department of education, you probably won't have direct access to the state superintendent or commissioner. In that case, you need to know who will have access. Present your case directly to that person, and make yourself available for a presentation to the ultimate approver(s) if that is possible. Where that access is absolutely denied, make your case to the intermediary clear and complete. If you have followed all the steps in the first section of this manual, you should be in good shape in this regard.

Sometimes, there will be one or more external observers looking over your shoulder during standard setting and making a report to the sponsoring agency. Find out if that is going to be the case and who that person or those persons will be. Make sure they get a copy of the original plan and all subsequent revisions and are completely conversant with the final plan prior to standard setting. In all fairness, the role of on-site reviewers is to make sure you follow the plan, not to nitpick your conduct of standard setting because they would have done it

differently. You really don't want to be in the middle of a bookmark standard setting and have the external observer ask you why you aren't using modified Angoff. Of course, that person is welcome to make comments during the plan review process, but when the final plan is approved, everyone is bound by it, including observers.

Identify the plan approval process. In a typical approval process, you will start with one person or a small group. This person or group will review your initial plan and make suggestions for changes. Rather than blindly make the changes, schedule a conversation to discuss them. This will be a time to clear up any confusion about what you meant or what they meant. Once you and they have a clear understanding of what the plan needs to address and how it needs to address it, revise and resubmit the plan.

The next step is usually review by an external party or parties. For K-12 plans, the external parties are usually members of a technical advisory committee (TAC). A certification and licensure group may or may not have a TAC but will likely have one or more consultants who will review the plan. In any case, find out who these people are and what their particular interests in standard setting are.

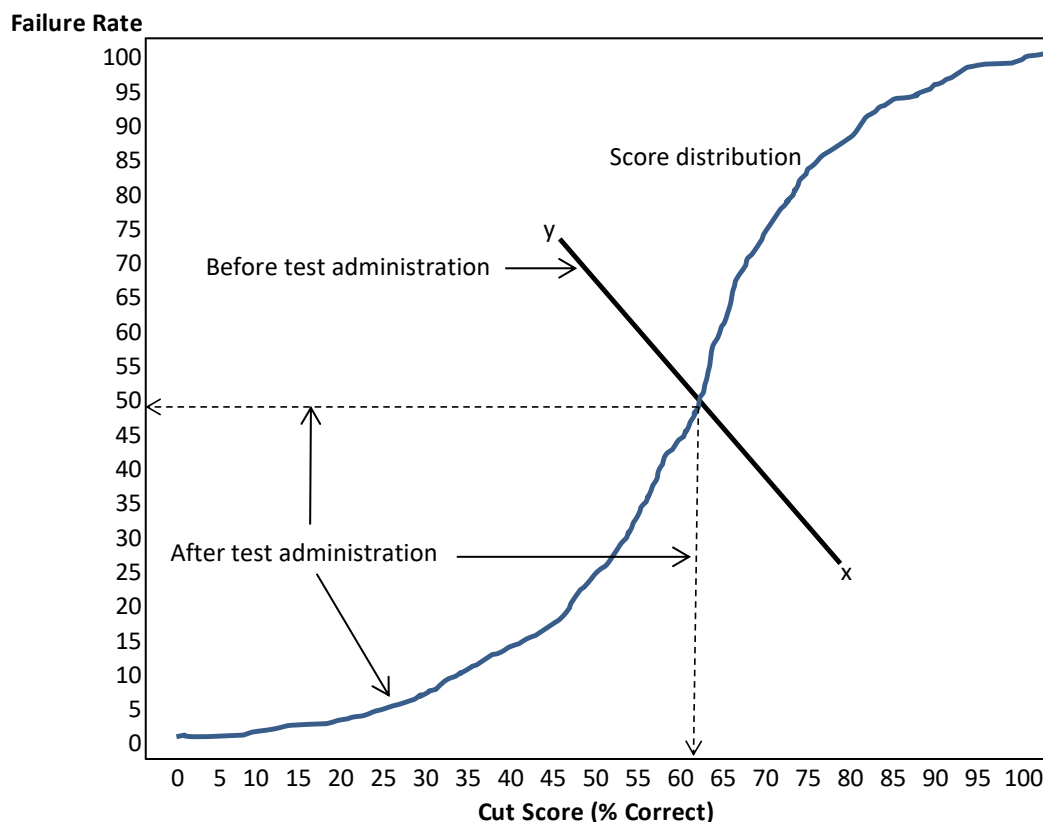
It is entirely conceivable that the plan will not ever reach the body responsible for approval of the cut scores. Find out early if that is to be the case. If it is, that body has delegated responsibility for approval of the plan to someone else, typically the director of assessment or executive director. Find out who that person is and exactly what his or her authority is. Also find out if the governing body has provided any insights into their desires for conduct of the standard setting activity or even its outcomes. Find out as much as you can about the desires and wishes of those who have ultimate authority to approve cut scores, and make sure you address those concerns in the plan.

Even if the governing body delegates responsibility for approving the plan to a staff member, they will surely be interested in the outcome. One way to engage them is to conduct a mock Hofstee (1983) or Beuk (1984) standard setting activity with them. For example, consider a high school science test with four levels, Level 3 being the one designated as Proficient for federal reporting purposes. Ask the assessment director (or whoever has been given responsibility by the board for approving the plan) to set up a half-day meeting with the whole board or a committee of that board. Allow them to review the test (or even take it and score it themselves, which is always enlightening). Then ask them these four questions:

1. What is the highest % correct cut score that would be acceptable?
2. What is the lowest % correct cut score that would be acceptable?
3. What is the maximum acceptable failure rate?
4. What is the minimum acceptable failure rate?

Questions 1 and 3 address concerns board members may have about defending the test in light of high failure rates. Questions 2 and 4 address just the opposite concerns: a test so easy that nearly everyone passes (i.e., reaches Level 3 or higher) and is therefore not rigorous enough to pass peer review or public scrutiny. That actually happens.

Figure 1.6 shows the results of a hypothetical application of the Hofstee method for the high school science test described above. The four percentages at the bottom of the graph indicate the means of the board's responses to questions 1-4 above. Point x indicates the intersection of the highest acceptable cut score and minimum failure rate. Point y indicates the intersection of the lowest acceptable cut score and maximum failure rate. The diagonal line from y to x shows all possible acceptable cut scores and failure rates. As long as the failure rate is somewhere along this line (between 25 and 75% and the cut score is somewhere between 45 and 80% correct), this board (on average) will be satisfied. The point where the eventual score distribution crosses this line will be the cut score and will reveal the failure rate.



- Question 1 Highest acceptable cut score: 80%
- Question 2 Lowest acceptable cut score: 45%
- Question 3 Highest acceptable failure rate: 75%
- Question 4 Lowest acceptable failure rate: 25%

Figure 1.6. Hypothetical application of the Hofstee method to a high school science test

After tests are administered, you can superimpose the graph of the score distribution (i.e., % scoring below each score point) on the decision line, as shown in Figure 1.6. In this example, the score distribution crosses the previously identified line of acceptability at (61,49); i.e., a Level 3 cut score of 61% correct, with 49% of students failing to reach that level. Both numbers fall within the previously identified range of acceptability, a fact you will no doubt want to point out to the board when you present the results.

Keep in mind that the test has not been administered yet, so standard setting has not taken place. However, we are now armed with an understanding of the acceptable parameters for the cut score and failure rates. We focus on just one cut score, the one most people will be concentrating on when the first score reports are delivered. Since this cut is for Level 3, we know that the cut for Level 2 has to be lower and the cut for Level 4 has to be higher. How much lower or higher can be worked out later.

If raw scores are not obtainable (e.g., in a computer adaptive environment), you will need to take a slightly different approach. Present to the board a fixed form of the test that conforms to the blueprint in all aspects. Construct a raw score to theta or scale score table for the fixed-form test. Get the board's answers to the four questions in terms of percent of total possible points on the fixed-form test, and convert that percentage to a theta or scale value. Then proceed as if you had started with percent correct scores.

Follow the process. That may sound simpleminded, but remember that the process will evolve. After you present the draft plan for the first round of review, start a log of all comments and changes. When you get feedback and start to make changes, save the first draft. There are two ways to move forward:

1. A new, clean copy with a summary of major changes to the original appended or included as a separate document;
2. A marked-up copy of the original, showing all changes and comments in line.

I prefer method 1 for the simple reason that method 2 becomes very difficult to follow after a couple of iterations with mark-up and comments in multiple colors, comments on comments, and corrections of corrections.

As you move through the process, note who suggests changes, what they suggest, and why. When you receive conflicting recommendations (e.g., from an agency representative and an external reviewer), take them to the highest authority to whom you have access, point out the conflict, and recommend a solution. Don't count on the person in charge to come up with a solution. Recommend one and be prepared to defend it with your expertise and airtight logic. If that approach doesn't work, present the conflicting recommendations to the people who made them, point out the conflict, and work it out.

Because the approval process may go on for months, it is helpful to keep track of changes and review them as new changes are made. There will be times when a change made early in the process gets reversed at a later stage. It is helpful to ask why the change is needed, as the answer often reveals changes that have occurred in the larger testing process or beyond. Standard setting may be your primary focus, but it is crucial that you keep up with everything that is going on in the total testing program as well as within the organization sponsoring the test. Circumstances change; be prepared to change with them.

As noted in the first section, set a date for final approval of the plan, and get a written commitment to that date from the sponsoring agency. Hold to that date. Make sure all concerned parties have a clean copy of the final, approved plan.

Final Steps

You now have an approved plan that all key players understand, so it is time to make sure other people understand the plan and their roles in implementing it. Primarily, that means your staff – the facilitators and support staff who will help you lead standard setting. Inasmuch as some of the details of the plan will be in flux right up to the last minute, concentrate on those fundamental elements of the plan that should be well established a month or so before standard setting, and familiarize your staff with them.

Contingency plans. As Robert Burns famously wrote:

The best laid schemes o' Mice an' Men,
Gang aft agley.
An' lea'e us nought but grief an' pain,
For promis'd joy!

There will be enough grief an' pain involved with standard setting that you won't need your plans going agley (*badly* to those of you not familiar with 18th Century Scots dialect). Panelists won't show up. People will get sick. Electricity or internet will go out. There will be another event across or down the corridor that conflicts with your access to services or that makes too much noise. The list goes on and on. A major part of your job will be to anticipate all the things that could go wrong *and that you can fix or prevent*. Even for things you can't prevent (e.g., fires or floods or a pandemic), you can have a reasonable response. For example, the Smarter Balanced standard setting plan had a 33-page contingency plan just for the communication component. There was a reported case of Ebola in Dallas the week before standard setting, and we were prepared – we made hundreds of phone calls that week, just as we had planned to do in case something like that (though not specifically an Ebola outbreak) happened.

[Worksheet 2](#) should help you identify things that could go wrong that you can fix or be prepared for. Keep it handy. Add to it as you find new things that go bump in the night (or day).

Part 2: Preparation

Now that you have a plan, it is time to begin preparing for the actual standard setting event: people, materials and equipment, and logistics. This process starts shortly after the planning process begins and ends just seconds before the opening session on the first day of standard setting.

Primary goal – Get Approved Cut Scores

Yes, that’s still your primary goal, and everything you do at this stage will be directed toward it.

People

We have already addressed interactions with sponsoring agency staff and others involved in the plan review process. Here we will concentrate more on your own staff, panelists, site staff, and others who will help you conduct standard setting. There will be more about each group in the next part of this manual when we get to the actual conduct of standard setting.

Your staff. This will likely be a fairly small group, one or two people for a certification and licensure engagement or a dozen or more for a large-scale K-12 engagement. While it may not have been necessary to name each staff member in the plan, it is necessary to identify everyone now. Depending on your relationship with the sponsoring agency, it may also be necessary to submit not only staff names but their credentials for approval. Find out if that is necessary, and follow through as needed.

Having identified staff who will participate in standard setting, identify each person’s role – in writing – and begin the training process. First, however, make sure everyone involved has the dates of standard setting and any related meetings on their calendars. These additional meetings will include, at a minimum, an initial orientation and a later final run-through prior to leaving for standard setting. There may also be one or more PLD development meetings or presentations to the sponsoring agency. Make sure everyone has cleared the dates for these events on their calendars.

The purpose of the initial orientation session is to clarify roles and responsibilities. Prepare a job description for each staff member and give them time to study them prior to the meeting. See [Appendix D](#) for samples. This meeting can normally be accomplished in a couple of hours. During the meeting, go over the plan and the basic flow of standard setting. How much time you spend on these topics will depend on how much experience the staff have with standard setting. If there is a wide range of experience, consider having two initial orientations, one for the more experienced, and one for the less experienced. At some point, however, you will need to bring them together for training so that they can get used to working with one another.

Final training will come a week or so before standard setting. By this time, you should have all the materials, directions, and scripts completed and approved. Give these to the staff, and allow them a couple of days to study them. Then bring them together for a final session to go over their roles and work flow. For facilitators (who may be content experts or psychometricians or a combination of both), conduct role play scenarios, and get other facilitators to critique. Offer your own suggestions, and give everyone an opportunity to participate and critique.

For data analysts and psychometricians who will not be serving as facilitators, go over each role and who performs it. It may also be necessary to do some cross-training at this point (as part of your contingency plan), so know in advance what particular skills each staff member has.

For support staff, identify all support activities in detail, and make sure everyone knows who is doing what. A written list of individual roles and responsibilities – by name – is essential.

If at all possible, try to schedule a dry run for all staff involved in standard setting. This event will include the opening session presentation, all training sessions, and mock standard setting rounds. If there are several facilitators, conduct one session in which one serves as facilitator and the others serve as panelists. Go through all steps in the standard setting process for that round. Do the same for inter-round discussions (with dummy data), logging in to software, and completing any necessary forms. From round to round, let facilitators take turns serving as panelists or facilitators. Note any problems that occur, and correct them prior to leaving for standard setting.

Once you reach the location where you will conduct standard setting, it is a good idea to gather all staff the evening before standard setting begins (or the morning before if standard setting starts in the afternoon) to make sure they have everything they need, that their breakout rooms are properly set up, and that they are ready to begin. Typically a half-hour meeting (largely to make sure everyone got there) is sufficient to go over roles and responsibilities. Afterwards, give facilitators time to inspect their breakout rooms, and give other staff time to check out their work areas and site layout.

Panelists. The plan will have specified not only how many panelists there should be but what characteristics they should possess as a group. Now comes the task of translating numbers in a grid into human beings. Getting the right people on the panels requires close cooperation with the sponsoring agency, diligence, and at least a little bit of luck.

Because final approval of the plan may go right up to standard setting time, it is best to get numbers and characteristics of panelists approved at the outset. That usually is not a problem. What you are going to do at standard setting may remain in flux for some time, but the people who will be helping you do it can be established months in advance. Get that approval out of the way as quickly as possible, and start identifying and securing panelists.

Some sponsoring agencies have standing committees for standard setting. These standing committees may be highly qualified, or they may simply be the same people who have been serving on a number of committees for years, in which case, it may be time to branch out.

For small organizations, particularly for small certification and licensure organizations, panels will be small, and the people available to serve on those panels will be fairly quickly defined. In this case, selecting panelists will be a quick and easy job. The same may be true for a local school district, particularly if it is small. Similarly, standard setting panels for individual college or university courses will generally be fairly small. Of course for an entire curriculum, with multiple panels, the numbers increase quite quickly.

For all other organizations, be prepared to spend a lot of time identifying, contacting, and securing panelists. For state and most local education agencies, there is a fixed protocol you must follow. Find out what it is, and follow it. It may be necessary, for example, for agency staff to make all contacts with potential panelists. Your job will most likely be helping to draft the letters, memoranda, and other communications that proceed from the agency. [Appendix A](#) contains some sample letters and e-mails we have used with good results. There are also [links](#) to other panelist recruitment materials located at the end of this manual.

The first letter should not actually invite the recipient to standard setting. Instead, it should explain the activity, give the dates, and ask if the recipient is available and interested. The letter should include a response form (either paper or electronic) that must be completed in full. It should include evidence of relevant content expertise (e.g., certifications, coursework, courses or grades taught) and key demographic characteristics. The approximate distribution of those characteristics should be worked out in advance between you and the sponsoring agency. Last but not least, there should be a hard deadline for receipt of the response. Give yourself enough time to go through the responses to put together representative panels.

As candidates respond to the initial inquiry, you will build a matrix of potential panelists like the one shown in Table 2.1. This particular hypothetical target matrix is for a statewide language arts test administered in grades 3-8 and 10. The plan called for 15 panelists per grade, representative of the statewide teaching population by race and sex. The plan specifically called for on-grade as well as above-grade classroom teachers. For grade 10, the above-grade teachers could be teachers of grade 11 or grade 12 language arts classes or representatives of two-year or four-year colleges. Teachers of English language learners (ELL) and of students with disabilities (SWD) are also included by design to make sure the cut scores take into account the needs of these populations. Administrators (building and district level) are included by design, as are members of the general public. By design, on-grade classroom teachers will constitute a majority on every panel. Also note that the cell numbers are approximate. Work out with the sponsoring agency how much wiggle room you will have with the final counts per cell.

Table 2.1
Sample Panelist Target Matrix for One Subject

By Role	Grade							Total
	3	4	5	6	7	8	10	
On-Grade Teacher	8	8	8	8	8	8	8	56
Above-Grade Teacher	2	2	2	2	2	2	2	14
ELL Teacher	1	1	1	2	1	1	1	8
SWD Teacher	1	1	2	1	1	1	2	9
Administrator	1	2	1	1	1	2	1	9
General Public	2	1	1	1	2	1	1	9
Total	15	15	15	15	15	15	15	105

By Sex								
Female	9	9	9	8	8	8	7	58
Male	6	6	6	7	7	7	8	47
Total	15	15	15	15	15	15	15	105

By Race								
Asian	1	2	1	1	2	1	2	10
Black	5	4	5	5	5	6	5	35
Hispanic	1	2	1	1	1	1	1	8
White	7	7	7	7	6	6	6	46
Other	1	0	1	1	1	1	1	6
Total	15	15	15	15	15	15	15	105

It is helpful to identify four or five times the number of people you will ultimately need. The sponsoring agency will supply the names and contact information and send out the letters you draft or help to draft. With any luck, about half the people contacted will respond. Duplicate the target grid, and start each cell at zero. As positive responses come in, add to each cell. For example, if a Hispanic female fifth grade teacher responds, you can add 1 to On-Grade 5 or Above-Grade 4 and also add 1 to grade 4 (or 5) Female and Grade 4 (or 5) Hispanic. Continue adding until all cells are filled. You may have to move some people around by grade (e.g., moving a 5th grade teacher from grade 5 to grade 4 as an Above-Grade teacher).

At some point, you may realize that you won't be able to fill some cells based on the responses you have received. At this point, you will need to start making phone calls. Start with the people who have not responded. Again, depending on protocol, you may be able to make these calls, or the sponsoring agency may have to make them. Find out, and follow the protocol. Give those people a week or so to submit their responses, and continue filling your matrix. If you are still short, another round of calls will be in order. Again, follow the sponsoring agency's protocol for identifying and contacting additional people. Keep at it until the matrix is filled.

Now, you are almost done. As noted in the previous section, people will drop out for various reasons. Plan on a list of back-up panelists. Keep that list separate from the primary panelists.

Submit the final matrix and list of names to the sponsoring agency for approval. After you receive approval (or revisions), you may send out invitations to the panelists selected. Make sure the invitation includes a deadline for responding. Do not send letters to those not selected just yet. Wait until the deadline for responding to the invitation has passed. Then call those who have not responded to the invitation. Only after confirming each invitee as a Yes or No should you contact the candidates not selected. If some of the people you invited have turned down the invitation, replace them with someone on the not-selected list with the same qualifications (i.e., by role, sex, and race). Continue doing that until you have a full matrix of confirmed panelists. This would also be a good time to negotiate for a bit of wiggle room in final counts. If the target is 15 per panel, would a panel of 13 or 14 be out of the question? What about 12? Under what conditions would 12 or 13 or 14 be acceptable? Get written confirmation (an e-mail will suffice) and keep it handy.

But you are not quite done yet. Prepare a different letter to send to people on your backup list. This letter needs to be very carefully written to make it clear that you plan to call them, perhaps at the last minute, to serve on a panel as a substitute. In this letter, you will simply ask if the person will agree to be available on the days in question and willing to travel on a moment's notice, or not, as circumstances dictate.

Now send the final letter to those candidates not selected to serve on a panel. This letter needs to be sensitively composed to avoid problems when you may want to include these people in another activity later on. Be sure to thank them for their willingness to serve, and note other activities coming up in the future that may interest them.

This whole process, except for converting backup panelists to actual panelists, should be completed a month or more in advance of standard setting. If there is to be a separate meeting to draft or review PLDs, you will actually go through this process twice unless you put both events on the same initial letter of inquiry (not invitation!) and ask candidates to indicate interest in and availability for one or both events. If that is the case, follow the steps outlined above for both events.

If there will be a vertical articulation of cut scores across grades, the panel for that activity will usually be a subset of the initial set of panelists. Returning to our example in Table 2.1, let us assume that we want three panelists from each grade/subject panel to remain for an extra day for vertical articulation. The vertical articulation committee (VAC) for this standard setting will therefore be 21 people. Selection of members of this group is just as important as the original selection, perhaps more so. Their composition might look something like Table 2.2. Negotiate cell counts with the sponsoring agency.

Letters of invitation to this group will be a little different from those sent to other selected panelists, since they will need to be available for one extra day. Be specific about their responsibilities and that they should let you know immediately if they are not available for the entire time period.

Table 2.2
Composition of VAC, Based on Table 2.1

	Grade							
	3	4	5	6	7	8	10	Total
By Role								
On-Grade Teacher	2	2	2	2	2	2	2	14
Above-Grade Teacher	0	0	0	0	0	0	1	1
ELL Teacher	0	1	0	0	0	1	0	2
SWD Teacher	0	0	0	0	1	0	0	1
Administrator	1	0	0	1	0	0	0	2
General Public	0	0	1	0	0	0	0	1
Total	3	3	3	3	3	3	3	21
By Sex								
Female	2	2	2	2	2	2	1	13
Male	1	1	1	1	1	1	2	8
Total	3	3	3	3	3	3	3	21
By Race								
Asian	0	0	0	0	0	0	1	1
Black	1	1	1	1	1	1	1	7
Hispanic	0	1	0	0	0	1	0	2
White	1	1	1	2	1	1	1	8
Other	1	0	1	0	1	0	0	3
Total	3	3	3	3	3	3	3	21

Site staff. This section assumes that you are conducting standard setting at a location other than your own facility. Even if you are conducting standard setting at your own facility, much of this section will apply to you.

Site staff will generally include staff of a hotel or conference center where you will conduct standard setting. You will need to have everything spelled out in a contract. Once you know the dates of standard setting, contact one or more sites and explain your needs. Select one, and start negotiating a contract. You, the sponsoring agency, and a hotel manager will need to be involved in this process. The contract should cover meeting rooms, guest rooms, meals, equipment, internet (if applicable), and incidentals. Most hotels have minimum charges for just about everything. Find out what they are, and make sure they are specified in the contract. Some hotels will provide one complimentary room or suite for every 10 or 20 rooms rented. Find out if the hotel you are considering will do that. Some hotels will forego meeting room fees if you guarantee a certain number of meals. In short, there will be a lot to address in negotiating a contract with a site. If that is not your cup of tea, make sure someone on your staff has that skill, and assign that task to him or her.

Once the contract is set, you will be dealing with reservations, catering, and front desk staff. For larger events, you will also likely be dealing with a manager assigned to your event to make sure everything goes smoothly. Find out who that person is, and start a dialogue well in advance of the meeting. Building a relationship with that person will make on-site activities go much more smoothly than might otherwise be the case.

Make sure someone at the meeting site understands the flow of your standard setting event. It is not necessary that they have a copy of the plan (actually, since the plan will contain some secure information, hotel staff should never see it). They do need to know how the event will flow, from check-in of your staff and panelists, to the walk-through the night or day before the meeting begins, to breakfast on Day 1, to the opening plenary session, to breakouts for the individual panels, and on to the final session on the last day. Prepare in advance a flow chart showing which meeting rooms will be needed at which times and the approximate numbers of people in each of those rooms during each block of time throughout the day. Include in the flow chart when and where audiovisual equipment and internet connectivity will be needed.

In addition to understanding the overall flow of your event, hotel staff will need to know how you want each meeting room set up. Get meeting room schematics from the hotel and mark them up to show how you want them set up (see Figure 1.1). For plenary sessions, do you want auditorium, classroom, or some other style setup? Hotel staff will generally ask you, but it is always good to know what you want so you can tell them.

Others. In the time between the first draft of the plan and conduct of standard setting, there will be many opportunities to interact with sponsoring agency leadership and staff, external reviewers, and colleagues. This is a time of relationship building. If you haven't already done so, get to know these people, particularly those who will be on site helping, observing, or evaluating you. Find out what their expectations are and what you will need to do to meet those expectations. Make sure they know your needs and expectations and how they can help you meet them. Keep them informed of your plans, especially as conditions and plans change. Enlist them in keeping the plan on track.

Materials and Equipment

Whether you conduct standard setting online or with paper, you will need to develop materials and get them approved for use in standard setting. These typically include the following:

- PowerPoint presentations
- Facilitator scripts
- Test materials for panelists to use
- Forms for panelists to use
- Data entry forms and processes
- Data analysis and output processes
- Miscellaneous signs, badges, and related materials

Details are provided below.

PowerPoint presentations. It would seem that you can't give a presentation without PowerPoint, Prezi, or some other presentation software. If you lack expertise in this area, now is the time to gain some. For every presentation, start with the task(s) panelists will need to complete. Break down those tasks into their component parts. What exactly will panelists do, and what will they have to know in order to do it? Focus only on those two things, and eliminate everything else. Sample presentations are included in [Appendix E](#).

Most standard setting activities involve three or four primary tasks for which panelists will need to be trained:

- *General orientation to the process* – Cover this in the opening plenary session that focuses on the basic issues listed below. The presentation should be concise and factual. In most instances, this presentation may be delivered by someone from the sponsoring agency, by you, or by the two of you. Having more than two speakers will make this more complicated than it needs to be. It is fine to introduce additional agency staff or your staff, but this is not a time for them to make speeches.
 - Welcome and thanks
 - Why we are doing this
 - How you happen to be here
 - Your primary role
 - The tasks you will complete
 - The schedule and logistics
 - General groundrules
 - Where you will go
- *Orientation to the test* – This may include taking and scoring the test or just reading through it. In either event, the presentation should clarify what you are going to do, what each panelist is going to do, and who is going to help them do it. This presentation can also include some very basic background on the test blueprint and development process. This is also a good time to point out that the test has already been through an extensive review and tryout process and that the items on the test are not up for further modification.
- *Orientation to the PLDs* – As with orientation to the test, this may include background on how the PLDs were developed. However, the primary purpose of this session is to acquaint panelists with how they will use them in standard setting. The presentation should be fairly short so that panelists can spend most of the session reviewing and discussing the PLDs with their facilitator and with one another.
- *Orientation to the standard setting procedure* – Over the years, we have pared down these presentations considerably. Where we once spent several slides and quite a bit of time explaining the different approaches to standard setting and the technical specifics of the procedure we would be using, we now focus just on the tasks panelists need to complete and how they will complete them. We do spend some time on the *why* of the process, but our primary focus is on the *how*. A typical presentation for this session will

run about 45 minutes and contain no more than about 30 slides, and that includes an overview of software if relevant. Rather than spending more time trying to make sure everyone gets your point, break after about 45 minutes and let panelists demonstrate what they have just learned in a trial run. We include a practice round in every standard setting we conduct. Learning by doing is far superior to anything else we have ever tried.

- *Round-by-round review* – These presentations are typically not PowerPoint based. Instead, they are based on output from the data analysis programs. However, they are presentations, and they should be scripted and rehearsed just the same as any PowerPoint presentation. The objectives are the same: Explain the task panelists will complete next, tell them how they will complete it, and let them discuss the results of the previous round to the point that they understand what is expected of them and how they will meet that expectation.
- *Vertical articulation* – Not every standard setting activity will have vertical articulation. Certification and licensure standard setting typically focuses on a single test for a single job title and would therefore not include vertical articulation. Academic tests given in non-contiguous grades generally will not require vertical articulation either. Under current circumstances, this task will only apply to language arts and math, and even then only for activities in which multiple, contiguous grades are involved. The presentation for this activity will typically focus on the tasks listed below and last only 30-45 minutes. The rest of the time, we have found that an interactive Excel spreadsheet similar to the one shown in Figure 2.1 will be the main focus. A working model of Figure 2.1 is included as [Worksheet 3](#).

Vertical Articulation for Grades 3-8

Final Round Thresholds (Page #)				Percent At or Above			Percent in Level			
Grade	Level 2	level 3	Level 4	Level 2	Level 3	Level 4	1	2	3	4
3	12	26	50	68.7	41.3	10.2	31.4	17.2	31.0	10.2
4	12	28	55	70.5	37.8	6.1	29.5	26.6	31.7	6.1
5	10	22	52	70.2	40.7	2.4	29.8	27.2	38.3	2.4
6	17	30	52	52.9	24.8	2.7	47.1	25.4	22.2	2.7
7	18	32	60	73.8	41.1	3.1	26.2	29.5	38.0	3.1
8	8	24	64	75.2	36.1	1.7	24.8	37.4	34.4	1.7

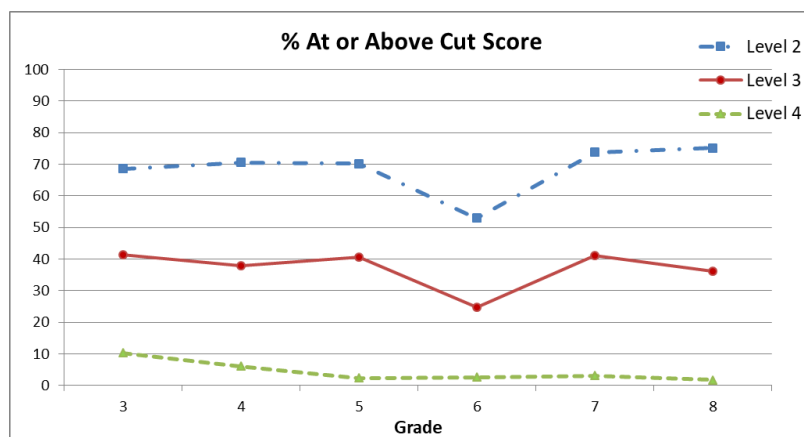


Figure 2.1. Sample vertical articulation interactive spreadsheet

Facilitator scripts. Facilitators will lead panelists in review of the tests and PLDs, a practice round, one or more rounds of standard setting, and possibly vertical articulation. Each of these activities should be scripted. I do not mean that every word should be written out and memorized; rather, the main talking points should be carefully spelled out so that every facilitator shares the same message, subject to the specific requirements of the panel they are leading. Sample scripts are included in [Appendix B](#).

Test materials for panelists to use. For paper-based standard setting, panelists will need a copy of the test, an answer document, and a scoring guide if they will actually take and score their tests. If they take the test, each panelist should score his or her own. Trading papers is time consuming and potentially embarrassing. Prepare sufficient quantities of these materials in advance. Since all these materials should be available months before standard setting, assembling them as soon as they are available leaves more time for preparing other materials closer to the time of standard setting.

In the case of bookmark, item descriptor match, or other standard setting procedures that require an ordered item booklet, the order of items may not be known until a week or less before standard setting. To expedite the process, set up the OIB production process well in advance – all eligible items, selection criteria, and ordering process. For constructed-response items that require one or more sample responses for each score point, we have found that by identifying key districts for early scoring, we can get an adequate sample of student responses to the operational test to satisfy this requirement of OIB construction. If that is not possible, responses from the field test may be satisfactory.

In the case of body of work or other procedures that require samples of student work, operational responses are preferred, but field test responses may be necessary for some or all of the samples. The key is to select a range of responses for each score point: short, long, in-between, mix of subscore combinations (for items scored on multiple dimensions), and a range of penmanship and spelling (if responses are handwritten) or signs of computer proficiency (if responses are key entered). Collect these samples as early as possible, let content experts (your own as well as those of the sponsoring agency) review them, and order them from lowest to highest total score. If conducting a paper-based standard setting, make sufficient copies of each set of responses and package them for shipment to the meeting site. For online standard setting, tests, support materials, student work samples, PLDs and other materials required for your specific standard setting procedure should be loaded into the software as soon as possible and tested.

Forms for panelists to use. Whether standard setting is done online or on paper, there will be forms for panelists to use. These will include not only the customary workshop materials (sign-in forms, security agreements, and reimbursement forms) but procedure-specific forms such as bookmark forms, body of work rating forms, readiness forms, and evaluation forms. For vertical articulation, there will be a special form for the facilitator to use to record motions and votes. Several sample forms are included in [Appendix C](#). Others are included under separate cover and listed in [Appendix E](#).

Data entry forms and processes. Recall that the objective of standard setting is to obtain input from panelists and turn that input into cut scores. Over the years, we have found that the more this process can be automated, the faster it will go and the fewer mistakes there will be. However, for now, let us consider data entry and processing in both paper-based and automated formats.

- *Paper-based* – Figure 1.3 showed a sample bookmark rating form. For this form, we created a spreadsheet similar to the one shown in Figure 2.2. A working model is included in [Worksheet 4](#), along with Excel spreadsheets for Angoff and body of work procedures. Facilitators collected the completed forms and handed them to data entry staff who entered them into the spreadsheet. The completed spreadsheet was set up to yield median page numbers. The OIB production program was set up to convert page numbers to theta values (Ach. In Figure 2.2) and ultimately into cut scores.

Results of Round 1 of Rating						
	Level 2		Level 3		Level 4	
Rater	Page	Ach.	Page	Ach.	Page	Ach.
1	11	0.273	27	0.900	36	1.200
2	9	0.193	23	0.616	31	0.998
3	13	0.420	25	0.810	27	0.988
4	9	0.193	25	0.810	40	1.489
5	12	0.286	26	0.891	38	1.333
6	7	-0.176	20	0.569	33	1.090
7	16	0.493	23	0.616	41	1.510
8	8	0.082	27	0.900	40	1.489
9	10	0.270	31	0.900	38	1.333
10	12	0.286	24	0.740	38	1.333
11	11	0.272	21	0.579	38	1.333
12	11	0.272	22	0.600	36	1.200
13	10	0.270	16	0.493	35	1.191
14	17	0.540	23	0.616	33	1.090
15	14	0.440	26	0.891	39	1.340
Median	11	0.272	24	0.616	38	0.988
Min	7		16		27	
Max	17		31		41	
Range	10		15		14	

Figure 2.2. Sample bookmark tally sheet in Excel

Data analysis and output processes. Figure 2.2 shows how one bookmark Excel spreadsheet is set up. Figure 2.3 shows how those tallies are translated into feedback given to panelists between Rounds 1 and 2.

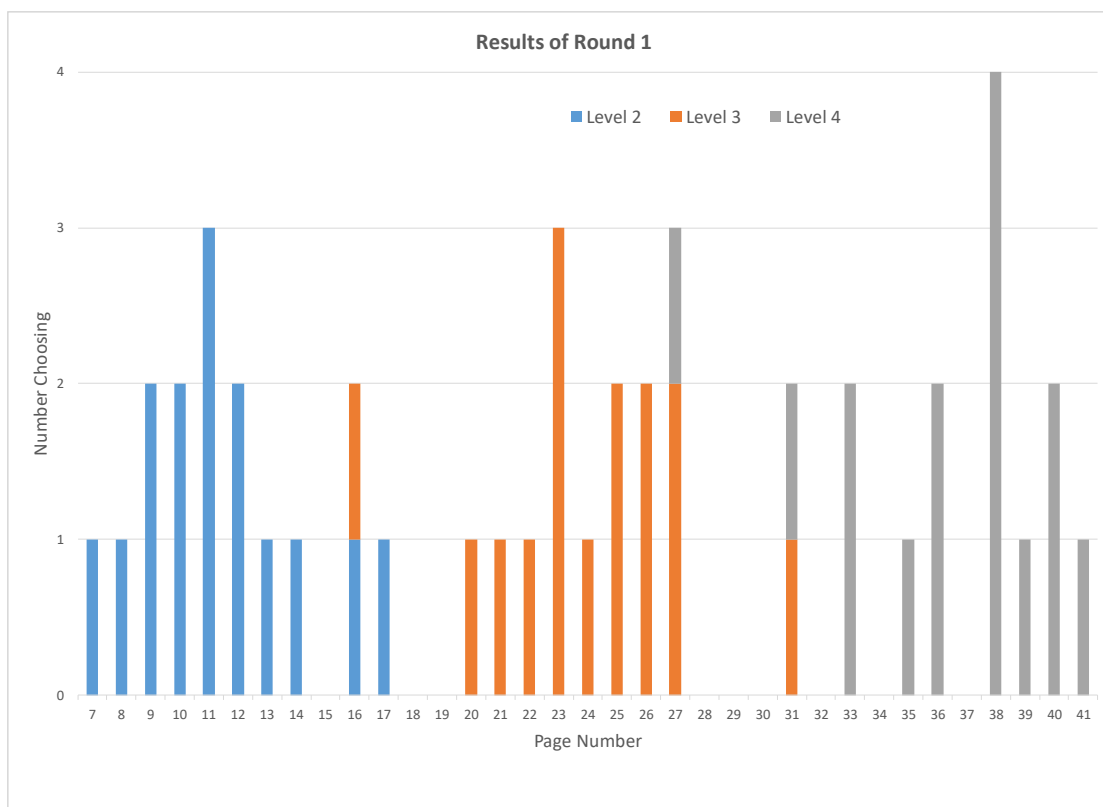


Figure 2.3. Sample feedback based on the tallies in Figure 2.2

Miscellaneous signs, badges, and related materials. Like any workshop, standard setting will require signage for rooms, tables, and participants. Workshop leaders and facilitators should be easily identifiable by name tags or badges. Panelist name badges should have not only their names but their panel designations as well (e.g., 5M2, indicating Grade 5 Math Table 2). Table tents are also helpful to indicate where presenters should sit and to identify table numbers in the breakout rooms. If there will be several breakout rooms, work with hotel staff to identify them, and get a map of the part of the hotel where these rooms are located. Although the rooms will likely have their own signage (e.g., Windsor Room, Grand Ballroom, etc.), each plenary session room and each breakout room should have an additional sign indicating the panel that meets there. Work with hotel staff to make sure they post appropriate signs or will allow you to put up your own signs.

At particularly large meeting sites, it may be a good idea to supply all participants with a map. In addition, for a meeting that will last two or more days, provide lists of eating establishments and other points of interest (shopping, entertainment, cultural events) for all participants. Hotels usually have such lists to distribute to guests. The meeting agenda will normally include breakfast and lunch, but not dinner. At the end of the day, panelists need to get out, eat dinner, and think about something other than standard setting for a few hours. Helping them do that is part of your job, so make sure you have good maps and lists.

Logistics

Whether the event is for a single licensure test or a multi-state, multi-grade battery of tests, people and things need to get from Point A to Point B with a minimum of confusion. This section address travel, lodging, shipping, and miscellaneous logistical issues.

Travel and lodging. Let's start with what looks like a very simple operation – seven subject matter experts meeting to set a cut score for a licensure test. This will most likely be conducted as a webinar, due to the geographic dispersion of the panelists and other demands on their time. Your main task will be to find a time (half day, full day) that all seven can participate in a webinar. Any one of a number of web-based calendar programs will work for this task. Once you have a set meeting time, send full directions and pre-meeting materials to each panelist. The day before the webinar, send out reminders that include log-in information and a number to call in case there are problems logging in. Make sure someone (not you, as you will be busy conducting a webinar) is available at that number to help.

Now let's take those same seven content experts and have them meet face to face for a day and a half. The meeting starts at noon on July 10 and ends at 4 P.M. on July 11. Six of the seven experts will fly to Chicago from six different locations. The seventh will be driving in from Peoria. You need to get them all to Chicago, house all of them for at least one night and one of them (coming from London) for two nights. Here's your task list:

- Find out if anyone has any special travel or lodging needs
- Make 6 airline reservations that comply with special travel needs
- Make 7 hotel reservations in accordance with board policy and special lodging needs
- Find out if anyone has any special dietary needs
- Arrange for lunch on July 10 and breakfast and lunch on July 11
- Send reminders to all 7 experts on July 8
- Conduct a site visit on July 9 to make sure all is ready
- Prepare reimbursement and honoraria forms

There will be other tasks, depending on the nature of the meeting and the characteristics of the seven content experts, but this list covers most of it.

Now consider a four-day standard setting activity with vertical articulation for a state department of education: 210 panelists (most of whom will drive in daily), 14 facilitators, support staff, and a dozen or so department of education officials, staff, and observers. The meeting will be held in a large hotel in the capital city, so travel and lodging arrangements for department of education staff will not be your responsibility. You will be responsible only for travel and lodging for yourself and approximately 90 other people (your staff plus 75 panelists who will not drive in every day). This is why travel agencies exist.

Start with the hotel. You should have worked out a contract shortly after you set the date for standard setting. That contract will have an approximate number of room nights, with definite beginning and ending dates. Confirm your block of guest rooms and meeting rooms and the schedule for using them. The vertical articulation committee will be a subset of the full set of

panelists, and they will stay one night longer than everyone else. Make sure those panelists are accounted for.

Create a guest list that you and a hotel banquet manager will work from. This list will include everyone who will stay at the hotel, panelists as well as you and your staff and anyone else coming in from out of town (e.g., the official reviewer). Get the information for your staff on the list right away. As individual panelists confirm their participation, add their names and contact information to the list. There will likely be changes along the way, so be sure to keep the list up to date when someone drops out or is added.

One feature of the guest list will be special requests: dietary restrictions, handicap-accessible guest rooms, handicap parking, and similar issues. The invitation letter should be very specific about these things, and panelists should be responsible for letting you know about their needs. However, some will fail to do so. Keep a small number of handicap-accessible rooms on hold, and notify the banquet manager to be prepared to provide a small number of special meals on short notice.

Travel arrangements come next. You really should have one person working on this full time and not try to do it yourself. Let that person read over the plan, discuss the big picture, and let them take it from there. They should update you on a regular basis.

For most states, most panelists will drive to the meeting. However, you still need to know who will drive and who will fly (or travel by train or some other public transportation). The invitation letters need to be explicit about limits on travel costs (e.g., no first class airline tickets, no seat upgrades, no duplicate mileage for people traveling in the same car, and other constraints imposed by state law or regulation).

Once you know who is flying (or taking other public transportation), give the travel agent their names and contact information. It is best to let the travel agent, working with a member of your staff, make the arrangements rather than ask panelists to make their own. We have tried it both ways, and it is always cheaper and more efficient to do it ourselves. In addition, we get a complete travel itinerary for everyone. The same is true for staff – most of you will go at the same time and probably on the same flight; there is no need for 14 different people to make travel arrangements on their own.

The final issue is in-transit communication. For each person attending standard setting, add their cell phone number to a master list and have a member of your support staff maintain that list. The cell phone number of that person should also be in the final letter to each invited panelist. They should be directed to call that number if anything goes wrong. If someone doesn't show up at the hotel at the expected time, call them. If you and your staff get separated in transit, call them to make sure everyone gets to the meeting site on time and in good shape.

Shipping

For paper-based standard settings in particular, there will be a lot of paper to ship from your location to the meeting site. If you are conducting on-site online standard setting, there will be a lot of equipment to ship. Let's take a look at both.

Paper. Table 2.3 summarizes the paper products that you will need to produce and ship to the meeting site. This table assumes standard setting for seven grades and two subjects with 15 panelists per grade/subject.

Table 2.3
Paper Products to be Shipped

	Modified Angoff		Bookmark or IDM		Body of Work	
Item	Pages	Copies	Pages	Copies	Pages	Copies
Test booklet	60	210	60	210	60	210
Answer document	4	210	4	210	4	210
Scoring guide	4	210	4	210	4	210
PLDs	6	210	6	210	6	210
OIB			72	210		
Passage booklet			24	210		
Student work samples					60	210*
Readiness form	2	210	2	210	2	210
Evaluation form	1	210	1	210	1	210
Nondisclosure form	1	210	1	210	1	210
Reimbursement form	1	210	1	210	1	210
Sign-in sheet	3	14	3	14	3	15
Total Pages	16632		36792		16635	
Cases of Paper	33		74		33	

* Plus replacement work samples for Round 2

Bookmark and IDM use a lot more paper because of the need for a regular test booklet as well as an ordered item booklet. A case of paper weighs about 60 pounds, so the total shipping weight for this standard setting would be about 4,400 pounds. For Angoff or body of work, the total weight would be closer to 2,000 pounds. UPS and FedEx have bulk shipping rates, and other carriers are even cheaper (but perhaps not as reliable). The point is that all this printing and shipping will be a significant part of your overall standard setting budget and logistics plan. Make sure you include them. In some instances, it may be more economical to ship or carry originals to the site and make copies there, particularly those items you will use on the second or subsequent days, or you may prefer to rent a van or truck and transport large quantities of paper to the site. For shredding of used secure materials, consider contracting on site rather

than shipping everything back home and shredding it there. Calculate the costs for all approaches and proceed accordingly.

Online. Online standard setting has its own shipping issues. For any of the workshops detailed in Table 2.3, there would be 210 panelists plus you and your staff, for a total of about 230 computers and monitors to ship to the site. Laptops are convenient and reasonably lightweight, but they are expensive. We have found it economical to use Chromebooks (see Figure 1.5) from the standpoint of cost to purchase or rent and cost to ship. This will be another area where comparison shopping will be a good idea.

Online standard setting done on site also requires cabling, power cords, and surge protectors to keep all 210 panelists online when they need to be online. You will essentially be setting up a computer lab for each panel. This becomes not just a shipping issue but an on-site logistical issue in itself, as you will have to determine the best way to set up the computers and monitors in each room to make best use of electrical and internet connections.

For any activity, you may also want to have your own copier and printer for inter-round reports. Factor the purchase or rental and shipment of all these devices into your plans.

Management Activities

Between the time you submit the draft plan and the time you start standard setting, you will have considerable interaction with the sponsoring agency. The exact nature and extent of these activities will depend on the size and complexity of the undertaking. In any event, they will include those listed below. Work out in advance with the sponsoring agency how often and how detailed each should be.

Plan revisions. Recall from Part 1, there will likely be ongoing revisions to the plan before its final approval. Work out in advance how that process will play out, and monitor it closely. See **Follow the process** (p. 23) and *Worksheet 1* for details.

Communications. As the plan makes its way through the review and approval process, recruitment and retention of panelists progresses, and arrangements with the meeting site are finalized, keep the sponsoring agency apprised of progress and problems. Agree in advance the nature (formal report, phone call, e-mail, text) and timing (monthly, weekly, daily, or accelerating as the date of standard setting approaches) of updates. Then stick to that communication plan. At the same time, there will be a series of communications with panelists. See **Panelists** (pp. 27-28) and [Appendix A](#) for details.

Part 3: Implementation

After months of planning and preparation, you are finally on site, ready to begin standard setting. Congratulations!

Primary goal – Get Approved Cut Scores

Yes, this is still your primary goal and will remain your primary goal until cut scores are approved. Everything you do on site will be directed toward this goal.

Preliminaries

Make sure you and your staff arrive at the site at least a day in advance of the opening plenary session. Check with the front desk to make sure all reservations are in order and that they have not heard from any panelists with last-minute cancellations. If you have not already done so, meet with the banquet manager for a tour of the facility. Have someone on your staff available to handle problems with reservations, even though hotel staff technically have that responsibility. You will have sent that person's phone number to all panelists, and he or she should be the one to resolve any issues a panelist may have.

Gather your staff for a brief meeting to go over roles and responsibilities one last time. Make sure everyone has arrived and has no guest room issues. Afterwards, inspect the meeting rooms. You should have made arrangements in advance to have the banquet manager or a member of the hotel staff open each meeting room. If they are in use, at least get a quick peek into each room and go over your schematics with hotel staff to make sure they understand how the rooms are to be set up the following morning.

Go over the agenda one last time, rehearse your opening presentation, and get a good night's sleep. You're going to need it.

Opening Day

Arrive at the meeting site at least 30 minutes prior to the scheduled start time. If breakfast is on the agenda, go to breakfast and meet people you will be working with – panelists and agency staff and officials. Then go check out the opening plenary session meeting room. Is it set up the way you asked that it to be set up? If not, get the banquet manager or designee down there immediately and correct the situation before people start filing in. Check all audiovisual equipment. Load and project any PowerPoint presentations you will use during the opening session, and make sure they all run properly.

Make sure there is a sign-in table or station near the plenary session meeting room, but not inside it. Most larger hotels have a pre-function area just outside their larger rooms for this

purpose. For a large group, have two people at the sign-in table. Have any panelist materials (e.g., reimbursement forms, nondisclosure forms, name badges) packeted and spread out neatly on the table with panelist names in large enough letters (28-36 point *sans serif*) to be read from a few feet away.

Meet briefly with anyone who has a speaking role in the opening session. Make sure everyone knows the order of presentations, who introduces whom, who issues the official welcome, and what each speaker is to cover. Go over the agenda for the day and the objectives of the opening session. Make sure all facilitators are present and visible, as you will introduce them at the end of the session.

Start the opening session on time. There will be stragglers, but you are not obliged to wait for them. Show respect for those who do arrive on time. Of course, if the agency official in charge says to wait, wait.

Make your presentation like a professional:

- Rehearse, so that you know your material.
- Practice advancing slides.
- Speak clearly and at a moderate rate.
- Look at the audience, not the screen.
- Stick to the script; avoid unnecessary and time consuming *ad libs*.

At the close of the session, direct panelists to their breakout rooms. Your final slides should have the panel names, room names, and facilitator names in large enough letters that people in the back of the room can read them, as shown in Figure 3.1. Call out panels one by one, introduce the facilitators, and dismiss that panel before introducing the next facilitator. It may be helpful to start with the panel that will be meeting farthest away from the plenary session room and end with the panels meeting closest.

After the last panel has filed out of the room, meet briefly with the on-site leader of the sponsoring agency for a quick debriefing on the session. Set a time at the end of the day to get back together for a complete debriefing on the events of the day and a preview of tomorrow. Gather and secure your things and start your first tour of breakout rooms to make sure every panel has gotten started and that facilitators are following the script. You will repeat this process several times throughout this and subsequent days, taking notes on what you see and preparing to share those notes at the end of the day with your staff as well as sponsoring agency staff.

Grade Room		Facilitator
3	Altamaha	Craig
4	Chattahoochee	Jennie
5	Flint	Dan
6	Ocmulgee	Jami-Jon
7	Oconee	Joe
8	Satilla	Winnie
10	Savannah	Chris

Figure 3.1. Room assignment slide from opening plenary session

Training Sessions

Generally, there are three areas of training:

- The test – general orientation plus background and history and possible administration and scoring;
- The PLDs – what they are, what they represent, and how they are to be used;
- The procedure – general orientation to the procedure plus a step-by-step explanation of what panelists will do, including orientation to the software being used, if applicable.

Responsibility for training will have been determined well in advance of the meeting. For orientation to the tests and the PLDs, training may be shared by agency staff and contractor staff. If that is to be the case, the presenters should have worked out in advance who leads on which part. As this training typically takes place in the individual breakout rooms, there may not be enough agency staff to cover all panels at one time. If there are general issues to address about either the tests or PLDs, it may be more appropriate to include those issues in the

opening plenary session, with agency staff providing most or all of the instruction. The remaining subject/grade-specific issues can then be covered in the individual breakout rooms by contractor staff.

Training for the specific procedure to be used should be a plenary session conducted by the lead facilitator for the event and should be the final training session. Ideally, the person leading this session will not be facilitating a group and can use the time that facilitators are leading their panels through the tests and PLDs to set up the presentation for the standard setting procedure.

As with the opening session, the presenter(s) should be well rehearsed, have a thorough understanding of the procedure, and be able to explain it in a professional manner. There will likely be several questions during this session, so the time officially allotted to it can be stretched by a few minutes. However, it is appropriate to point out that there will be a practice session immediately following training, and there will be more time to ask questions and get answers then. The presenter and facilitators should work out in advance which types of questions are most appropriate for each session.

After orientation to the standard setting procedure, panelists will return to their individual breakout rooms for a practice round. During this session, panelists will examine a miniature version of the product they will review during the first round of standard setting. For example, in a bookmark procedure, panelists will examine a short (6-8 items) OIB and set one cut score, after which time they will discuss where they set their marks and why. For a body of work procedure, panelists will examine a small batch (again, 6-8) of work samples and assign them to proficiency levels, after which time they will discuss their rationales. In a modified Angoff procedure, panelists will evaluate a small group of items (6-8) and assign probabilities of correct response for multiple-choice items or predicted score for constructed-response items. In each instance, the purpose of the exercise is to allow panelists to experience the procedure in a no-stakes setting, realize that they can actually complete the task, and discover that there will be differences of opinion, which is quite all right.

Presentation of external data. If you are using external data to validate cut scores, consider carefully how you will present the data. In the example shown in Figure 3.2, we had scaled items from two college admissions tests to the test for which we were setting cut scores. The College Board and ACT have identified score ranges on their tests that correspond to college ready. We included these score ranges in the item map for a bookmark standard setting activity. Panelists saw these score ranges (highlighted in the last two columns of Figure 3.2) so that when they set their bookmark for Level 3 (Proficient), they could see if it were in, below, or above those ranges. In this particular example, the panelist has placed a bookmark for Level 3 on the row for item 22 in the OIB. This item has a theta (RP67) value of 2.61, which is in the range of theta values corresponding to a scale score of 18 on the ACT and 500 on the SAT.

Page	ID	Standard	DOK	Theta	Bookmark	Comment	Reviewed	ACT	SAT
15	1234	2.1.1	2	2.11			x		
16	2468	2.2.1	2	2.19			x		
17	3691	3.2.1	2	2.27			x		
18	4812	1.2.2	3	2.33			x		
19	5101	1.2.3	2	2.41			x		
20	6121	1.2.1	3	2.48			x		
21	7142-3	3.2.2	3	2.53			x		
22	8162	3.3.3	2	2.61	3		x		
23	9182	2.3.1	3	2.70			x		
24	1949	2.2.2	3	2.78			x		

Figure 3.2. Display of external data in an item map

Evaluation of training. After each training session, it is important to know if panelists have understood what they just learned and are able to apply it to the standard setting task they will eventually undertake. This rule applies not only to the training on the specific procedure but to all aspects of training. Therefore, it is a good idea to create and use a readiness form after each training session. On this form, panelists are able to indicate that they participated in the training, that they understood the content, that they believe they can apply the content, and that they are ready to move to the next step. You will continue to use the readiness form right up to the last event on the last day of standard setting. Sample readiness forms are included in [Appendix C](#).

Rounds

There is no set number of rounds for any standard setting procedure. Convention appears to favor three, but particularly when there is to be a vertical articulation, two rounds may be sufficient. For the purposes of this section, I will assume three rounds plus vertical articulation.

Round 1. Facilitators discuss the practice round and ask panelists to complete the portion of the readiness form indicating that they are ready to begin Round 1. They then orient them to the task (OIB, collection of work samples, test booklet, etc.) and invite them to begin. During this round, discussion among small groups (i.e., those sitting at the same table) is not only permitted but encouraged. However, each panelist is to enter his or her own ratings, not the collective judgment of the group. Of course, for remote online standard setting, panelists will be entirely on their own. In either event, panelists work through the assignment and submit their Round 1 input. If the event is paper based, the facilitator collects all secure materials and dismisses panelists as they finish. For remote online standard setting, the facilitator checks the submitted responses and contacts the panelist if there are errors or if the submission is incomplete. There is no need for everyone to stay until the last person finishes; there will be nothing to discuss until the beginning of Round 2. However, it is important that each panelist actually complete the task before leaving. Making sure they do so is the responsibility of the facilitator.

Between Rounds 1 and 2. As each panel completes Round 1, the data analysts analyze the data to produce round summaries similar to those shown in Figures 2.2 and 2.3. Each report should be independently verified by someone other than the person who produces it. The lead facilitator should see and approve all Round 1 output and share summaries with designated agency staff.

In the special case of body of work, it will be necessary to remove some of the Round 1 work samples and replace them with other work samples with scores in the general range of cut scores derived from Round 1. Round 2 will then be limited to this more focused group of work samples. See Cizek & Bunch (2007, Chapter 9) or Kingston & Tiemann (2012) for guidance in this regard.

Round 2. The facilitators begin with a brief review of Round 1 – what panelists thought of the task, whether or not they had difficulties, how they resolved those difficulties, and how they approached the task, using the training they had received. They will answer any questions the panelists may have.

When it is clear that there are no more questions, facilitators introduce the Round 1 results and go over them step by step, beginning with a focus on the dispersion of ratings. There is usually quite a range of ratings during the first round. It is good for panelists to see that. During this discussion, the facilitator will ask different panelists to explain one or more of their ratings, and encourage roomwide discussion. Everyone should have a chance to speak during this period; the facilitator may need to draw some panelists out or caution others to allow their fellow panelists a chance to speak.

After examining the dispersion of panelist input, move on to the dispersion of resulting cut scores. Focus on both central tendency and dispersion. Elicit panelist reactions to their individual cut scores and the panel mean or median cut score. Are there any surprises? If so, how big, and in which direction?

With three rounds, I generally recommend withholding impact data until after Round 2. However, some sponsoring agencies prefer to present impact data after Round 1. This is an extremely important point, as the timing of impact feedback can affect cut scores. Work that out with the sponsoring agency and highlight it in the plan.

After presenting and thoroughly discussing the results of Round 1, facilitators will invite panelists to complete the section of the readiness form that indicates their readiness to begin Round 2. Panelists will then proceed just as they did in Round 1. As in Round 1, panelists will discuss the task with others in their small groups (for face-to-face meetings only), and facilitators will dismiss them as they complete the task. As in Round 1, there is no need to keep everyone until the last person finishes, but the facilitator needs to make sure everyone actually completes the assignment.

Between Rounds 2 and 3. As before, the data analysts analyze the data to produce round summaries similar to those shown in Figures 2.2 and 2.3. Each report should be independently verified by someone other than the person who produces it. Review and approve each report, and share a summary with agency staff. Spend a few moments with agency staff reviewing the impact data.

Round 3. Round 3 begins as Round 2 did, with a discussion of dispersion of panelist ratings and cut scores. The new bit of information will be impact data – the percentages of examinees at each proficiency level based on the cut scores established by the panel in Round 2. Sharing of impact data is often called the “moment of truth.” Facilitators should give panelists a moment to digest the impact data before starting a discussion of it. They should answer all questions that arise and when there are no more questions invite panelists to complete the portion of the readiness form that indicates readiness to begin Round 3. Round 3 then proceeds exactly as Rounds 1 and 2, although usually much more quickly. As this is the final round, facilitators must be extra vigilant to make sure everyone validly completes the rating task and returns all secure materials before departing.

Because some panelists will complete Round 3 very quickly, it is a good idea for the facilitator to thank the entire panel for their participation before they begin their final task. Agency staff may also want to visit each panel at the beginning of Round 3 to offer their thanks.

Between Round 3 and vertical articulation. If there is to be a vertical articulation activity, prepare a spreadsheet similar to the one shown in Figure 2.1 for each subject, showing impact across grades for each performance level. By now, you should have a good idea how the cut scores and impacts line up across grades and where the trouble spots (if any) may be. By meeting daily with your staff, you should know what the issues are. Since this will be a new experience for all panelists, they may be reticent to recommend changes to any cut scores. Based on your previous conversations with staff (and the client), have a couple of recommendations of your own or at least one or two cut scores you would like the VAC to examine.

Vertical articulation. In Rounds 1-3, each panel has considered cut scores for a single grade (or perhaps two adjacent grades), without knowledge of where cut scores for other grades were being set. For this round, the lead facilitator is usually in charge. Orient panelists to the task by covering the following topics:

- The purpose of vertical articulation – to make sure cut scores and resulting impact make sense across grades
- Reasonable and unreasonable expectations – steady increase or decrease across grades in percentages of examinees at or above a given cut score or very nearly equal percentages at or above a given cut score rather than an up and down shape to the graph of percentages across grades
- Your task – review and adjust individual cut scores to achieve a continuum of impact that makes sense
- Groundrules – how to recommend changes, how to validate recommended changes

Allow time for questions and answers. Afterwards, invite panelists to complete the portion of the readiness form that indicates they are ready to begin vertical articulation. When everyone has so indicated, begin the process.

For this activity, the facilitator and panelists will need to have access to all PLDs and all tests or work samples used in the three rounds of standard setting. If this activity is paper based, it will

not be necessary that every panelist have every piece of paper. Indeed, it would be quite cumbersome to take that approach. Instead, arrange the seating so that each table or small group has representatives from every grade or at least every grade pair (e.g., 3-4, 5-6, 7-8, 10), as shown in Figure 3.3.

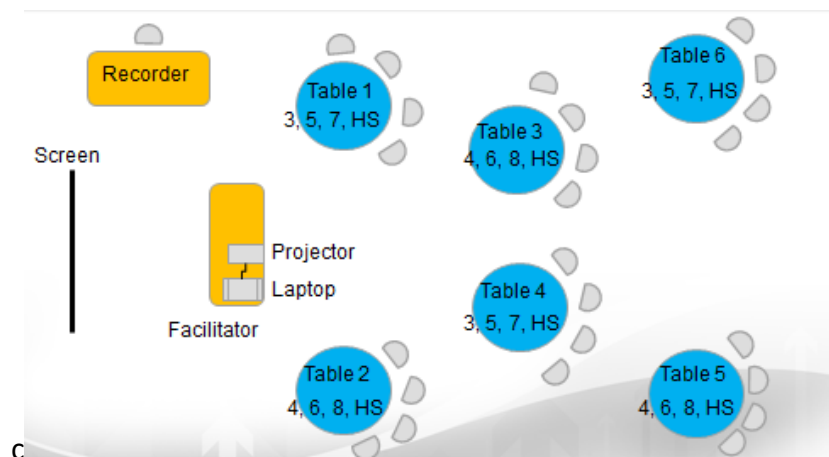


Figure 3.3. Sample VAC seating chart

The facilitator should have all necessary tests and PLDs for that grade to share with the rest of the table or small group. If the activity is computer based, then everyone can and should have access to all materials. Part of the orientation to vertical articulation should be a demonstration of how to access those materials. If Rounds 1-3 were computer based, this orientation should go rather quickly, as it is merely an extension of what they have already done but with more materials available. A sample vertical articulation orientation session PowerPoint presentation (Generic VAC Webinar.pptx) is listed in [Appendix E](#) under PowerPoint Presentations.

After orientation, display an interactive spreadsheet similar to the one shown in Figure 2.1 on a large screen. Explain what the numbers and graphics mean, and give panelists a moment to absorb what they are seeing. Then ask if anyone sees anything out of place, based on what they have just learned about vertical articulation. Someone may suggest a particular bump or dip in one of the graphs. Ask them to explain what that means to them. Ask others to comment on what it means to them. If the panel seems interested in pursuing the issue, ask if someone would like to recommend a closer look at that cut score for that grade, or you may make that recommendation yourself.

For any cut score so examined, direct panelists to open the appropriate test booklet to the appropriate page or the work samples at the appropriate score point as well as the appropriate PLD. They should then compare what they see in the test item or work sample with what they see in the PLD. Now take a look at the adjacent items or work samples. Would one of them be just as good a marker for this cut score as the one approved by a panel in Round 3? If so, is someone willing to recommend such a change?

To illustrate this process, we will work through a scenario for four recommendations. Returning to the VAC of 21 panelists suggested in Table 2.2, Jill S. has recommended that the Level 3 cut score for grade 7 be moved from a raw score of 42 to a raw score of 41. Table 3.1 shows how we would proceed with that recommendation, as well as with three other motions.

Table 3.1
Vertical Articulation Motions and Actions

Motion	Time	Second	Vote	Result/ Action
Jill S.: Move grade 7 Level 3 cut score from RS 42 to RS 41.	2:15 P.M.	David B.	19-2	Passed
Albert G.: Move grade 8 Level 2 cut score from RS 27 to RS 26	2:31 P.M.			Failed for lack of a second
Monica F.: Move grade 4 Level 3 cut score from RS 49 to RS 51	3:17 P.M.	Gerald P.	12-9	Failed; not a 2/3 majority
Sam B.: Move acceptance of all remaining cut scores as well as revised cut score for grade 7 Level 3.	3:49 P.M.	Hubert L.	20-1	Passed

In the case of the motion brought by Jill S. the facilitator asked for a second, which David B provided. The facilitator recorded the motion, second, and time and then directed the VAC to the materials for grade 7 and the PLD for Level 3. If this had been for a bookmark procedure, members of the VAC would have been directed to the page in the OIB corresponding to the original cut score (e.g., RS=42; page 22) as well as the proposed new cut score (e.g., RS=41; page 21). The discussion would have focused on whether the item on page 21 would be a better indicator of the threshold for Level 3 than the item on page 22.

If this had been a modified body of work procedure, members of the VAC would have been directed to the work samples with raw scores of 41 and 42 and asked to determine which would be a better example of where Level 3 starts. In each instance, panelists would be asked to defend their responses in terms of the PLD for Level 3. For this motion, the vote column indicates 19 panelists in favor of the change and 2 opposed. The motion passed, and the cut score was changed from a raw score of 42 to a raw score of 41. The facilitator changed the cut score in the spreadsheet and put the graphic back on the screen with the new cut score and resulting changes in impact.

After the vote on this cut score, Albert G. recommended moving the grade 8 Level 2 cut score from 27 to 26, and the facilitator recorded the motion and the time it was made. After a moment, no one had seconded Albert's motion. The facilitator then asked if anyone would second the motion, and no one responded. The motion therefore failed for lack of a second. The facilitator then recorded this motion and result/action.

A moment later, Monica F. offered a motion that the cut score for grade 4 Level 3 be raised from 49 to 51, and Gerald P. seconded the motion. The facilitator recorded the motion, time, and second and directed the members of the VAC to examine the items or work samples in the

same manner as before with respect to grade 7 Level 3. In this case, after some debate, the facilitator took the vote: 12 in favor, 9 opposed. In the orientation session, the facilitator had pointed out that a 2/3 majority would be required for any motion to pass. The rationale is that any change the VAC makes is essentially an override of a decision made by another panel and should meet a higher standard than a simple majority. We have used the 2/3 rule in all VAC work, based on recommendations in *Robert's Rules of Order Newly Revised* (11th Edition) and the rules of the U.S. Congress in overriding a presidential veto.

Finally, Sam B. moved adoption of all remaining cut scores plus the one change made previously in the afternoon. Hubert L. seconded. The facilitator noted the motion, time, and second and opened the floor for discussion. Hearing none, the facilitator called for the vote: 20 in favor, 1 opposed. The motion passed and was duly recorded. At this point, vertical articulation was over. The facilitator asked VAC members to complete the final evaluation and thanked them for their participation. Upon receipt of the evaluations, the facilitator formally ended the session.

This last motion is a very important one. Later in the process of approving cut scores, policymakers will need to know that the VAC considered every cut score, not just a few. By making and passing this last motion, the VAC certifies that it has considered all cut scores and modified one and approved all others.

Management Activities

The lead facilitator and/or project manager should be able to visit each breakout room throughout the day each day. Therefore, they should not also facilitate a panel. After the opening plenary session, the lead facilitator or project manager should spend a few minutes with sponsoring agency staff to make sure their expectations for that session were met and to address any concerns they have for the rest of the day. Then they should begin making rounds to the breakout rooms. To avoid repeating “lead facilitator and/or project manager” several times, I use “workshop leader” throughout the remainder of this section.

Making rounds. Each day of standard setting, the workshop leader should visit each room once or twice. The purpose of these visits will be to make sure all facilitators are following their scripts and to answer any questions the facilitators have not been able to answer. There are not likely to be very many on the first day (other than perhaps logistical ones related to guest rooms, lighting, temperature, and the like), but they will come up over the course of standard setting. The workshop leader can answer these questions as well as those related to procedure. However, any question related to policy should be escalated to agency staff. It is also important that all panels hear the same thing, so if something comes up in one room, it should be shared with other rooms if it is relevant. For example, in a bookmark procedure, it is appropriate to share across rooms questions and answers about how to set a bookmark. An answer to a question about the grade 6 Level 3 PLD would not need to be shared across rooms. A brief checklist for making the rounds would include the following:

- *Is the room set up the way the plan showed it to be set up?* If not, don't try to rearrange the room while activity is underway; wait until a break, particularly if the room arrangement is supposed to be different for the next activity.
- *Is the facilitator following the script?* Different facilitators may work at different paces, but they should all be following their assigned scripts. Be familiar with these scripts and carry copies with you, either on paper or electronically. Of critical importance is noting whether or not each facilitator has asked panelists to complete the appropriate section of the readiness form and that each panelist has complied.
- *Is the facilitator connecting with the panelists?* This will probably be difficult to discern the first morning when they are still getting to know one another. By the end of the first full day, facilitators and panels should have made a solid connection. Panelists should be paying attention to the facilitator, and the facilitator should be hearing their comments and concerns and responding appropriately.
- *Are panelists engaged and on task?* Answering this question satisfactorily will require several minutes of observation. If you observe during a discussion session, are all panelists participating? Are one or two monopolizing the conversation? Is the discussion staying on topic? Are the comments focused on the PLDs and test items or work samples? If you observe during a rating session, are panelists working alone or with others at their table or in their small group? Are they working through the materials in accordance with instruction? Are they progressing, stuck, doing something other than the assigned task?

In a very large standard setting with a dozen or more panels meeting simultaneously, it may not be possible to visit each panel twice each day and do all the other things that need to be done. The first day is particularly important, and each panel should get at least one visit from the lead facilitator. On subsequent days, the workshop leader should still visit each panel at least once, but the visits may be shorter than on the first day. For smaller events, with seven or fewer panels, it should be possible to conduct two meaningful visits with each panel each day.

At most standard settings, there will also be sponsoring agency staff visiting each panel. For some standard settings, there will be an external reviewer responsible for observing and writing a formal evaluation report on the activity. Keep up with who is in which room, and try to avoid having multiple observers in one room at the same time. If you peek into a room and see someone else observing, move on to the next room, and come back to this one later.

Daily debriefing. At the end of each day, and more often if necessary, the workshop leader, sponsoring agency staff, facilitators, and official external observers/reviewers should meet for 30-45 minutes to review the events of the day and discuss plans for the next day. The following is a brief outline of how these sessions usually go:

- The workshop leader thanks everyone for coming and outlines the purpose and agenda, starting with comments from the sponsoring agency.

- Sponsoring agency comment on what they have seen and voice any concerns they have. If they have noticed anything particularly commendable, they should mention that as well.
- The external reviewer should have time to report his or her observations and offer comments.
- Facilitators may also offer comments or concerns, but it is not necessary for each one to give a full report of the day's activities. Most of the time, one or two will offer an observation, and the rest will acknowledge that they had the same or a similar experience.
- Finally, the workshop leader should summarize the positives and negatives of the day, provide an overview of the next day's activities, ask for final comments, and, if there are none, dismiss the group.

The project manager should assign someone to take notes of this debriefing, type them up before dinner, and e-mail them to facilitators, agency staff, and the external reviewer before breakfast.

If there were concerns with one or more panels, the lead facilitator should meet with the facilitators involved and address that concern before the next session begins. If the concerns were about logistics or the meeting site, the project manager or on-site coordinator should contact staff of the meeting site to resolve those issues before the next session begins. It is also important for the lead facilitator to solicit feedback from the external reviewer and agency staff during the course of the activity and address any concerns as they arise.

Care and feeding. The evaluation forms in [Appendix C](#) contain specific references to meeting room conditions and meals. These are important. While first-class accommodations and gourmet meals will not guarantee a desirable outcome, uncomfortable accommodations and low-budget meals almost guarantee an undesirable outcome. Make sure panelists are comfortable and reasonably well fed. Meeting rooms should be large enough, and work areas should be laid out in a way that allows everyone to work comfortably and efficiently. For onsite, online activities, panelists should not have to struggle with screen resolution or clarity, connectivity, or navigation. In short, you are responsible for the care and feeding of panelists for the duration. Make sure everyone is well rested and eager to start each morning and still engaged at the end of each day.

Coordination. As if you don't have enough to do already, spend some time each day with the person designated by the sponsoring agency as the person in charge. Don't wait until the daily debriefing to discuss an issue that arises. If it needs to be addressed immediately, discuss it with the agency lead and resolve it. If it becomes necessary to shorten or lengthen a session or modify an approach, do so only after consultation with the sponsoring agency lead. If there is an official external reviewer, notify that person as well. It is not necessary, and generally not appropriate, to involve the external reviewer in the decision, but it is vital that he or she be informed so that there is not a glaring variance between what was expected and what was observed.

Part 4: Follow-Up

Standard setting is now over, but there is still a lot of work to do. Keep in mind the primary goal: **Get Approved Cut Scores**. The remaining tasks are to write a final report and then present cut scores to the appropriate authorities for approval.

Final Report

Ideally, the final report should simply be the plan with all verb tenses changed from future to past, plus results. Where the plan said, “We will do A, B, and C,” the final report will say, “We did A, B, and C, and here are the results.” Thus, a well-written plan can become a well-written final report. More importantly, the plan has already been approved, so by being able to report that you followed the plan, you should be half-way home with regard to getting approved cut scores.

Components of the final report. The primary focus of any report is the final set of cut scores and their impact. How these results are presented is critical. We have found the following general outline very helpful.

- *Title Page* – Even for a relatively brief report, a title page is important, not only for initial presentation but for later reference when stakeholders are trying to find the report.
- *Table of Contents* – Particularly if the report is lengthy, a table of contents is a tremendous help to the reader. It should list each major heading and associated page number as well as a list of tables and figures and their page numbers.
- *Executive Summary* – This is what the real decision makers will focus on, so make sure it tells the whole story in a concise and understandable way. It should even be possible to prepare an executive summary with a brief overview of the procedure and results prior to leaving standard setting. Five or six pages should suffice. We have followed this practice for years and have found that agency staff and officials really appreciate it. The key content of the executive summary is the final results plus impact. That report can be prepared almost entirely ahead of time, with empty data tables to be filled in after the final round. Moments after the final round, the author of the report enters those numbers and hands or e-mails the report to the sponsoring agency’s designated recipient. *Note: Don’t lead with the results. Many people will stop there and not understand where those results came from or what they mean.*
- *Introduction* – This should be a reiteration or summary of the introduction included in the plan.
- *Methodology* – This should also be a reiteration or summary of the method section in the plan.
- *Round-by-Round Description* – Here, the report will likely be more detailed than the plan.
- *Results* – Results from each round plus final impact, by pre-identified subgroup, go here.

- *Recommendations* – Strongly recommend that the cut scores be accepted as they are, without modification
- *References* – The introduction and methodology sections will likely reference previous reports or technical guidelines. List each one with a proper citation.
- *Appendices* – These will include the plan, forms, scripts, data tables and analyses, and reference to PowerPoint or other media-based presentations included under separate cover. The sponsoring agency should receive a copy of any and all presentations in their original formats (e.g., PowerPoint, Excel) rather than .pdf.

Rather than include a complete final report in an appendix to this manual (they can run hundreds of pages, with appendices), I have placed immediately after the References section [links](#) to four exemplary final reports. I have included them because they cover just about anything that would ever need to be covered (and more), and they are available online.

While the executive summary should be ready to hand off at the close of standard setting, the full report will take longer – but not much longer. A week to two weeks after completion of standard setting is fairly standard for a final report deadline. Even for an extensive report, one or two weeks should be plenty of time. The key is to have nearly all of the appendices prepared even before standard setting begins. The only pieces missing will be data collected and processed onsite – round-by-round judgments, readiness data, and evaluation form responses. Even those can be processed on a daily basis. Round 1 data analyses should be ready before Round 2 begins, so those data can be dropped right into the appropriate appendix as soon as they are reviewed and approved. For computer-based standard setting, all data processed on site can flow directly to the report's author, whether on site or not. Advance planning makes short work of final reports.

Review and Approval

While you are completing the full final report, agency staff can be reviewing the executive summary and lining up the people who have the authority to approve the cut scores, as well as others who will need to advise them. Recall that you identified all these people when you were planning the standard setting activity. Just a reminder – several months or even a couple of years may have passed since you made those lists; keep up with developments in the agency, and update your lists as stakeholders come and go. There's nothing like an election between planning and execution to make standard setting more exciting than it already is.

Confer with the agency lead about his or her plans for navigating the review and approval process. If there is a report from the external reviewer, ask for a copy, and read it before you submit your report, if that is possible. Certainly read it before you appear before any authorizing body if they will have read it. If you are to appear at any meeting to support agency staff and leadership, find out exactly when and where that meeting will be and what your role will be. Prepare. Rehearse in front of three or four other people, and get them to ask you questions about your presentation. If they ask you a valid question you can't answer, prepare some more. Write out your main talking points, based on the final report, and commit them to memory.

If you are to meet with agency staff or others at the agency, allow plenty of time to get there. If the meeting is in a different state, plan to go the day before and stay overnight as close to the meeting site as possible. Arrive early so you can confer with the agency lead and/or staff before going into the lions' den. Go over the order of presentation and who is responsible for what.

Once the final meeting is set for the governing body, prepare once again. Know in advance who will be there and what kinds of questions they are likely to ask you. If permitted to make an opening presentation, prepare it in conjunction with the agency lead, and rehearse the whole presentation, not just your part. If any of the cut scores have been modified by the agency in response to input from individuals or groups other than the standard setting panels, know exactly what those changes are and how and why they were made. If there are questions about those changes, don't be shy about turning to the agency lead to respond. However, if you are presenting solo, know the answer in advance, and give it as if it is yours. Do not throw the agency under the bus by saying the changes were made after you exited the scene. That response serves no one.

Make your presentation, and recommend that the cut scores be adopted as presented. If anyone asks you if you believe these are valid cut scores, don't hesitate; answer "Yes" immediately. If pressed, you may need to point out that deciding how much an examinee knows in order to be considered proficient or advanced or minimally qualified is a matter for elected or appointed officials to decide with the help and advice of content experts. Your job was to make sure those offering advice did so in accordance with best standard setting practices. This may be a good time to remind the body that you followed to the letter the plan they approved several months ago (if they did, in fact, approve the plan).

Answer any and all questions directed to you. Feel free to confer with the agency lead or whoever is there with you to make the presentation before answering particularly complicated questions. Remember that your contribution is technical, not policy. Defer all policy questions to the agency unless, of course, you are presenting solo. Understand that under those circumstances, it is perfectly acceptable to respond to policy questions by saying, "It is my understanding that..."

Be familiar with any law or regulation impinging on standard setting, and feel free to quote it. For example, "Why did you think it necessary to set three cut scores for every grade for every test? We only have to report who's proficient and who's not." The answer to that question can be found in federal law: The Every Student Succeeds Act of 2015 (PL 114-95; Section 1111). Answers to other "Why?" questions can be found in state law or regulation. Know them. Don't get cocky, though; this very group may have written those laws or regulations.

At the end of your presentation and after you have answered all questions put to you, you may be dismissed, or you may be asked to stay. Either way, thank the group for allowing you the opportunity to present the results of standard setting, and make yourself available for any additional questions they may have.

You may have a chance to hear the discussion, or you may not. If not, at least make arrangements with the agency to find out whether the governing body approved all the cut scores or not. If they changed any or all, find out which ones they changed. If possible, find out

why. At some point, the governing body will make a choice. With any luck, that choice will be to accept your recommendations. However, they know that they have an obligation to approve something by a fixed date, so they will approve something. At that point, you have met your primary objective, but there are still a few loose ends to tie up.

Other Activities

There are five more major activities to complete and several minor ones. These are detailed below.

Develop or update a score scale. Cut scores may have been established in terms of raw scores or ability (theta) scores. Such scores are rarely useful as reporting metrics. For a new test, a new score scale will be needed that reflects the cut scores in some meaningful way. There are several ways to derive these scores. An Excel spreadsheet (Scaling.xlsx) listed in [Appendix E](#) under Other Files contains examples of each of the scenarios outlined below.

- *Single test, one cut score* (e.g., certification/licensure test), cut score based on raw score
- *Single test, two or more raw cut scores* (e.g., academic test with four levels and three cut scores)
- *Single test, two or more cut scores, theta values available*
- Multiple grades, two or more cuts, IRT

In all instances of scaling, make sure the algorithm you derive does not produce negative scale scores. For example, if the difference between the lower and higher cut scores is too large, then scores at the bottom of the range could easily translate to negative scale values, and scores at the top end of the range could greatly exceed desirable limits (e.g., scale scores above 1000 when the score report program is set up to report 3-digit scores). If at first you don't succeed, try, try again.

Revise the final report. The report you wrote a few weeks ago is missing one key piece – the final cut scores as approved by the governing body. Even if the governing body accepted all cut scores as they were presented, you will need to add a narrative about the review and approval process and reiterate the cut scores and impact. If any changes have been made, highlight those and explain when and how they were made. Provide any rationale you have been given for those changes, since you probably were not privy to the discussions preceding them. Submit the revised report to the agency, and be prepared to edit it one or more times. Since this will be the official report for the public to see, it may get more attention and polishing than the one you submitted earlier. Be prepared to spend some time on this task.

Forward the final cut scores. Cut scores were set for a reason. People will take tests and be classified according to them. The final cut scores need to find their way into the hands of the people producing the score reports. The who and how of that process need to be worked out well in advance so that the cut scores can be dropped into the score report programs without error or delay.

Upload cut scores into the score reporting programs. Run sample reports, using enough records to generate a set of scores that spans the range from low to high. Check a sample of reports to make sure that the proficiency level associated with each score is correct. Take nothing for granted.

Review/revise PLDs. If cut scores were adjusted during the final review/approval process, do they still match the PLDs? Now is the time to examine them closely. Specifically, a group of content experts (a combination of yours and those of the sponsoring agency) should return to the test booklets or work samples on which the cut scores were set by panelists and consider whether or not the final, approved cut scores still fall within the range of cuts provided by panelists and whether they still match the PLDs. Since the cut scores have been officially approved, they should not be changed. Therefore, if anything gets changed, it will be the PLDs.

Now that officially approved cut scores are in hand, it is time to finalize the reporting PLDs. These may have been drafted some time ago. However, now that there are *bona fide* cut scores to stand behind them, it is definitely a good time to review them one more time before generating score reports. As with all previous sets of PLDs, editing and approving these PLDs will be a joint effort between your staff and that of the sponsoring agency, with the sponsoring agency having final approval authority.

Settle up. Panelists who participated in standard setting probably have submitted expense claims. If they have not already been paid, pay them now. But don't stop there. Send each one a thank-you note, either with their check or separately. Do the same with the meeting site. Make sure that bill gets paid promptly; you may want to go there again. However, if there were serious issues with the facility, the food, or anything else that detracted from standard setting, address those issues in a follow-up letter to the management. Include panelist feedback. Be professional about it.

Finally, thank and congratulate your staff for a job well done. Recognize them publicly if possible. Share feedback from panelists with them one by one. If there are concerns in the evaluation responses, schedule time with the individuals in question (again, one by one) to review them, get their side of the story, and initiate corrective action if necessary.

Miscellaneous. There will be a host of other activities that may or may not be relevant to your particular situation:

- *Who will be responsible for posting the results on the agency website?* That's usually the agency unless they ask you to do it. If a few months go by, and you don't see it on their site, ask about it.
- *Will you write a paper for a professional meeting or to submit to a journal?* If you have done something new and different, by all means ask the agency lead for permission to do so. You may find a new writing or presenting partner. Never presume to publish or present any client data without written permission.
- *Did you learn anything?* We learn something new every time we do standard setting, usually something totally unrelated to the methodology. Mostly, we learn more about managing standard setting and interacting with all the people involved. As soon as

possible, write down lessons learned. Ask the facilitators and other staff to contribute to the document. Keep it handy for the next standard setting plan.

References

- Angoff, W. H. (1971). Scales, norms and equivalent scores. In R. L. Thorndike, (Ed.), *Educational Measurement* (2nd Ed.). Washington, DC: American Council on Education.
- Beuk, C. H. (1984). A method for reaching a compromise between absolute and relative standards in examinations. *Journal of Educational Measurement*, 21, 147-152.
- Cizek, G. J. and Bunch, M.B. (2007). *Standard Setting: A Guide to Establishing and Evaluating Performance Standards on Tests*. Thousand Oaks, CA: Sage.
- Egan, K. L., Schneider, M. C., and Ferarra, S. (2012). Performance level descriptors: History, practice, and a proposed framework. In G. J. Cizek (Ed.), *Setting Performance Standards: Foundations, Methods, and Innovations*, (2nd Ed.). New York: Routledge.
- Hofstee, W. K. B. (1983). The case for compromise in educational selection and grading. In S. B. Anderson and J. S. Helmick (Eds.) *On Educational Testing* (pp. 109-127). San Francisco: Jossey-Bass.
- Kingston, N. M. and Tiemann, G. C. (2012). Setting performance standards on complex assessments: The body of work method. In G. J. Cizek (Ed.), *Setting Performance Standards: Foundations, Methods, and Innovations*, (2nd Ed.). New York: Routledge.
- Phillips, G. W. (2012). The benchmark method of standard setting. In G. J. Cizek (Ed.), *Setting Performance Standards: Foundations, Methods, and Innovations*. New York: Routledge.
- Robert, H. M. (2011). *Robert's Rules of Order Newly Revised*, 11th ed. Cambridge, MA: Da Capo Press.
- Smarter Balanced Assessment Consortium (2015). Achievement Level Setting Final Report (with appendices). Olympia, WA: Author. Downloaded 01/27/20 from: <https://portal.smarterbalanced.org/library/en/achievement-level-setting-final-report-with-appendix.pdf>
- United States. (2015). Every student succeeds act of 2015: H. R. S1177, 114th Cong., 1st sess., Public law 114-95. Reports, bills, debate and act. [Washington]: [U.S. Govt. Print. Off.].
- Wyse, A. E., Bunch, M. B., Deville, C. & Viger, S. G. (2014). A body of work standard-setting method with construct maps. *Educational and Psychological Measurement*, 74(2) 236–262.

Links to Key Standard Setting Reports

[If links do not work immediately, cut and paste it into your browser.]

American Diploma Project Standard Setting Report

American Diploma Project Algebra II End-of-Course Exam Standard Setting Briefing Book, written by staff of Pearson and Achieve. This 247-page report from 2009 documents one of the first applications of the briefing book (now evidence-centered) method. It is extremely well done.

http://images.pearsonassessments.com/images/tmrs/tmrs_rg/AmericanDiplomaProjectAlgebraII.pdf

PARCC Standard Setting Report

Performance Level Setting Technical Report, written by Laurie Davis and Eric Moyer (2015) and hosted by ERIC Document Reproduction Service. This 212-page report describes in detail the planning, conduct, and reporting of standard setting conducted in the summer of 2015.

<https://files.eric.ed.gov/fulltext/ED599257.pdf>

Smarter Balanced Standard Setting Report

This report describes in detail the preparations for, conduct of standard setting and activities following standard setting to make cut scores available to Smarter Balanced states. The report has several appendices with sample materials and procedures that can be adapted for use in almost any standard setting. Two areas in particular – recruitment and vertical articulation – are highlighted below. [Very large file; it may take a while to download.]

<https://portal.smarterbalanced.org/library/en/achievement-level-setting-final-report-with-appendix.pdf>

Panelist recruitment. Panelist recruitment is described in Section 5, starting on page 43.

Appendix D (p. 542) contains several documents used in the panelist recruiting process for that standard setting (see below). OP stands for Online Panel, TIC stands for Teacher Involvement Coordinator, HE stands for Higher Education, and IPP stands for In Person Panel.

Online Panel

• D01_OP K-12 Leads Online Panel Recruitment	543
• D02_OP K-12 Leads Recruitment Information	557
• D03_OP TIC Information	559
• D04_OP HE Lead Information	561
• D05_OP Educator FAQ	563
• D06_OP He Leads E-mail	567

• D07_OP K-12 Educator E-mail	568
• D08_OP HE recruitment Reminder	569
• D09_OP K-12 Recruitment Reminder	570
• D10_OP Outreach E-mail Newsletter Higher Education	571
• D11_OP Outreach E-mail Newsletter K-12	573
• D12_OP Parent Outreach E-mail	575

In-Person Workshop/Cross-Grade Review Panel

• D13_IPP HE Leads E-mail	577
• D14_IPP K-12 TIC E-mail	579
• D15_IPP HE Leads Information	581
• D16_IPP K-12 Information Packet	586
• D17_IPP HE FAQ	594
• D18_IPP K12 Educator FAQ	597
• D19_IPP HE Educator E-mail	601
• D20_IPP K-12 Educator E-mail	603
• D21_IPP HE Nomination List Template	605
• D22_IPP Parent-Community E-mail	606
• D23_IPP K-12 Participant Application	608
• D24_IPP State Nomination List Template	609

Facilitator Scripts	678-734
----------------------------------	----------------

Vertical Articulation.....	925
-----------------------------------	------------

G01_Cross-Grade Review Agenda	926
G02_Cross-Grade Review (PowerPoint Presentation)	927
G03_Cross-Grade Review Results	952

PowerPoint Presentations

PowerPoint Presentation to State Chiefs.....	964-1009
--	----------

Texas STAAR Standard Setting Report

The Modified Briefing Book Standard Setting Method: Using Validity Data as a Basis for Setting Cut Scores. Paper presented at AERA/NCME in 2010, written by Julie A. Miles, Jennifer N. Beimers, and Walter D. Way of Pearson Assessment & Information

http://images.pearsonassessments.com/images/tmrs/tmrs_rg/11_ncme_usingvalidityevidence_setstds_42210.pdf

This paper documents the use of the use of a briefing book standard setting procedure in Texas. It is an excellent example of a briefer form of report for a professional audience (AERA and NCME members).

Appendix A: Sample Panelist Selection Letters and Forms

Sample	Page
Sample Teacher Interest Letter	64
Application to Participate in Standard Setting Activities	65
Sample Acceptance Letter and Nominee Acceptance Form	67
Sample Backup Letter and Response Form	69
Sample Rejection Letter	71
Sample Thank-You E-mail to VAC Panelists	72

For additional sample letters, see Cizek & Bunch (2007), Chapter 13, pp. 226-230. Those letters are also available online at <https://us.sagepub.com/en-us/nam/standard-setting/book227536#preview>



Disciplines Products F

[Chapter 13 - Materials 1, letters \(pdf\)](#)

Sample Teacher Interest Letter

Southeastern Assessment Consortium
State Office Building
1 Tower Drive
Atlanta, Georgia, 30339

October 18, 2019

Dear Colleague:

The Southeastern Assessment Consortium (SEAC) has created a new generation of tests to administer to students in six southeastern states in grades 3-8 and 11, in accordance with federal law. This spring, we will administer those tests for the first time. Before we can report scores on those tests, we must set performance standards. The process will involve teachers, parents, business and community leaders, and representatives of higher education. I am writing to solicit your interest in participating in a process to recommend performance standards on one of those tests.

There will be two events:

- Performance Level Descriptor development on April 22, 2020 (in Columbia, SC)
- Standard Setting on July 21-24, 2020 (in Atlanta, GA)

Descriptions of both events are described in the enclosures. You may apply for one or both by completing the enclosed form.

Please respond by December 2, 2019, using the enclosed postage-paid envelope, to confirm your interest in participating in one or both of these events. If you are selected to participate, our contractor, Measurement Incorporated, will send you a link to an online registration form to complete by January 31, 2020.

SEAC will pay for travel, lodging, meals, and other allowable expenses for all participants.

Sincerely,

Amanda Havenhold, Ph.D.
Assistant Commissioner

Application to Participate in Standard Setting Activities

First Name:	
Last Name:	
Primary Email Address [Work]:	
Secondary Email Address [Personal – For summer contact]:	
Contact Phone Number:	
Alternate Phone Number where you can be contacted during the summer:	
Sex:	<input type="checkbox"/> Female <input type="checkbox"/> Male
Race:	
<input type="checkbox"/> Native American or Alaska Native	<input type="checkbox"/> Asian
<input type="checkbox"/> Black or African American	<input type="checkbox"/> Native Hawaiian or Other Pacific Islander
<input type="checkbox"/> White	<input type="checkbox"/> More than One Race
Please select the Role that best fits your experience	
<input type="checkbox"/> Educator: General	<input type="checkbox"/> Educator: 2-Year or 4-Year College
<input type="checkbox"/> Educator: Students with disabilities experience	<input type="checkbox"/> Parent
<input type="checkbox"/> Educator: English language learner experience	<input type="checkbox"/> Business or Community Leader
<input type="checkbox"/> Educator: Administration/Non-Teaching	
This section is to be filled out by educators. All others complete the other side of this form.	
Name of district and state where you currently teach	
Name of school where you currently teach	
School Community Type : <input type="checkbox"/> Urban <input type="checkbox"/> Suburban <input type="checkbox"/> Rural	
Content Area(s) Taught or Supported within the last three years: <input type="checkbox"/> ELA <input type="checkbox"/> Math <input type="checkbox"/> Other	
Grade Level(s) Taught or Supported within the last three years:	
Are you certified to teach English language learners? <input type="checkbox"/> Yes <input type="checkbox"/> No	
If yes, have you taught English language learners in the last three years? <input type="checkbox"/> Yes <input type="checkbox"/> No	
Are you certified to teach students with disabilities? <input type="checkbox"/> Yes <input type="checkbox"/> No	

If yes, have you taught students with disabilities in the last three years? <input type="checkbox"/> Yes <input type="checkbox"/> No
This section is to be filled out by parents.
Grade level(s) of your child or children:
Grade level(s) in which you would like to participate
This section is to be filled out by representatives of higher education.
Type of college you represent: <input type="checkbox"/> 2-Year <input type="checkbox"/> 4-Year <input type="checkbox"/> Other: _____
Your area of expertise: <input type="checkbox"/> English <input type="checkbox"/> Languages <input type="checkbox"/> Math <input type="checkbox"/> Other: _____
This section is to be filled out by business or community leaders.
Your role: <input type="checkbox"/> Public Official <input type="checkbox"/> Representative of an Interest Group: _____
<input type="checkbox"/> Other: _____
Grade level(s) in which you would like to participate

Event for which you are applying

☐ Performance Level Descriptor Development (April 22, 2020; Columbia, SC)

☐ Standard Setting (July 21-24, 2020; Atlanta, GA)*

☐ Either

☐ Both

* July 24 will involve only a subset of those participating on July 21-23. Please indicate whether you are interested in participating in this follow-up group. If you are not, you are still eligible to participate on July 21-23.

☐ Yes, I am willing and able to stay through July 24.

☐ No, I cannot participate on July 24.

Please return this form by December 2, 2019 to

Paige Turner
Measurement Incorporated
423 Morris Street
Durham NC 27701
(919) 683-2413 x 549
Fax: (919) 683-9999

Sample Acceptance Letter

[SEAC Letterhead]

March 13, 2020

[Recipient Name and Address]

Dear _____:

I am pleased to inform you that you have been chosen to participate on the Southeastern Assessment Consortium (SEAC) Language Arts Literacy (LAL) standard-setting committee for grade 8. You and your peers will be making judgments involving expectations of student performance based on your experience and expertise, specifically in terms of the *Next Generation Language Arts Content Standards*, and your knowledge of eighth-grade students' skills and abilities. District superintendents were asked to nominate outstanding teachers to serve on a panel that is representative of experienced public school teachers across the region. Thank you for agreeing to work with the 2020 SEAC language arts standard-setting committee.

The standard-setting committee will meet from Tuesday, July 21 to Thursday, July 23, 2020. You have also been selected to participate in a group that will meet on Friday, July 24, to review recommendations across all grades. Tuesday's session will begin at 9 a.m. and continue until 4:30 p.m.; on Wednesday through Friday, the sessions will run from 8:30 a.m. to 4:30 p.m. Approximately 30 educators from across the region will be involved in activities that will help determine the SEAC cut scores in language arts. These activities, including an orientation and training session, will take place at the Buckhead Holiday Inn in Atlanta (directions enclosed). Information available from this process will be used to develop recommendations for the proficiency-level standards, which will be presented to the SEAC commissioners for consideration.

The staff of Measurement Incorporated will be responsible for the logistics for this activity. Please complete and return the enclosed nominee acceptance form no later than Friday, **June 5, 2020**. You may fax the form to Paige Turner at (919) 683-9999 if you choose to assist us. It is imperative that you participate in the process in its entirety. Therefore, you must be able to attend every meeting day in order to participate. I appreciate your commitment to further enhance New Jersey's commitment to educational excellence for our students. I hope that you are looking forward to working on this exciting process.

Sincerely,

Amanda Havenhold, Ph.D.
Assistant Commissioner

SEAC
Standard Setting
Nominee Acceptance Form

Name: _____

State _____

District: _____

School _____

____ **YES**, I will be able to attend the SEAC standard-setting committee meeting on July 21-24, 2020

____ **NO**, I will be unable to attend the SEAC standard-setting committee meeting on July 21-24, 2020

Please return this form via fax no later than Friday, June 5, 2020 to

Paige Turner: (919) 683-9999.

Sample Backup Letter and Response Form

[SEAC Letterhead]

March 13, 2020

[Recipient Name and Address]

Dear _____:

I am pleased to inform you that you have been chosen to an **alternate** for the Southeastern Assessment Consortium (SEAC) Language Arts Literacy (LAL) standard-setting committee for grade 8. Although all panels have been filled, there will typically be one or two members who ultimately are unable to participate, due to unforeseen circumstances. The role of alternates is to fill in for those panelists.

If you are able to serve in this capacity and are available to be in Atlanta on July 21-24, please indicate your willingness to do so by returning the enclosed acceptance form. In the event that we do have a vacancy, we will contact you no later than Monday, July 20.

The standard-setting committee will meet from Tuesday, July 21 to Thursday, July 23, 2020. Standard setting will begin at 9 a.m. on Tuesday, July 21, and continue until 4:30 p.m.; on Wednesday through Friday, the sessions will run from 8:30 a.m. to 4:30 p.m. The meeting, including an orientation and training session, will take place at the Buckhead Holiday Inn in Atlanta (directions enclosed). Information available from this process will be used to develop recommendations for the proficiency-level standards, which will be presented to the SEAC commissioners for consideration.

The staff of Measurement Incorporated will be responsible for the logistics for this activity. Please complete and return the enclosed Alternate Panelist acceptance form no later than Friday, **June 5, 2020**. You may fax the form to Paige Turner at (919) 683-9999 if you choose to assist us. It is imperative that you participate in the process in its entirety. Therefore, you must be able to attend every meeting day in order to participate. I appreciate your commitment to further enhance New Jersey's commitment to educational excellence for our students. I hope that you are looking forward to working on this exciting process.

Sincerely,

Amanda Havenhold, Ph.D.
Assistant Commissioner

SEAC
Standard Setting
Alternate Panelist Acceptance Form

Name: _____

State _____

District: _____

School _____

____ **YES**, I will be able to attend the SEAC standard-setting committee meeting on July 21-24, 2020

____ **NO**, I will be unable to attend the SEAC standard-setting committee meeting on July 21-24, 2020

Please return this form via fax no later than Friday, June 5, 2020 to

Paige Turner: (919) 683-9999.

Sample Rejection Letter

[SEAC Letterhead]

March 20, 2020

[Recipient Name and Address]

Dear _____:

Thank you for your interest in participating in the 2020 SEAC standard setting activities. We received nominations from many districts across the region and, as a result, the number of qualified individuals nominated has exceeded the number of individuals needed to fill the committees.

The selection of participants was performed in a manner to ensure that the standard setting committees would be comprised of expert educators demographically representative of schools and districts in this region. While I regret to inform you that you were not chosen for the Performance Level Descriptor Panel or Standard Setting Panel, we would like to keep your materials on file in the event that new committee openings develop. Standard setting is one of several assessment processes for which input from educators is crucial to ensuring the validity of the results.

We are grateful that your district remains supportive of individuals willing to assist us in the implementation of the statewide assessment programs.

Sincerely,

Amanda Havenhold, Ph.D.
Assistant Commissioner

Sample Thank-You E-mail to VAC Panelists

To: VAC Panelists
CC: Amanda Havenhold
From: pturner@measinc.com
Subject: Thank You

Dear VAC Panelists:

Thank you for your participation in the Southeastern Assessment Consortium Standard Setting Panel. We are particularly appreciative of the extra contribution of those who stayed for a fourth day to serve on the Vertical Articulation Committee. Your time and diligent efforts were critical to this essential task of recommending achievement levels for SEAC assessments.

We have received many requests for several of the resources that were used in Atlanta. Attached you will find the PowerPoint presentation on the bookmark procedure; a list of links to helpful resources on the SEAC website; and Performance Level Descriptors (PLDs) for all grades and content areas. Please note that the attached PLDs are also available on the SEAC website; we simply reformatted them for ease of use at the standard setting.

If you have any questions, please contact Paige Turner at pturner@measinc.com.

Thank you again for your participation and extraordinary contributions!

Attachments:

SEAC Standard Setting OIB and Bookmark Procedure Presentation

SEAC Resource Links

Panelist Certificate of Participation

Performance Level Descriptors

Appendix B: Scripts and Work Forms

Sample	Page
Sample Modified/Extended Angoff Facilitator Script	74
Sample Body of Work Facilitator Script	78
Sample Bookmark Facilitator Script	85
Facilitator Notes for Inter-Round Discussion	93
Sample Vertical Articulation Script	97

Sample Modified/Extended Angoff Facilitator Script

9:00 Introduction to the Tests

1. Put name on first inside page of notebook (with # 1-15) [Put ID # later.]
2. Describe test development process
 - a. Developed by MI to content standards
 - b. Reviewed last summer by teachers
 - c. Field tested this spring with students
 - d. Reviewed and selected by Department staff
 - e. Assembled from approved field test items by MI – 6 equivalent forms
 - 4th grade: 10 2-pointers and 4 3-pointers
 - 7th grade: 10 2-pointers and 3 3-pointers
3. Describe notebook organization
 - a. Each item with a tab in order that it appears in final test booklet
 - b. Scoring rubric on first page (explain what a scoring rubric is if necessary)
 - c. Sample responses representing minimal and solid scores: 0, 1, 2, and 3
 - d. Final tab is Performance Level Descriptors – we will get to those next.
4. Take 5-10 minutes to look through the notebook and become generally familiar with it.
 - a. Score is circled at the bottom of the page.
 - b. Some scores are marked “M” for minimal (or maximum if 0)
 - c. Some scores are marked “S” for solid – showing that there is range even for single score points
 - d. General comments and questions about organization of the notebooks, tests, or scoring. [Note: These are the tests, and these are the scoring rubrics. We are not here to change either. If you take issue with an item or a score, we will discuss later a way to address that issue in standard setting.]
5. In the time remaining, go through item by item pointing out how student responses evolve from 0 to 1 to 2 (to 3).
6. Note that we will have two days to look more closely at each item because we will be trying to figure out how students at different proficiency levels will respond to them, one by one.

10:00 Introduction to the PLDs

1. Direct panelists to last tab
2. Purpose: to review, revise (if necessary), and apply these PLDs to student performances
3. Describe role of PLDs in standard setting – to provide a picture of a student who embodies the standard. The stated PLDs describe typical; we are looking for borderline, or “just barely.” Read the PLDs and start thinking about what it would mean to be just barely “Meets” or just barely “Exceeds.”
4. Consider the test items relative to the PLDs:

- a. Is anything missing from the PLDs?
 - b. Is there anything in the PLDs that is not relevant to the test items? [Note that these PLDs have to be relevant to all forms of the test, not just this one; therefore focus on the content standards measured by these items, not just the items themselves.]
 - c. Is there anything in the PLDs that needs to be changed?
5. Start with "Meets"
 - a. Read the entire PLD.
 - b. Ask for comments regarding appropriateness and clarity.
 - c. Write any suggestions for improvement on chart paper
6. Examine "Does Not Meet"
 - a. Same as "Meets"
 - b. What distinguishes "Does Not Meet" from "Meets?"
 - c. What would the borderline between "Does Not Meet" and "Meets" look like? [Write on chart paper.]
 - d. What would "just barely Meets" look like? [Write specific limitations on chart paper.]
7. Examine "Exceeds"
 - a. Same as with "Meets"
 - b. What distinguishes "Exceeds" from "Meets?"
 - c. What would the borderline between "Meets" and "Exceeds" look like?
 - d. What would "just barely Exceeds" look like? [Write specific limitations on chart paper.]
8. Summarize "Just Barely" modifications to definitions.
9. Note all suggestions for changes. Get room consensus on each specific recommendation.

Afternoon: Extended Angoff Practice

1. Select final 3-point item.
2. Ask panelists to study the item and each scored sample.
3. Ask panelists to consider 10 students who "just barely Meet" the standard. Now imagine that those 10 students responded to this item.
 - a. How many 3s do you suppose there would be? Why?
 - b. How many 2s do you suppose there would be? Why?
 - c. How many 1s do you suppose there would be? Why?
 - d. How many 0s do you suppose there would be? Why?
 - e. Given these 0s, 1s, 2s, and 3s, what would be the average score for these 10 students? Write that number down.
 - f. [Write down each panelist's estimate, calculate mean. Discuss range of scores and why some gave low or high scores.]
4. Ask panelists to consider 10 students who "just barely Exceed" the standard. Now imagine that those 10 students responded to this same item.

- a. [a-f as above]
5. Direct panelists to click on their spreadsheet icon. Explain that each person has a different numbered spreadsheet. They should write this number (701-720) on the front page of their notebook next to their name. For the rest of the time, we will distribute materials based on this number.
6. Explain that they will examine each item in this manner and enter two scores for each item. Pull up Tab X01 on the big screen and show how to enter the numbers. Purposely put in a number for one item that is too big. Show what happens. Purposely put in a number for Meets that is larger than the number for Exceeds. Show what happens.
7. Get a sense of panelists' understanding of the process. Ask for questions. Then ask if everyone is ready to begin Round 1. Complete Round 1 of the Readiness Form.

Discussion at the beginning of Round 2

1. Distribute notebooks with Round 1 rating sheets.
2. Start with **Summary** on screen. Show average value for each item so panelists can compare their ratings with room average. Note 2 cut scores. No need to dwell on these because both are shown in more detail under Meets and Exceeds.
3. Go to **Range**. Point out the 2-3 items with the largest Meets ranges, then the 2-3 items with the largest Exceeds ranges. Make a note of these on the board, and then go back to Meets.
4. Show **Meets**. Highlight the item with the largest range. Invite panelists at extremes to comment on their ratings. Then invite panelists closer to the mean to comment. Note different approaches and assumptions. Make sure all comments are couched in PLDs and test content.
5. After looking at specific items, go to *Cut Score* column. Again note extremes and invite panelists to comment on their strategies or assumptions. Try creating a low group (e.g., Mean – 1 SD), a middle group, and a high group (e.g., Mean + 1 SD), rather than putting the most extreme panelists on the spot. Then call on some of the panelists closer to the mean to discuss their strategies and assumptions.
6. Finally, translate the Meets cut score into a % of total points. (After Round 1: 42.5 for grade 4 and 38.0 for grade 7). Does this seem reasonable? Why or why not?
7. Show **Exceeds**. Same as #4-6.
8. Questions – answers.
9. Complete Readiness form for Round 2.
10. Round 2

Discussion at the beginning of Round 3

1. Repeat Steps 1-7 as in Round 2. Between Rounds 2 and 3, we will enter the item means in the Score row in the Exceeds tabs. After Step 7, draw attention to the item field test means. Note that these means are based on field test values and point out motivational

issues and newness of some of the curriculum. Do these scores make sense? Do you have reason to believe they will go up in spring 2009, spring 2010? Do these numbers make you want to change any of your item ratings? Why or why not?

2. Questions – answers
3. Complete Readiness form for Round 3.
4. Before beginning Round 3, thank the group for their work and remind them that when they complete Round 3, they are not quite done. We have an evaluation form to complete, and the Department may have other things to discuss.
5. As panelists complete Round 3, check data as they complete evaluation. Collect evaluation, thank them again, collect secure materials, and dismiss individually.

Sample Body of Work Facilitator Script

Note to the reader: For this application of the body of work procedure, we employed a recent modification in which panelists review an ordered set of student work samples to identify the first one to meet the criteria for each performance level, rather than assign each work sample to a particular level. Please see the article cited below for details.

Wyse, A. E., Bunch, M. B., Deville, C., & Viger, S. G. (2014). A body of work standard-setting method with construct maps. *Educational and Psychological Measurement* 74 (2) 236–262.

Navigation

[Allow 20 minutes; longer if there are many questions.]

Let's open the OPLS software and take a look at the Practice Test. Your entry into the system will look just like this, and your packets will look a lot like this one, but they will be much longer. This Practice Test has only six essays; yours will have 20.

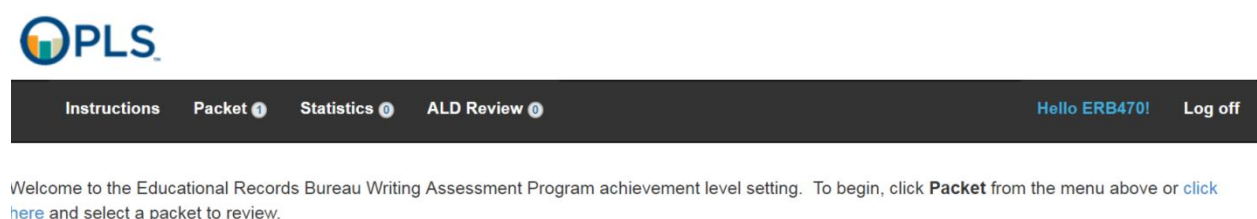
Go to <https://opls.measinc.com/InPersonLogin>

User ID: xxx470

Event Code: xxxxxx2018

Table: 1

This is the opening screen. You will have to answer a couple of security questions to get in, but since I've already done that, we're going straight in.



Once you get to this page, click Packet in the ribbon at the top of the page or "click here" in the text to select a packet to review. That will take you to the Packets screen.

Packets

 View Selected

Name

Practice

When you get to the Packets screen, you will see a list of available packets. You should have six. Here we only have one, because the Practice packet is the only one available right now. When you open your Packets page, you should start with the first one and work your way down, one genre at a time, lower grade first and upper grade second. Complete one and submit your two thresholds before opening the next one. By clicking on the Packet name, we go directly to the Construct Map for that packet.

Sample Reviews


Construct Map for Packet: Practice

7-8

Round: 1

Review each essay in the ordered Response Packet by clicking on the sample number below, starting with sample 1.

Enter Comment

 View Selected

Sample	Threshold	Comments	Reviewed	Genre	
1			<input checked="" type="checkbox"/>	Informative	
2			<input checked="" type="checkbox"/>	Informative	
3			<input type="checkbox"/>	Informative	

This is the construct map I've been telling you about. There is a lot going on here. I will walk you through its features one at a time:

- Sample – Each student essay is called a sample, in keeping with the language of the Body of Work method. You are going to be evaluating these “samples” in order to identify two specific ones – the first one you believe minimally meets the criteria for Level 2, and the

first one you believe minimally meets the criteria for Level 3. For this packet, we are only looking for the threshold for Level 2, but you will have many more “samples” and will need to identify the one marking the threshold for Level 2 and the one marking the threshold for Level 3.

- **Threshold** – In this column, you will enter your two thresholds. Actually, when we click here to go to sample 1, you will see that you can set a threshold right on that page, and the software will enter a mark in this column. Thus, for example, if you identify Sample 7 as the threshold for Level 2, the software will place a mark in this column on the Sample 7 row. I’ll show you what I mean. [Go to Sample 1 and set a threshold and come back.] And you see that the software has entered Level 2 next to Sample 1. Let’s move on to the rest of the construct map before looking at any more samples.
- **Comments** – As you go through your packet, you can use the Comment icon to enter comments about any or all essays. For example, you may have narrowed your Level 2 threshold to Samples 8, 9, and 10. You’ve gone back and forth between each one and the Level 2 PLD, and there’s just too much information to hold in your head. You can write your notes on paper, or you can enter them here. The nice thing about entering them here is that they will be here when you come back together on Wednesday instead of in the trash or in a folder you can’t find.
- **Reviewed** – As you open each Sample, the software enters a mark. That way, if you have to stop for a while and come back later, you can check the construct map to find the last essay you opened. You can pick up right where you left off. Also, if you want to enter a threshold on the construct map and you haven’t reviewed the sample you’re identifying as the threshold, the software won’t let you do that. [Demonstrate.]
- **View Selected** – As with most software, there is more than one way to do some things. If you highlight a sample and then click View Selected, you will go straight to that sample. The same is true for Enter Comment at the top right of the construct map. [Show]

Now let’s go to Sample 1.

Back to Construct Map

Set Threshold Enter Comment Next

Response Prompt ALD

Final Draft

Both of the Teens They want to Help People

Here we see a sample student essay. Note that the Sample number is located right at the top of the page. [Show] Some of these are fairly short, and others are quite long, perhaps up to three pages. [Scroll down to the bottom of the essay.] Some are easy to read, and some are not (like this one). We have used the best print quality we could find for the score points we selected. For some score points, we didn't have a lot to choose from, so there will be some you will have to work with a little bit to get through.

Besides the student response itself, there are more icons and buttons on these pages. Let me show you how each one works.

- **Back to Construct Map** – At any time, you can click here to go back to the Construct Map for any reason.
- **Set Threshold** – When you get to an essay you believe minimally meets the criteria for Level 2 or Level 3, click here to set your threshold. You will get a drop-down box that asks if you want to set a threshold here. You will have a choice to set the Level 2 threshold or Level 3 threshold. Once you select one, you will get another drop-down box [Show] that asks if you are sure. We've made it as difficult as possible to do something by mistake.
- **Enter Comment** – If you click this icon, you get another drop-down box. [Show] If you Save the comment, it will show up on the Construct Map. If you close without saving, the comment will not save, so always remember to click Save.
- **Response** – This icon will always be highlighted because you are on the Response page. If you click out to Prompt or PLD, you will see these same three icons with Response not highlighted. Click it, and you are right back on the Response page.
- **Prompt** – Click here to see the prompt students were writing to. [Show] Since this packet only has six samples, I limited them to a single prompt, the same one I showed you a little while ago. You will recall that this prompt is associated with two passages the student has to read before writing an essay. When we start the practice round, I will give you a quick summary of both passages so we won't have to spend a lot of time reading them. When you get to your Round 1 packets, please take some time to read each prompt thoroughly. In your packets of 20 or so essays, you will probably see essays written to about three different prompts. After you've read each prompt a couple of times, you will probably have them pretty well internalized. Keep in mind that we are primarily interested in the essays, but there will be times when you want to check the prompt if there's an issue about factual accuracy of the essay. And, of course, you can click right back to the Response or just about anywhere else in the packet.
- **PLD** – Click here to see the PLDs for this grade/genre. [Show] We've included all three levels, but you will be concerned only with Levels 2 and 3. In one or two instances, you may want to check the Level 1 PLD if you are really on the fence about whether a particular essay is a minimal Level 2 response or a high Level 1 response. For the first several samples in your packets, you will want to pay close attention to the descriptor for Level 2.
- **Next/Previous** – To get to the next sample or the previous one, you could go back to the Construct map, or you could just click Next or Previous. [Demonstrate]

That's pretty much the grand tour of the software. Let me pause here for questions, and I'll try to answer them. Please use the Chat feature to enter your questions. If it's complicated, raise your hand, and I'll unmute you. [Answer questions and move on to the Practice Round.]

Practice Round

[Allow 45-50 minutes; conclude by 2:30 p.m.]

All right, let's go back to the Construct Map and take a look at these six essays – work samples – and see if we can identify the one that marks the beginning of Level 2. Let me tell you a few things about these six samples first:

- Most of them have low to intermediate scores.
- Given that there are only six of them, we could not represent every score point.
- Therefore, the threshold for Level 2 may actually lie between two samples; i.e., you may see one that is clearly below Level 2, and the next one will be a solid Level 2, and you would really like to see one in between. In that case, mark the one just over the threshold, and we'll talk about it later.

Let's start with Sample 1. [*Click Sample 1.*]

Before we read this or any other sample, let's take a look at the prompt. [*Click*]

To move this along, I've prepared a short synopsis of the plots of the two passages. Let me share that with you:

Redacted – Secure Material

Students read both passages and then composed their own essays. Now, before we go back and look at Sample 01, let's take a look at the Level 2 PLD, so we'll know what we're looking for. Those of you in the Grade 7-8 group have a head start on the others, so I'm going to give you all a few minutes to read just the Level 2 PLD. [*Pause while they read the Level 2 PLD. Scroll down as necessary.*]

In summary, Level 2 requires that a student

- Establish a context with some elaboration

- Provide an introduction that identifies the topic
- Use multiple facts/details from the stimulus
- Use simple and compound sentences
- Use precise, descriptive words
- Use grade-appropriate conventions

So now, let's go back to Sample 01 and apply this PLD. *[Click Response to go to Sample 01. Hold top portion on screen to allow time to read, then scroll down as appropriate.]*

I'm not going to ask you yet what you think of this one. Let's go on to Sample 02. *[Repeat the process, scroll through the sample.]*

And now Sample 03. *[Repeat the process, scroll through the sample.]*

All right; let's take another look at the Level 2 PLD. *[Click back to PLD.]* Has anyone seen anything that looks like it might just barely be in Level 2 or solidly in Level 2? If so, raise your hand, using the Raise Hand function on your screen. *[Count hands.]* Thank you, you can put your hands down now. *[Make sure all hands have been lowered.]*

Let's go on to Sample 04. *[Repeat the process, scroll through the sample.]*

And now, Sample 05. *[Repeat the process, scroll through the sample.]* I'll ask again - Has anyone seen anything that looks like it might just barely be in Level 2 or solidly in Level 2? If so, raise your hand. *[Count hands.]* Thank you, you can put your hands down now.

And now, just for the sake of argument, let's take a look at Sample 06. *[Repeat the process, scroll through the sample.]*

By now, I hope everyone has seen an essay that either marks the threshold for Level 2 or is the first one clearly at Level 2, so I'm going to give everyone an opportunity to tell me what they think. I'm going to open a poll that allows you to vote for the sample you think is the best representative of the threshold of Level 2. I will also briefly display each of the six essays and keep the poll open until everyone has voted. Afterwards, I will tally the votes and show you how we did. I need to tell you that this polling function only allows for up to five choices, so I've lumped Samples 05 and 06 together. If you would like to place your threshold on either of those samples, use the fifth option. *[Open polling and show each essay for about one minute or until everyone has voted.]*

Here are the results of the polling. *[Show results, noting range of samples selected.]* If this were an actual round, we would be starting a discussion now about the range of thresholds you set. But since this is the Practice Round, I can tell you that we have achieved our basic goal: You

looked at some essays, you applied the PLD, and you identified an essay you believed marked the threshold to Level 2. We didn't get marks on all six essays, and that's a good thing. We did have some distribution, and that is to be expected. It looks like you know what you are doing. However, before we turn you loose, I'm going to open the floor one more time for questions. If you have one, use the Chat function or raise your hand, and I will call on you in the order that your hands go up. *[Answer any questions and close the session.]*

That concludes the official portion of the training and practice round. I am going to dismiss in a moment, but first a couple of things:

1. If you log out of this webinar and want to start Round 1, you must first send an e-mail to your facilitator stating that you have completed the training, understand the task, and are ready to start Round 1. When we get that e-mail, we will open your Round. I'll show you their e-mail addresses in just a moment.
2. If you have additional questions, stay online. When everyone else has logged out, we will answer as many questions as we can. Feel free to leave at any time or stay until the end. I will keep the webinar open until 4 P.M. Eastern Time or until there are no more questions, whichever comes first.

Panelists then logged out of the webinar and logged into OPLS to begin Round 1 of their body of work rating exercise. They had a full day to complete the exercise and submit their work in OPLS. A follow-up webinar was scheduled for the next day.

Sample Bookmark Facilitator Script

Morning Day 1 (10:15am -12:00 noon)

Goals: Introduce panelists to the Test Blueprint and the Threshold PLDs.

Materials of Importance: Test blueprint, Threshold PLDs, Non-Disclosure Agreements, Readiness Form, Demographics Form

Facilitator Outline:

1. Assist panelists with table assignment
2. Conduct group introduction (15-30 seconds per person).
3. Have panelists sign non-disclosure agreements and demographics form.
4. Allow panelists to become familiar with the blueprint and threshold PLDs.
5. Lead panelists in a discussion of the blueprint and threshold PLDs.
6. Have panelists complete the threshold PLD portion of the Readiness Form.
7. Dismiss panelists for lunch.

Facilitator Talking Points:

- Identify yourself as the facilitator, along with relevant information about yourself, and ask panelists to identify themselves with their names, districts, and job titles.
- Discuss the purpose of the test blueprints and PLDs. Ensure that panelists are aware that PLDs are simply descriptions of students' abilities at a particular performance level. **Stress** that all decisions concerning cut scores **must be firmly** grounded in the PLDs.
- Ask panelists to read the PLDs carefully and to contemplate what it means to be Level 1, Level 2, Level 3, and Level 4, respectively.
- Encourage panelists to imagine students they have known who might have fit each of the Level descriptors.
- Ask panelists to highlight and underline the differentiating characteristics of each performance level.
- Lead panelists in a room-wide discussion of the differentiating characteristics of each performance level.
- Next, narrow the focus and ask panelists to focus on the **Just Barely** Level 2 students and what differentiates them from the Level 1 students. Next, ask the panelists to focus on the **Just Barely** Level 3 students and what distinguishes them from the Level 2 students' performance. Finally, ask the panelists to focus on the **Just Barely** Level 4 students and how they differ from the Level 3 students.

Afternoon Day 1 (1:00-4:00pm)

Goal: Introduce panelists to the test taking experience.

Materials of Importance: Readiness Form, Chromebooks to access the online assessment, Username and Password Information, Answer Keys, and both hard copy and online Threshold PLDs

Facilitator Outline:

1. Assign/distribute log in information.
2. Assist panelists with logging onto the computers and the assessment site.
3. Briefly explain/demonstrate how to navigate within the assessment instrument.
4. Assist panelists throughout the testing experience.
5. Give panelists an answer key after each unit is completed.
6. Engage in a room-wide discussion of the testing experience at the end of Unit 3 (three).
7. Ask panelists to complete the portion of the Readiness Form pertaining to understanding of the tests.
8. Dismiss panelists for the day.

Facilitator Talking Points:

- Explain to panelists how to navigate within the online testing environment.
- Ensure that panelists are aware that they are taking the test so that they can fully understand test contents and gain a greater understanding of the difficulties students may encounter as they take the test.
- **Reinforce the notion that critiquing the items and test development process are not the panelists' responsibility or purpose.**
- Ask panelists to discuss their impressions of the test content. What did they think would have been easy or difficult for students at Levels 2, 3, and 4, respectively? What types of skills did they notice would be needed to successfully answer the items on the test?

Morning Day 2 (10:15am-12:00 noon)

Goal: Introduce panelists to the OPLS system

Materials of Importance: Chromebooks to access OPLS, Username and Password Information, and Readiness Form

Facilitator Outline:

1. Assign/distribute log in information.

2. Project OPLS on a screen visible to the entire room.
3. Assist panelists with logging onto the computers and OPLS (content specialist).
4. Demonstrate the following tasks:

Finding and selecting the appropriate test (Practice and Operational)

Navigating to the item map

Features of the item map

Going from the item map to a page in the OIB

Returning from a page in the OIB to the item map

Calling up the PLDs from item pages

Calling up stimulus materials from item pages

Entering comments

Entering thresholds

Submitting cut scores

5. Answer panelists' questions.
6. Ask panelists to complete OPLS portion of the Readiness Form.
7. Dismiss panelists for lunch.

Facilitator Talking Points:

- Explain to panelists how to navigate within the OPLS software.
- Answer questions posed by panelists throughout the presentation.
- Ensure that everyone has responded positively to Readiness Form, if not, address concerns.

Afternoon Day 2 (1:00-4:00pm)

Goals: Introduce panelists to the Bookmark procedure by leading them through the practice Ordered-Item Booklet (OIB) on OPLS. Ensure all panelists are prepared to begin the Standard Setting process. Complete Bookmark Round 1.

Materials of Importance: Chromebooks, Username and Password Information, Bookmark Practice OIB in OPLS, Readiness Form, Round 1 OIB in OPLS

Facilitator Outline:

1. Assist panelists to navigate in OPLS to the Practice Round.
2. Instruct panelists to set one cut score (Level 3 only) for the Practice Round.
3. Use the Facilitator Role Function of OPLS to determine where panelist set their cut scores.
4. Discuss Practice Round results.

5. Ask panelists to complete Round 1 portion of the Readiness Form.
6. Have panelists navigate to the Operational OIB in OPLS and begin Round 1.
7. Monitor Round 1 and be available to answer panelists' questions.
8. Ensure that each panelist submits his/her cut scores, upon completion of Round 1.
9. Collect any secure materials and dismiss panelists for the evening.

Facilitator Talking Points:

- Before beginning the Bookmark Practice Round remind panelists of the following:
 - The items in the OIB are ordered from easiest to hardest based on actual student performance on the items.
 - If there is a constructed response item on the assessment, it will appear in the OIB multiple times, once for each score point.
 - They will place their practice bookmark on the first item that the ***Just Barely*** Level 3 students would have a less than 2/3 chance of answering correctly.
- Work through the first two item in the Practice OIB as a group asking the panelists to specifically discuss the following questions:
 - What types of skills and abilities must students possess to correctly answer this item?
 - How do those skills and abilities relate back to the PLDs?
- Ask panelists to complete the Practice OIB. They will place one practice bookmark that differentiates between the Level 2 and Level 3 performance levels.
- Discuss the results of the Practice Round with the group. Focus the discussion on:
 - Range of pages on which panelists set thresholds
 - Rationales and strategies used by panelists
 - Threshold placement issues
 - Questions about key aspects of the threshold PLDs and/or items
 - Navigation issues.
- Explain to panelists that they will engage in three rounds of standard setting, with discussion between rounds. During each round, they will be asked to place three bookmarks, representing the division between Levels 1 and 2; Levels 2 and 3; and Levels 3 and 4, respectively. Their cut scores will be recommendations and not necessarily the final cut score. The State Board will make the final decision.
- Briefly review the characteristics of the ***Just Barely*** Level 2, 3, and 4 students, respectively.
- Before beginning Round 1 remind panelists to consider the following questions for each item as they progress through the Ordered-Item Booklet:
 - What do you know about students who correctly answer this item?

- What makes items progressively more challenging?
- Would the ***Just Barely*** Level 2 students have a 2/3 chance of answering the item correctly?
- After placing the Level 2 bookmark, would the ***Just Barely*** Level 3 students have a 2/3 chance of answering the item correctly and so on for the ***Just Barely*** Level 4?
- Explain to panelists that once they identify an item that they believe the ***Just Barely*** Level 2, 3, or 4 students, as the case may be, have less than a 2/3 chance of answering correctly that they should take a look at the next few items in the Ordered-Item Booklet to confirm that they have reached the best page on which to place their bookmark.
- Remind panelists that all their decisions concerning their placement of bookmarks must be firmly grounded in the PLDs.
- Encourage panelists to consult with the other people sitting at their tables during each round. Reassure them that it is ok to feel a little confused or uncertain about the process at this point.
- Answer any questions the panelists might have about the process and ensure all panelists are prepared to begin Round 1.
- Remind panelists to pace themselves.
- Remind panelists to submit cut scores for all 3 levels; log out; and turn in any secure materials before leaving for the day.

Day 3 Morning (8:30-11:45am)

Goals: Discuss Bookmark Round 1 Results and complete Round 2.

Materials of Importance:: Chromebooks, Username and Password Information, Round 1 Results (bookmark median and distribution), Round 1 OIB in OPLS, Readiness Form, Round 2 OIB in OPLS

Facilitator Outline:

1. Display and discuss Round 1 results.
2. Be prepared to project both the OIB and the results on a screen in order to toggle between the two as necessary.
3. Have panelists complete Round 2 Readiness Form and Log into OPLS.
4. Begin Round 2.
5. Allow panelist to take a break after starting Round 2 as needed.
6. Monitor Round 2 and be available to answer panelists' questions.
7. Ensure that panelists submit all three bookmarks before leaving the room.
8. Dismiss for lunch.

Facilitator Talking Points:

- Show the distribution of Round 1 bookmarks and the median bookmark placement.
- Conduct a room-wide discussion concerning the Round 1 results. Highlight the following topics:
 - What were the challenges panelists faced in Round 1?
 - What factors influenced panelists' decisions in placing their bookmarks?
 - How did the panelists use the PLDs in their decision making process?
 - Group consensus is not necessary.
 - Ask panelists to explain their thought process concerning the placement of their bookmarks in the OIB. Specifically ask panelists from both ends of the spectrum to explain their reasoning.
- Explain to panelists the Round 2 process:
 - Round 2 will be more targeted. Panelists will start Round 2 on the lowest recommended Level 2 bookmark recommended in Round 1. Similarly, the last page in the OIB that they will review for Round 2 will be the highest recommended Level 4 bookmark.
 - Panelists should place three bookmarks using the same process employed in Round 1.
 - Panelists are free to discuss the items and PLDs with their tablemates, but not across tables.
- Before beginning Round 2, reiterate to panelists that they should consider the following questions for each item they examine in the OIB:
 - What do you know about students who correctly answer this item?
 - What makes items progressively more challenging?
 - Would **Just Barely** Level 2 students have a 2/3 chance of answering the item correctly?
 - After placing the Level 2 bookmark, would **Just Barely** Level 3 students have a 2/3 chance of answering the item correctly?
 - Same for Level 4.
- Ask if the panelists have any questions and ensure they are ready to begin Round 2.

Day 3 Afternoon (1:00-4:00pm)

Goals: Review Round 2 results. Ensure all panelists are prepared to begin Round 3.

Materials of Importance: Chromebooks, Username and Password Information, Round 2 Results (bookmark median and distribution), Round 2 OIB in OPLS, Readiness Form, Round 3 OIB in OPLS, Evaluation Forms

Facilitator Outline:

1. Project and discuss Round 2 results.
2. Have panelists log in to OPLS; complete Round 3 Readiness Form; and begin Round 3.
3. Monitor Round 3 and be available to answer panelists' questions.
4. Ensure that panelists submit three bookmarks and evaluation form before being dismissed.

Facilitator Talking Points:

- Project the Round 2 results and the Round 2 OIB.
- Conduct a room-wide discussion concerning the Round 2 results. Highlight the following topics:
 - What were the challenges panelists faced in Round 2?
 - What factors influenced panelists' decisions in placing their bookmarks?
 - How did the panelists use the PLDs in their decision making process?
 - Group consensus is not necessary.
 - Ask panelists to explain their thought process concerning the placement of their bookmarks in the OIB; specifically those panelists who placed bookmarks at each end of the spectrum.
- Review the Round 2 impact data. Highlight the following topics:
 - The data are being presented to the panelists to give them perspective concerning the effect of their ratings.
 - Do the percentages of students in the four performance categories seem realistic?
- Explain to panelists the Round 3 process:
 - Round 3 will be even more targeted. Panelists will start Round 3 on the lowest recommended Level 2 bookmark recommended in Round 2. Similarly, the last page in the OIB that they will review for Round 3 will be the highest recommended Level 4 bookmark from Round 2.
 - Panelists should place three bookmarks using the same process employed in Rounds 1 and 2 with all Bookmarks based on the threshold PLDs, not impact data.
 - Panelists are free to discuss the items and PLDs with their tablemates, but not across tables.
- Before beginning Round 3, once again reiterate to panelists that they should consider the following questions for each item they examine in the OIB:
 - What do you know about students who correctly answer this item?
 - What makes items progressively more challenging?

- Would ***Just Barely*** Level 2 students have a $\frac{2}{3}$ chance of answering the item correctly?
 - After placing the Level 2 bookmark, would ***Just Barely*** Level 3 students have a $\frac{2}{3}$ chance of answering the item correctly?
 - Ask the same for the ***Just Barely*** Level 4 students.
- Ask if the panelists have any questions and ensure they are ready to begin Round 3.
- Remind panelist to complete the evaluation form before leaving.

Facilitator Notes for Inter-Round Discussion

This script accompanies the Body of Work script and is used between Rounds 1 and 2.

Materials Needed

- Round 1 Table/Chart loaded
- PLDs
- List of panelists

Procedure

Read all comments from Round 1 before starting the round. Get a general sense of the major themes of panelist comments plus any correspondence you have had with them between Monday afternoon and Wednesday morning. Review Round 1 output before you start the webinar and have a general sense of the median and range of the cut scores. Note that the Round 1 Table/Chart will reveal individual panelists' thresholds if you hover your mouse over the bar. Try not to do that during the webinar, as that may be embarrassing for them. It's OK to do that as you review, but not during the webinar.

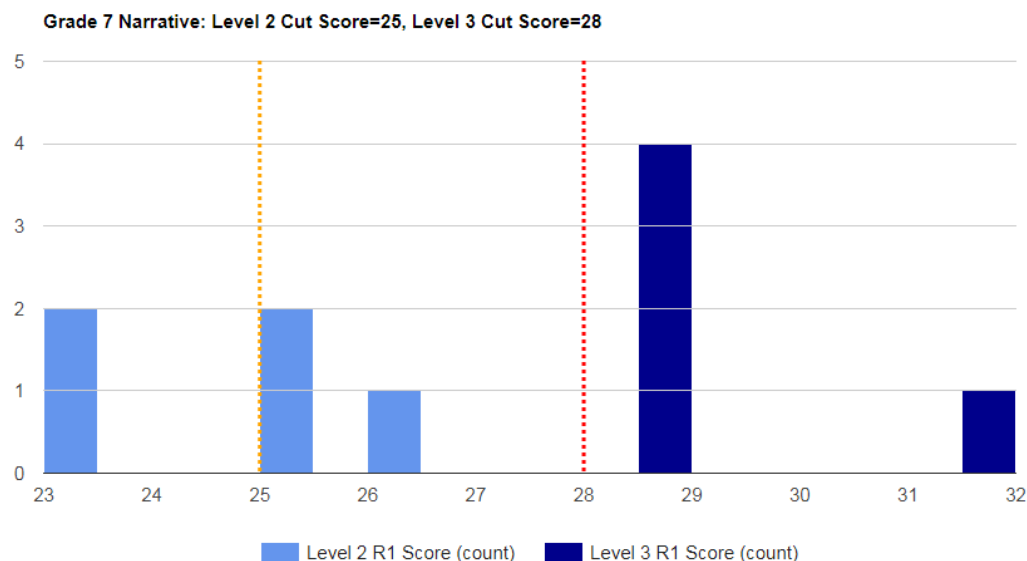
- Open the TurboMeeting at 1:30. Make sure it is working properly. Call for help (Kathleen) if you need it.
- Open OPLS to Facilitator role for access to PLDs, essays, and R1 Comments.
- Minimize OPLS.
- Check log-in information to make sure everyone is present and accounted for.
- Welcome everyone back.
- Call roll if you are uncertain about anyone's presence. Delete names of those who were not on the Monday call and who did not start Round 1.
- Make sure everyone has their PLDs handy (this will be emphasized during the Monday webinar).
- Ask for general comments about the Round 1 process, problems, etc. Address as necessary.
- Present R1 Table/Charts for each genre (Argument/Opinion, then Informative, then Narrative). There are a couple of issues about the graph you will need to point out. Because we had to allow for the possibility of having thresholds for both Level 2 and Level 3 on the same page, the spacing will look a bit odd in the graphs. The values for Level 2 are to the far left, and the values for Level 3 are to the far right. It looks like the Level 3 bar for score point 28 in the example below actually represents a score of 29, but it doesn't. It is simply occupying the far right half of the space for score point 28.

This may take some getting used to, but you can verify the proper placement of the bar by rechecking the table at the top of the page.

- Median – the middle score point (4th out of 7 panelists or halfway between the 3rd and 4th score point out of 6 panelists) for this threshold
- Min – minimum score point assigned by any panelist for this threshold
- Max – maximum score point assigned by any panelist for this threshold
- Range – the difference between the minimum and maximum score points assigned for this threshold

Grade 7-8 Narrative

Grade	Median		Min		Max		Range	
	Level 2	Level 3	Level 2	Level 3	Level 2	Level 3	Level 2	Level 3
7	25.0	28.0	23	28	26	31	3	3
8	27.0	31.0	25	29	28	32	3	3



- Note distribution of cut scores (thresholds). Depending on distribution, begin with discussion of cut score with the greatest range. Ask why there would be such a wide variance of opinion. Panelists will not have access to the essays or their own R1 entries; therefore, maximize OPLS to begin the discussion of R1 comments, cut score distribution, and specific essays selected as threshold representatives. Show one of the extreme essays a panelist selected to represent Level 2 of the lower grade (5, 7, or 9). Ask someone to volunteer to tell why this is a good candidate for Level 2 for this grade, using the PLD text as supporting evidence. Continue with other essays in the same

manner and ask for others to comment on why that particular essay should or should not be the one to represent the threshold of that level for that grade.

- Continue in this manner for all cut scores, particularly for those for which there is considerable variability among the panelists. If there is a cut score for which there is general agreement, point that out and ask why that might be. Don't spend a lot of time on those.
- At this point, the discussion should be primarily on whether or not the relationship among cut scores across grades makes sense or is close to what panelists thought it might be. Discuss any oddities in the relationships and answer questions that arise.
- At the conclusion of the discussion of distribution of cut scores, ask for any questions panelists may have about procedures or logistics (navigation, placing two cut scores on the same score point or essay, etc.). Remind panelists that the software won't accept more than one cut score on a given point but that in Round 2, the essays will be more tightly grouped, and there will be some score points with more than one essay. If there are two essays with scores of 27, for example, it would be possible to set one threshold on the first 27 and the other threshold on the second 27.
- Return to your main screen and show the login directions for Round 2.
- Describe the process for Round 2. It is essentially the same as Round 1 except for the fact that the essays are more closely packed. In essence, we have eliminated all the Round 1 essays no one identified as a potential threshold representative and replaced them with others with scores closer to the ones someone did identify. There may well be two or more essays with the same score, which will be shown in the score column. Their task will be the same, however: identify the first essay to meet the criteria for Level 2, lower grade, then Level 3 lower grade, then Level 2 upper grade, then Level 3 upper grade. For grades 5-6 and 7-8, start with the first packet in the list, complete it, and then complete the second packet. Enter and submit all cut scores for one packet before moving on to the second packet.
- Ask for questions or comments; answer all questions. If you can't, get the panelist's name, and promise to e-mail an answer within the hour. If it is a general concern, copy everyone on your list.
- Remind panelists that their Round 2 entries are due by 9 A.M. Friday (EDT). We will close the application at that time.
- Thank panelists for their participation, and end the webinar.

- Obtain answers to any questions you were unable to answer and send an e-mail to the panelists within the hour.
- Begin monitoring Round 2 activity. Continue to monitor throughout the next day. Check your e-mail as well. If there are panelists who have not completed all packets 5 P.M. Thursday, send them an e-mail asking them to complete by 9 A.M. Friday. Remember that for west coast panelists, that's 6 A.M., and you will be calling them at 2 P.M. their time. Use your own best judgment as to when to call those panelists.

Sample Vertical Articulation Script

Points to consider in the data include the following:

- Reversals – higher theta cuts in lower grades
- Differential pacing – with a vertical scale, uneven size of increase from one grade to the next (e.g., a small difference between theta cuts for grades 6 and 7 but a large difference between grades 7 and 8)
- Differential opportunity to learn – better implementation of the CCSS at certain grades
- Differential test difficulty – potentially interacting with Pacing
- Idiosyncrasies in one or more panels – systematically stringent or lenient interpretation of the PLDs; differences in interpretation of the charge

Keep all of these possibilities in mind, but be prepared for other suggestions that may come from the Vertical Articulation Committee (VAC).

Once VAC training is over, display the actual Round 3 impact graphic. Review the process by which these impact data were obtained, and give panelists a few minutes to size it up. Then ask:

Do these results look more like the generally increasing, stable, or generally decreasing examples we saw earlier?

Give them time to form an opinion. Depending on the overall shape of the graphics, they could be any of the above or no particular shape at all. Say:

Keep in mind that each point on this graph represents the performance of a whole grade of students in a particular grade. For example, the point for Grade 6, Level 3, represents the percentage of grade 6 students scoring at or above Level 2. Looking just at the graph for Level 3, do you see any points that seem out of place; that is, a point that seems higher or lower than you would expect, given performance in the adjacent grades?

Give panelists time to identify any unusual looking points on the **Level 3 graph only**. If anyone identifies a point on the Level 2 or Level 4 line, note that we will get to Levels 2 and 4 later; we are focusing on Level 3 now. It is possible that two or more points will be identified. Ask each panelist who identifies a point to tell how it relates to adjacent grades' points and how it seems out of place. After each panelist explains his or her data point, select one for the group to discuss. The following is a hypothetical example in which one panelist identified a dip at grade 6 and another identified an unusual rise at grade 4.

Then reiterate the ground rules:

We can change any one of the cut scores set yesterday or none of them. To make any change, we will need a formal motion to change a specific cut score for a specific level for a specific grade to a specific value, associated with a specific page or set of pages in the ordered item booklet. For example, if you want to raise the cut score for grade 7, Level 2, you would say, “I move that the cut score for grade 7, level 2 be moved from XXX to YYY,” and give the corresponding page numbers in the grade 7 OIB. Then someone will need to second the motion. Once we have a motion and a second, we will discuss the pros and cons of the motion. I will call the question, and we will vote on the change. Given the level of effort that has already gone into each of these cut scores, we will require a 2/3 majority for passage of any motion. If a motion fails to get a 2/3 majority, it fails; if it receives a 2/3 majority, it passes, and the cut score is changed. Are there any questions about the procedure?

Let’s take a look at grade 6. From grade 5 to grade 8, there is a general increase in the percentage of students performing at Level 3 or higher, except for grade 6. Does it seem logical that there should be such a “dip” at grade 6?

Wait for responses, and in each instance, ask the panelist to tell why it does or does not seem logical for such a “dip” to occur, without any further information about the tests or students who took them. Use panelists’ comments to guide the next portion of the discussion.

It looks like most of you found this “dip” unusual. Let’s take a look at the actual Level 3 cut score for grade 6, relative to those for grades 5 and 7, for example. We can see from the first table at the upper left of the screen that

[Scenario 1: Differential pacing] *The difference between the cut scores for grades 3 and 4 is much larger than the difference between the grade 4 cut score and the grade 5 cut score or between the grade 5 cut score and the grade 6 cut score. In other words, we have asked fourth graders to make more progress from grade 3 to grade 4 than we have asked fifth graders or sixth graders to make from the previous grade. Is it reasonable to have such an expectation?*

Allow panelists to respond to this question. If such an expectation is reasonable, it may not be necessary to consider this part of the graph at all. However, if most panelists believe it is not reasonable, move forward:

All right; it seems most of you believe that it is unreasonable to expect students to make larger gains grade 3 to grade 4 than from grade 4 to grade 5 or from grade 5 to grade 6. We could conclude that the grade 3 cut score is on target, but the others are off target; or we could conclude that grade 3 is off target, and the others are on. For the moment let’s assume the others are on target, and grade 3 is off. Let’s see how we arrived at this level 3 cut score for grade 3. Can someone from the grade 3 panel give the rest of us a brief summary of the last round of discussion you had surrounding the Level 3 cut score?

[Scenario 2: Reversal] *When we look at the Level 3 cut score for grade 3, we see that it is actually higher than the Level 3 cut score for grade 4. We have placed these tests and items on a common scale, so what this seems to be telling us is that we expect the third grader just entering Level 3 to know more and be able to do more than the fourth grader just entering Level 3. Is it reasonable to have such an expectation?*

Allow panelists to respond to this question. If such an expectation is reasonable, it may not be necessary to consider this part of the graph at all. However, if most panelists believe it is not reasonable, move forward:

All right; it seems most of you believe that it is unreasonable to expect third graders just entering Level 3 to know and be able to do more than fourth graders just entering Level 3. Can someone from the grade 3 panel give the rest of us a brief summary of the last round of discussion you had surrounding the Level 3 cut score?

For both scenarios (plus others that might be imagined), if no one responds immediately, call on a couple of people from the third grade panel, using your roster and making sure not to put any one person on the spot. After they have given a review of the panel's rationale, ask all panelists to call up the grade 3 OIB and turn to the region in which the Level 3 cut score was set. This should include 2-4 items on consecutive pages. Direct the panel to consider these items for a moment. Then say,

Given the Level 3 cut score set by the third grade panel, we are saying that the performance of students just entering Level 3 would be best characterized by these items and tasks or others similar to them. In fact, it comes closest to the item on page [X, the page with the RP value closest to the cut score but not below it]. I would also point out that half the panel placed their Level 3 bookmarks between pages [XX and YY, (the interquartile range for this panel for this cut)]. Now, considering the items in the vicinity of the cut, would anyone care to suggest that one of these other items on a previous page would be a better indicator of what students just entering Level 3 can do?

Wait for someone to suggest a page number. If no one suggests a page number, ask again if the items on the two pages surrounding the cut seem to be reasonable representations of what students just entering Level 3 know and can do. If there are still no recommendations, ask,

Do I hear a motion that we alter the cut score for grade 3, Level 3?

If someone responds, make sure the motion is specific, as described above. Record the motion and call for a second. If there is no second, the motion fails for lack of a second. If there is a second, record the second, and proceed to the discussion.

It has been moved and seconded that the Level 3 cut score for grade 3 be changed from [XX] to [YY]. Is there any discussion? Raise your hand if you would like to comment. We will have [AA] minutes to discuss this cut score, after which I will call the question.

Call on those who raise their hands, in order. The assistant will help with this task. The key feature of this discussion will be to urge all respondents to ground their comments in the PLDs and the contents of the items at or near the cut scores they are discussing. During the discussion, panelists will have access to all the OIBs and item maps, all the content standards, and all the PLDs. Facilitators will have the same access and the ability to project their own screens on the large screen in the front of the room. For every comment not based in the PLDs or content standards, politely ask the respondent to check the appropriate document in defense of the point he or she is making. At key points, it may be necessary to show on the big screen a particular page of the content standards, PLDs, or some other document to make sure everyone is focusing on the same information.

Gauge in advance the likely number of cut scores that may need to be considered and pace the discussion accordingly. Level 3 will generate the most discussion, so if it seems to be running a bit long, keep this in mind. Levels 2 and 4 will go considerably faster. Call the question at the appointed time. Ask for a show of hands of those who favor the motion. Have the assistant and one other person count hands. Then ask for a show of hands of those who oppose the motion. Have the assistant and one other person count hands. Tally the votes and announce the result. If the motion received a 2/3 majority, it passes; otherwise, the motion fails. If the motion fails, ask for another recommendation, and repeat the process. If no one recommends a new cut score, discussion of that cut score for that grade ceases. At that point, call the panel's attention back to the chart and ask if there are other cut scores that should be considered. For each such cut score, follow this same procedure.

When it is apparent that there will be no more changes for Level 3, ask for a motion to cease making changes for Level 3. On obtaining a motion and second, call for discussion. Call the question, and vote as before. Again, a 2/3 majority is required for passage of this motion. It is important to include this vote simply to preclude later charges that someone wanted to consider a particular cut score but was denied the opportunity. A formal vote by the full panel will show that the panel itself, not the facilitator or client, foreclosed further discussion. If the motion to cease discussion of any level fails, open the floor for additional motions for any cut score for that level. Give a full minute for the panel to think about it. If no one offers a motion in that time, it is entirely appropriate to call that level closed.

This action on the part of the facilitator may later be reversed if changes in another level make it clear that discussion at this level should be reopened. For example, if a Level 2 cut score is raised so much that it is very near the previously approved (or unchallenged) Level 3 cut score,

it is entirely appropriate to reopen discussion of Level 3 for that grade. However, this could reopen the entire set of Level 3 cut scores. Exercise caution and point out such consequences when the recommendation to raise the Level 2 cut score that high is made.

Take a 15-minute break after Level 3 and prepare for Level 2. While panelists are on break, check the voting records for Level 3, and check your notes for Level 2. Upon resumption of the vertical articulation, follow the same procedures as above.

Take a 15-minute break after Level 2 and prepare for Level 4. Follow the same procedures as in Level 2.

At the close of voting for Level 4, display the final cut scores and impact data for all grades and levels. On the cut score chart, highlight those that have been changed. Thank panelists for their assistance, and ask them to turn to the final evaluation form on their screens.

Appendix C: Readiness and Evaluation Forms

Sample	Page
Southeastern Assessment Consortium Science Standard Setting Readiness Form	103
Graduation Tests Standard Setting Readiness Form	106
VAC Readiness Form	107
Southeastern Assessment Consortium Science Standard Setting Evaluation Form	108
Graduation Tests Standard Setting Workshop Evaluation	109

**Southeast Assessment Consortium
Science Standard Setting Readiness Form**

Part 1: The Test

I understand the content and reporting categories for the test that students took this spring.

Yes No

I have taken and scored the test.

Yes No

I have discussed the test with my group.

Yes No

Part 2: Proficiency Level Descriptors

I have reviewed the Proficiency Level Descriptors for the grade assigned to me.

Yes No

I have discussed with my small group the Proficiency Level Descriptors for the grade assigned to me.

Yes No

I understand the progression of student achievement from Level 1 to Level 4.

Yes No

I understand how the Proficiency Level Descriptors relate to the tasks I have been asked to perform.

Yes No

Part 3: Bookmark Round 1

I have received instruction in OPLS.

Yes No

I have received instruction in the bookmark procedure.

Yes No

I have participated in the Practice Round for the bookmark procedure.

Yes No

I placed a bookmark for Level 3.

Yes No

I participated in the discussion following the tally of bookmarks.

Yes No

I understand how my bookmark will be combined with the bookmarks of other panelists to derive a cut score.

Yes No

I understand the procedure I am to follow and am ready to begin Round 1.

Yes No

Part 4: Bookmark Round 2

I have seen the results and heard the discussion of Round 1.

Yes No

I know where our panel set the Round 1 cut scores for Level 2, Level 3, and Level 4, based on the group's median bookmark.

Yes No

I understand the procedure I am to follow and am ready to begin Round 2.

Yes No

Part 5: Bookmark Round 3

I have seen the results and heard the discussion of Round 2.

Yes No

I participated in the discussion of percentages of students who would be classified at each level, based on our Round 2 cut scores.

Yes No

I understand the procedure I am to follow and am ready to begin Round 3.

Yes No

Graduation Tests

Standard Setting Readiness Form

Rater Number_____

Practice Test: I have completed the practice test, and I understand what I need to do to complete Round 1.

(Circle one): Yes No

Round 1: I have discussed the results of Round 1, including my ratings, the ratings of others, and the impact data, and I understand what I need to do to complete Round 2.

(Circle one): Yes No

Round 2: I have discussed the results of Round 2, including my ratings, the ratings of others, and the impact data, and I understand what I need to do to complete Round 3.

(Circle one): Yes No

VAC Readiness Form

I have seen the presentation about vertical articulation of cut scores.

☐ Yes

☐ No

I understand the purpose of vertical articulation.

☐ Yes

☐ No

I understand that recommendations to change a cut score will be done in accordance with a recognized democratic process.

☐ Yes

☐ No

I understand that there will be other reviews of the cut scores after the Vertical Articulation Committee completes its work.

☐ Yes

☐ No

Which of these patterns of student performance is/are expected? (Check all that apply.)

☐ Generally increasing over grades or time

☐ Generally decreasing over grades or time

☐ Generally stable over grades or time

☐ Up and down over grades or time

Which type of majority will be required to change a cut score?

☐ Simple Majority

☐ 2/3 Majority

I am ready to begin vertical articulation.

☐ Yes

☐ No

Southeast Assessment Consortium Science Standard Setting Evaluation Form

Panel (circle one) Grade 4 Grade 7 Grade 10

Indicate your agreement with the statement by circling one of the following:

SD for Strongly Disagree **D** for Disagree **?** for Uncertain
A for Agree **SA** for Strongly Agree

Statement	SD	D	?	A	SA
Taking and scoring the test was helpful.	SD	D	?	A	SA
The discussion of performance level descriptors was useful.	SD	D	?	A	SA
The bookmark presentation was clear and helpful.	SD	D	?	A	SA
The bookmark practice round was helpful.	SD	D	?	A	SA
I was able to navigate the standard setting software successfully.	SD	D	?	A	SA
My facilitator was able to answer my questions.	SD	D	?	A	SA
The discussion after Round 1 was helpful.	SD	D	?	A	SA
The discussion after Round 2 was helpful	SD	D	?	A	SA
The process was fair.	SD	D	?	A	SA
The process was orderly.	SD	D	?	A	SA
My group's final cut score for Level 2 is reasonable. [If you disagree, should it have been higher or lower? Circle one.]	SD	D	?	A	SA
My group's final cut score for Level 3 is reasonable. [If you disagree, should it have been higher or lower? Circle one.]	SD	D	?	A	SA
My group's final cut score for Level 4 is reasonable. [If you disagree, should it have been higher or lower? Circle one.]	SD	D	?	A	SA
The meeting site was a good place to conduct this activity.	SD	D	?	A	SA
Food service was good.	SD	D	?	A	SA
My personal needs (travel, lodging, accommodations, dietary restrictions) were met.	SD	D	?	A	SA

Comments on the back

Graduation Tests Standard Setting Workshop Evaluation

Please respond to the statements below by indicating your agreement or disagreement. Check one box for each statement to indicate whether you strongly agree, agree, disagree, or strongly disagree with each statement. Room for comments has been provided at the bottom of the form.

Subject: ☐ Science ☐ Social Studies ☐ Writing

	Statement	Agree	Disagree
1	The workshop leaders clearly explained the purpose of the meeting.		
2	The workshop leaders clearly explained my task.		
3	The examples and exercises helped me understand how to perform my task.		
4	The large and small group discussions helped me understand the process.		
5	I was able to follow the instructions and complete the rating sheets accurately.		
6	The discussions after the first round of rating were helpful to me.		
7	The discussions after the second round of rating were helpful to me		
8	The information showing the distribution of student scores was helpful to me.		
9	The facilities and food service helped to create a good working environment.		

Comments [Continue on the back if necessary.]

Appendix D: Job Descriptions

Job Description	Page
Project Manager	111
On-Site Coordinator	112
Lead Facilitator	112
Facilitator	113
Data Analyst	114
Clerical Support Staff	114
Technical Support Staff	114
Panelists	115
Sponsoring Agency Lead	115
Official Reviewer	116

Project Manager

The project manager (PM) will generally be the person who has overall responsibility for all aspects of the project, from test design to delivery of score reports. This person will be the primary point of contact for the client and will be directly involved in all aspects of the project, including standard setting. Prior to standard setting, the project manager will be responsible for gathering requirements for developing the standard setting plan but will not necessarily be responsible for writing the plan.

The PM will be responsible for securing resources for the standard setting: Panelists, Lead Facilitator, Facilitators, Clerical and Technical Support Staff, meeting site(s), equipment, and supplies. The PM will be responsible for delivering the standard setting plan and, with the help of the Lead Facilitator, shepherding the plan through the review and approval process.

Between plan submission and the start of standard setting, the PM will be responsible for regular communication with the client, panelists, other stakeholders, and site staff. If there are regularly scheduled calls or meetings with the client, the PM will be responsible for scheduling, facilitating, and documenting them. The PM is also responsible for all logistics associated with standard setting: travel arrangements for staff and panelists, shipment of materials and equipment to the site, securing on-site services such as secure shredding or internet access, and other elements spelled out in the standard setting plan. The PM may delegate some of these responsibilities to others, primarily the Lead Facilitator or On-Site Coordinator.

During standard setting, the PM may be on site or may delegate on-site responsibilities to the On-Site Coordinator. On-site responsibilities include interaction with site staff to make sure meeting and guest rooms, meals, and services are satisfactory. Additional responsibilities include interacting with client staff on security matters, assisting the Lead Facilitator in work flow, and time management. The PM is responsible for daily briefings and debriefings, with the assistance of the Lead Facilitator.

After standard setting, the PM is responsible, along with the Lead Facilitator, for producing the final report and forwarding all necessary documentation to whoever will produce score reports. The PM, with the help of the Lead Facilitator, is also responsible for assisting the client with preparing for and presenting cut scores to the governing body for approval if the client so wishes.

The Project Manager should be an experienced manager familiar with all aspects of test development, administration, scoring, and reporting. He or she should have excellent oral and written communication skills and a thorough understanding of the client's needs and operating processes. For certification and licensure projects, the PM should understand the larger process of how individuals apply for and obtain certificates or licenses. For projects in education, the PM should understand the legal and technical mandates under which their clients operate, as well as the impacts of cut scores on students, teachers, and administrators.

On-Site Coordinator

The On-Site Coordinator acts in the place of the PM. On-site responsibilities include interaction with site staff to make sure meeting and guest rooms, meals, and services are satisfactory. Additional responsibilities include interacting with client staff on logistical and security matters and assisting the Lead Facilitator in work flow, and time management. If the PM is not present, the Lead Facilitator will generally be responsible for daily briefings and debriefings, with the assistance of the On-Site Coordinator.

The On-Site Coordinator should be an experienced manager familiar with logistics and meeting management. He or she should be able to interact positively and successfully with large numbers of people of varying backgrounds and dispositions and resolve logistical problems gracefully and quickly.

Lead Facilitator

The Lead Facilitator is responsible for writing the standard setting plan, preparing all training materials, training Facilitators and other on-site staff, and preparing the final report. The Lead facilitator assists the PM in overall coordination of standard setting activities with other project activities. The Lead Facilitator works directly with Data Analysts to develop and document data analysis programs. He or she also interacts with technical advisors, internal and external reviewers, and others responsible for guaranteeing the technical adequacy of the standard setting plan and conduct.

Upon arriving at the standard setting site, the Lead Facilitator conducts a briefing with Facilitators and inspects the whole-group meeting rooms and other arrangements for the opening session. During standard setting, the Lead Facilitator conducts the whole-group training exercises (including vertical articulation) and oversees the Facilitators as they conduct small-group activities. At the beginning of each day, the Lead Facilitator assists the PM in briefing staff and clients regarding the events of the day. At the end of each day, he or she assists in or leads the daily debriefing with staff and clients. At the end of standard setting, the Lead Facilitator prepares the draft final report for review by the PM and client.

After standard setting, the Lead Facilitator presents results to internal and external technical advisors and may assist the client in preparing for and presenting results to the governing board, at the request of the client. The Lead Facilitator prepares most of the final report for internal and external review, interacts with reviewers, and revises the report as necessary.

The Lead Facilitator should have an extensive background in psychometrics and a thorough understanding of standard setting techniques. He or she should be able to select a standard setting approach appropriate to the test for which cut scores are to be set and defend that selection to both technical and non-technical stakeholders. The Lead Facilitator should have excellent oral and written communication skills and the ability to interact successfully with both technical experts and governing boards.

Facilitator

Each standard setting activity will have one or more Facilitators, typically one for each panel. Thus, for a single certification and licensure test event, there may be only one Facilitator. For an event for two subjects and seven grades, there may be 14 Facilitators or even more if some subject-grade combinations have more than one panel or if some panels have more than one facilitator.

Prior to standard setting, each Facilitator studies the test for which he or she will lead a panel, as well as all other materials pertinent to the setting of cut scores on that test; i.e., job skill analyses, content standards, test blueprints, performance level descriptors, and technical reports. The Facilitator also participates in training offered by the Lead Facilitator and helps to develop scripts to be used during standard setting.

Content specialists serving as Facilitators play an important role in the development of PLDs. They research the content standards and learning progressions to develop preliminary PLDs, work with other educators to revise the PLDs, and prepare working PLDs to use during standard setting. During a PLD development workshop, they lead panels of educators to review and revise draft PLDs.

Upon arriving at the standard setting site, the Facilitator participates in the initial briefing with the Lead Facilitator and then inspects the breakout room in which his or her panel will work. During standard setting, the Facilitator attends whole-group training exercises and provides follow-up instruction to a single panel. During those training exercises, the Facilitator follows the script prepared in advance and answers panelists' questions. Upon determining readiness to begin each round of standard setting, the Facilitator directs panelists to begin and circulates throughout the room to check progress. The Facilitator reports any problems or unusual activity to the Lead Facilitator or PM and implements solutions they recommend. The Facilitator participates in daily debriefings, reporting progress and concerns and listening to those of other Facilitators as well as those of the Lead Facilitator, PM, and client staff.

For remote standard setting, the Facilitator monitors panelist progress on a computer screen and responds to chat, e-mail, and text messages from them. Facilitators also lead inter-round webinars to review results of the previous round and prepare for the next round. They make sure panelists properly log in, follow directions, submit their cut scores, and log out properly.

Facilitators may be psychometricians, content specialists, or subject matter experts (in the case of certification and licensure). In some instances, a panel may require both a psychometrician and a content specialist. This requirement must be worked out between the Project Manager and client well in advance of standard setting. Each Facilitator should be familiar with the particular standard setting procedure being used and/or the subject matter of the test. Facilitators should have experience leading small groups and have excellent oral communication skills.

Data Analyst

The Data Analyst oversees data entry and analysis of standard setting data and produces reports for Facilitators to share with Panelists. The Data Analyst prepares and checks the data analysis programs, checks data going into and out of the program, and confers with the Lead Facilitator on the accuracy and relevance of the reports going back to Facilitators and Panelists. The Data Analyst works directly with the Lead Facilitator in developing and modifying analysis programs and reports and participates in the daily debriefing. The Data Analyst may create the analysis programs or may use programs developed by others. Data Analysts may also be involved in creating ordered item booklets, item maps, or other specialized materials under the supervision of the Lead Facilitator.

The Data Analyst should have a background in psychometrics or statistics and be familiar with the statistical/psychometric principles behind the standard setting method being used. Close attention to detail is an essential characteristic.

Clerical Support Staff

Clerical Support staff provide support to the PM and on-site staff, particularly the On-Site Coordinator in the absence of the PM. Specific responsibilities include correspondence with Panelists and maintenance of a Panelist database, notifying the PM when Panelists need to be replaced, producing training materials, and arranging for and monitoring shipment of materials and supplies to the event site. On site, Clerical Support staff help with taking attendance and distributing materials to Panelists. Clerical Support staff should be able to respond efficiently to multiple, simultaneous requests and keep track of people and materials.

Technical Support Staff

Technical Support Staff make sure that hardware and software work properly and handle internet connectivity and webinar setup and operation. They may or may not travel to smaller standard setting events, but particularly for large-scale events, their on-site presence is essential. They work with site-based staff to make sure all hardware is properly installed, that all software is tested, and that all connections are secure and fully functional. Technical Support Staff work directly with the PM and Lead Facilitator to respond to technical issues that may arise in any of the meeting rooms.

Members of the Technical Support Staff should be well-grounded in the hardware and software they are asked to maintain and have a demonstrated capability of solving technical problems quickly and accurately. Ability to troubleshoot the specific standard setting software is welcome but not required.

Panelists

Panelists are the individuals who actually provide the raw data from which cut scores are derived. For certification and licensure, they are typically subject matter experts; for educational tests, they are typically classroom teachers of the subject and grade for which cut scores are to be set. However, in K-12 settings, panelists may include building- or district-level administrators, parents, community leaders, or other stakeholders identified by the client.

Panelists study the tests and other materials needed for standard setting (e.g., job descriptions and KSAs for certification and licensure), test blueprints, and performance level descriptors (for K-12 standard setting), and receive instruction in the standard setting procedure to be used. They participate in focused training, apply the principles they have been taught during training sessions, independently enter item or work sample ratings as directed, discuss their ratings with other Panelists, and independently enter final item or work sample ratings

Some Panelists will also participate in vertical articulation of cut scores across grades in K-12 settings. In vertical articulation, they consider cut scores for tests they did not originally review, share their views with other Panelists, recommend changes, and discuss and vote on recommended changes.

In all instances, Panelists comply with security protocols provided by the Project Manager or sponsoring agency staff. They comply with groundrules explained during orientation and complete all tasks as prescribed in training materials.

Since Panelists may include individuals who are not content experts (e.g., parents, community leaders), there are no universal prerequisites. However, for content matter experts (SMEs) and educators, there are a few basic requirements. SMEs should be bona fide experts in their fields, as evidenced by certification or licensure in the field and several years of successful experience in that field. Teacher Panelists should have considerable experience in teaching the grade and subject for which they are empaneled. Educational administrator Panelists should have experience in supervising teachers or creating curriculum for the subject/grade for which they are empaneled.

Sponsoring Agency Lead

The sponsoring agency lead is the one person charged by the agency with the responsibility for getting standard setting done. This person will therefore oversee the work of those actually conducting standard setting. Specific responsibilities include approving or arranging for approval of the standard setting plan, contracting with someone outside the agency to serve as an Official Reviewer, overseeing the work of the Project Manager and Lead Facilitator during standard setting, interacting with the Official Reviewer and Lead Facilitator during standard setting, approving any deviations from the plan as necessary, participating in daily debriefings, and approving or arranging for the approval of cut scores once they are obtained. The sponsoring agency lead will typically present the results of standard setting to external reviewers and ultimate approvers of the cut scores.

Official Reviewer

Official Reviewers may be agency staff or independent contractors hired by the sponsoring agency. They differ from observers in that they have a contractual obligation to produce a written report. They monitor the standard setting process and report to the agency. They should become thoroughly familiar with the standard setting plan and may participate in its development, review, and approval. Once the plan is approved by the sponsoring agency, the role of the Official Reviewer is to make sure standard setting is conducted in accordance with the plan. To that end, the Official Reviewer attends and takes notes at all whole-group training sessions, visits individual panels to witness some or all training in each room, observes Panelists carrying out their assigned tasks, participates in daily briefings and debriefings, and prepares a formal report on his or her observations to the sponsoring agency. Ideally, the Official Reviewer interacts with sponsoring agency staff, the PM, and the Lead Facilitator during standard setting and points out issues that should and could be corrected before they become problems, rather than waiting until the end of the process to report them.

Official Reviewers should have a thorough knowledge of psychometrics in general and standard setting in particular. They should have a solid reputation in the field and the ability to interact effectively with other psychometricians and governing boards.

Appendix E: Additional Files Under Separate Cover

- **Worksheets**
- **PowerPoint Presentations**
- **Other Files**

Worksheets

Number	Document Name	Purpose/Contents
1	Worksheet 1 Plan Template.xlsx	Spreadsheets to enter who, what, when, where, and how elements of standard setting
2	Worksheet 2 Contingencies.docx	List of things that could go wrong and what to do about them
3	Worksheet 3 VAC Example.xlsx	Complete VAC interactive example
4	Worksheet 3 Data Entry.xlsx	Data entry and analysis spreadsheets for Angoff, Bookmark, and Body of Work

Worksheets 1, 3, and 4, are Excel workbooks with several spreadsheets. These are set up in Excel for two reasons:

1. Many were conceived and executed years before we began doing standard setting online. They can be used as they are or easily converted and integrated into online systems.
2. Excel is simple, easy, and nearly universal. These spreadsheets will work when sophisticated proprietary software fails.

Included with each Excel workbook is a **ReadMe** tab that explains each of the other tabs. Be sure to read that first.

Worksheet 2 is a Word document with a table set up in three columns: Contingency, Threat, and Mitigation. It sets out a number of things that can go wrong before, during, or after standard setting and describes how you can prevent them or lessen their impact.

Additional resources are available at <https://us.sagepub.com/en-us/nam/standard-setting/book227536#preview>



Disciplines Products F

[Chapter 13 - Cert scheduling \(xls\)](#)

[Chapter 13 - Ed schedule \(xls\)](#)

PowerPoint Presentations

File Name	Contents
Angoff Overview.pptx	Introduction to the modified and extended Angoff procedures
Board Pres. .pptx	Generic presentation of standard setting results to a state board of education
Body of Work Webinar.pptx	Slides for an online standard setting using a modified body of work procedure; introduction to software included
Conducting Vertical Articulation.pptx	Slides from an NCME presession I did in 2018; more oriented to what a Lead Facilitator needs to do rather than what VAC panelists will do
Generic Opening. pptx	Sample slides to start an opening session of any standard setting
Generic VAC Webinar. pptx	Introduction to vertical articulation done online

Other Files

File Name	Contents
1_ Designing and Implementing.pdf	Plan for the 2014 Smarter Balanced standard setting, included in a symposium presented at the annual meeting of the Council of Chief State School Officers, San Diego, CA in 2015
2_Standard Setting Ground Level .pdf	Report on preparations for and conduct of 2014 Smarter Balanced standard setting as part of an invited symposium presented at the annual meeting of the National Council on Measurement in Education, Chicago, IL in 2015
3_ Extended Angoff .xlsx	Example of how to set up an extended Angoff spreadsheet
4_ Bookmark Round 2.xlsx	Sample spreadsheet for calculating and presenting Round 2 data in a bookmark standard setting
5_ VAC Motion Form .docx	Form to use to record motions made during a vertical articulation
6_ Standard Setting to Go .pdf	Paper presented at NCME in 2017, describing an online standard setting for a nationally administered writing test
7_ Scaling .xlsx	Excel spreadsheet showing how to set up score scales after standard setting for various scenarios