

Designing and Implementing a Large-Scale Web-Based Achievement Level Setting Activity

Michael B. Bunch
Measurement Incorporated

Citation: Bunch, M. B. (2015). Designing and Implementing a Large-Scale, Web-Based Achievement Level Setting Activity. In M. Bunch (Organizer), *Innovations in Achievement Level Setting for Smarter Balanced Assessments*. Symposium presented at the annual meeting of the Council of Chief State School Officers, San Diego.

Designing and Implementing a Large-Scale, Web-Based Achievement Level Setting Activity

Michael B. Bunch
Measurement Incorporated

Abstract

This paper lays out the overall plan for the Smarter Balanced achievement level setting project, including design and implementation of the online and in-person activities, communications support, interaction with test developers and other contractors, the contributions of various advisory panels, field testing of methods, and logistics. It also focuses on the process by which cut scores were vetted and ultimately presented to the Chiefs of the Smarter Balanced governing states for approval.

Introduction

In a request for proposals (RFP) issued in October 2013, the Smarter Balanced Assessment Consortium (Smarter Balanced) called for a contractor to provide services for a multi-phase standard setting process and to plan and execute a comprehensive Communication Plan. The standard setting plan was to be executed in five phases from conducting distributed standard setting to finalizing achievement level descriptors. The communication plan was to “proactively explain the rationale for setting common achievement standards tied to the Common Core State Standards, describe the standard-setting process in layman’s terms, and make the case for approval of the performance standards derived from the standard setting process.” (State of Washington, OSPI, 2013, p. 10).

Measurement Incorporated (MI) responded to that RFP and was ultimately awarded a contract to carry out the work, using a Bookmark procedure (cf. Lewis, Mitzel, Mercado, & Schulz, 2012). MI completed the project with the assistance of two subcontractors, McGraw-Hill Education (CTB) and Hager Sharp, a Washington, DC communications firm. This paper provides an overview of the goals, activities, and outcomes of the project. Other papers in this symposium will focus on two specific aspects of the project: software design and the in-person standard setting activity that took place in Dallas, Texas, in October 2014. Links to additional material about the project are included in Appendix A.

Background

The project for establishing cut scores was just one of many projects carried out by Smarter Balanced over the past four years. Achievement level setting was officially known as Project 21, indicating that contracts for 20 other projects were already in progress or completed. Many of the mechanisms of support for this project were therefore already in place. For example, there were ongoing weekly (and sometimes more frequent) cross-contractor conference calls. Mechanisms for communicating with K-12 and higher education stakeholders in each Smarter Balanced state were also in place when the project began.

Communications support. For this project, the stakes were high enough and the communications challenges great enough that a communications subcontractor was deemed appropriate. For this

reason, Hager Sharp was added to the project team. Hager Sharp is perhaps best known for its work with the National Assessment of Educational Progress (NAEP). Its primary role on this project was to develop an overall plan for coordinating all incoming and outgoing communications. Hager Sharp also prepared a communications contingency plan, worked with media outlets throughout the country and conducted dozens of interviews of panelists in Dallas to produce a video recording of the proceedings to share with Chiefs and others in Smarter Balanced states. The contingency plan was fairly generic, but when a crisis did occur (report of a single Ebola virus case in Dallas two weeks before the in-person activity began there), the plan proved quite effective.

Hager Sharp and the rest of the project team achieved these goals through a series of interrelated activities:

- **Environmental scan** – Starting with source materials from the “Common Core State Standards, Public Opinion Research” primer from the Bill and Melinda Gates Foundation, the “The Language of Deeper Learning in America” focus group research conducted by Luntz Global on behalf of the William and Flora Hewlett Foundation, and the polling conducted by the Tarrance Group and David Binder Research on behalf of the Collaborative for Student Success (“Findings From National Survey On Common Core Standards”), Hager Sharp staff created and tested messages for stakeholder groups such as statewide organizations, districts, colleges and universities, education associations, and PTA affiliates to inform them about the upcoming achievement level activities and encourage them to participate in them.
- **Focus group study** – To gather feedback from key audiences on messages and materials for the Smarter Balanced standard setting activities, Hager Sharp conducted online focus group sessions among parents, teachers and school administrators from California and Michigan. The research focused on parents, teachers and school administrators. Feedback from the focus groups informed the team’s outreach efforts for the standard-setting activities. Out of these studies came the terms “achievement level setting” to replace “standard setting” and “threshold scores” to replace “cut scores.”
- **Social media campaign/paid promotion** – Hager Sharp created a variety of social media posts highlighting the online panel opportunity. These posts used messages intended to resonate with different target audiences, including teachers, higher education faculty, parents and content experts. In addition, Hager Sharp worked closely with *Education Week*, an online magazine focusing on education.
- **Communications planning for key activities and state approval** – One of the most critical events in the process was the vote of the Governing States, originally scheduled for November 6, 2014, and ultimately taken on November 14. Hager Sharp planned and carried out a series of external briefings and conducted a run-of-show for a press conference and staffing plans for each day’s events.
- **Intake interviews: Development of contingency plan, messaging and protocols** – As part of a contingency planning exercise, Hager Sharp conducted a kick-off meeting, a series of telephone interviews with key MI and Smarter Balanced staff, and a contingency training session.
- **Ebola contingency response** – America’s first reported case of the Ebola virus was in Dallas, Texas, two weeks prior to the in-person achievement level setting activity in that city. Hager Sharp monitored media outlets throughout the event, including daily review of stories in the

media and on social media sites, and maintained continuous contact with the CDC and local Dallas health officials. Hager Sharp's monitoring of the Ebola crisis also included crafting messaging to send out to panelists both prior to the in-person panel and during the event. Dr. Willhoft delivered these messages to panelists each morning during the in-person panel event.

- **Video production** – Hager Sharp developed a generic video as well as 18 additional versions of the video customized for use by Smarter Balanced member states. The video production crew spent three days on-site at the in-person panel, collecting over 60 interviews from panelists from all of the Smarter Balanced members and a range of categories, including teachers, higher education faculty, parents, representatives from business and the community, staff and advisors.
- **Achievement level naming research** – Hager Sharp interviewed nationally recognized testing experts and conducted online and in-person focus group sessions with parents, teachers, and students. They synthesized the results of their research and presented options to Smarter Balanced, primarily for the naming of Level 3 but secondarily for the other levels, which had previously been named Levels 1, 2, 3, and 4, with Level 3 designated the level to use for accountability reporting purposes.

Interaction with test developers and other contractors. While most of the activities for this project could be carried out independently of other projects, there were key handoffs that required not only split-second timing but careful planning and execution. The weekly cross-contractor conference calls (noted above) helped to identify these handoffs and the activities leading up to them.

Construction of the ordered item booklets for the Bookmark procedure was perhaps the most critical handoff of the project. Through the spring of 2014, American Institutes for Research (AIR) administered the Smarter Balanced field tests online and collected student responses. McGraw-Hill Education (CTB) and its subcontractors received and scored student responses and passed the raw data back to AIR. Educational Testing Service (ETS) took the raw data from AIR and calibrated both items and students using item response theory (IRT).

MI and ETS negotiated a final data handoff date of September 11, 2014. MI used those data to create 14 OIBs, which had to be reviewed, approved, and uploaded for the online panel to review by October 4. Given the extensive list of tasks to perform, MI, AIR, and ETS worked out a plan to obtain all item images, metadata, and graphics files in June and July. There was a dummy data handoff in August, which allowed MI to create dummy OIBs and test all systems that would be necessary for the production of actual OIBs in September. The advance preparations allowed all parties to complete their parts of the handoff smoothly and in a timely fashion.

TAC and Advisory Panel involvement. Smarter Balanced had a technical advisory committee (TAC) from its inception. This same TAC provided technical oversight for this project as well as others in progress at the time. MI staff met with the TAC five times during the life of the project:

- March 6 (webinar): Present plan
- May 30 (webinar): Present plan supplement
- July 16 (face to face): Present project update, field test plans, Chiefs presentation plans
- August 27 (webinar): Report on field test
- October 30 (face to face): Review in-person and online results

Project staff also met with an Advisory Panel, convened just for this project, three times over the life of the project. The charge of the Advisory Panel was to provide both technical and policy advice to Smarter Balanced and the project team. They reviewed the overall plan as well as the long-term validation plan.

Designing

Online panel. The purpose of the online panel was to allow as wide an audience as possible to see and comment on the Smarter Balanced assessments. From the outset, it was the goal of Smarter Balanced to use feedback from the online panel to inform, but not necessarily direct, the deliberations of the in-person panels.

Recruitment of the online panel began in April, with primary emphasis on educators. In June, recruitment opened up to the general public. Prior to any recruitment activities, however, the project team tested various recruitment messages. Specifically, Hager Sharp drafted and tested messages on focus groups of parents, educators, and other stakeholders. In addition to learning about the efficacy of various messages, Hager Sharp identified a number of concerns about Common Core, Smarter Balanced assessments, and other issues that might affect recruitment efforts. They brought those issues to the attention of the larger project team who, in turn, crafted messages to address them. Sample findings are highlighted below (Hager Sharp, 2014):

- Parents are confused about the plethora of assessments; they need easy-to-understand information about the objectives of the various assessments.
- They are also interested in specific examples of how information included in the standard is incorporated into the curriculum, and then how it is tested.
- Parents trust their local schools – they rely on teachers and school administrators for information. We need to make sure these intermediaries are fully prepared for the questions they will get from parents.
- Parents want assurance that the standard setting is based on the authority of experts and a scientific process.
- Parents are concerned about the length of time required to participate and having to reserve that time so far in advance. They need more flexibility in their schedules. Moreover, they are concerned about what they would contribute to the process and how their contribution would be used.
- In terms of messaging for the online opportunity, the best motivation for parents to participate would be more along the lines of “experience what your children will experience” rather than “help us set the standards your children will have to meet.”
- Parents have questions and skepticism about technological challenges; these need to be addressed in the messaging.
- Parents are concerned about perceptions related to declining test scores. They need informative and encouraging messaging about this.

Once the project team understood the issues involved, the recruitment process commenced along multiple lines. State Leads for higher education and K-12 education were identified and brought into the process. Hager Sharp and Smarter Balanced staff crafted specific messages for each potential audience and prepared toolkits for leads in each state to use to recruit teachers and higher education faculty and staff. They produced additional materials, deposited on the Smarter Balanced website, to answer

frequently asked questions. From April to October, project staff sent thousands of e-mail messages and employed various social media to get the message out.

In-person panel. The purpose of the in-person panel was to review ordered item booklets (OIBs) in a Bookmark procedure and recommend cut scores for the 14 Smarter Balanced tests. There were 16 panels in all: one for each subject-grade combination for grades 3-8 and two per subject-grade combination for grade 11. Each panel for grades 3-8 was to have 30 members, while the four grade 11 panels were to have 36 members each. Therefore, the recruitment goal was to select 504 panelists (360 for grades 3-8 and 144 for grade 11) with the following representation:

- Representation from every Governing State
- Minimum of one on-grade general education teacher per grade level from every Governing State
- Diversity of teaching experiences among panelists, including panelists who have worked with special populations
- Diversity of demographic characteristics among panelists
- Representation from two- and four-year colleges and universities for Grade 11 panels

To meet these goals and select participants, Smarter Balanced implemented a nomination process in which Governing States nominated potential panelists based on grade level and subject panel targets specific for their state. Table 1 shows the overall panel targets by role used to recruit and select panelists. The target was 252 panelists for ELA and 252 for mathematics.

Table 1
Target Counts by Role In-Person Panels

Role	Grade							Total
	3	4	5	6	7	8	11	
Educator: ELL experience	2	2	2	2	2	2	4	16
Educator: General	17	17	17	16	17	17	18	119
Educator: Non-Teaching	6	6	6	7	6	6	10	47
Educator: SWD experience	2	2	2	2	2	2	4	16
General Public	3	3	3	3	3	3	6	24
Higher Education							30	30
Grand Total	30	30	30	30	30	30	72	252

It is noteworthy that the grade 11 panels were designed to have 30 higher education participants out of a total of 72 (43%). Moreover, each panel for grades 3-8 was to have five above-grade educators. The rationale for the higher education personnel and above-grade teachers on the panels was to focus the panel on next steps. The focus of the grade 11 panels was college and career readiness. The focus of the panels for grades 3-8 was preparation for the next grade. Who better to provide that focus than those who would receive those students a year or two after testing?

Recruiting for the in-person panel followed the same process and timeline as did online panel recruitment. CTB staff used the information from the online Nominee Registration Form. Only nominees

with complete registrations (those who were nominated by the state and completed online registration form) were considered for selection. Each panelist was selected to achieve coverage of the necessary content expertise and state representation.

The selection process began on May 27, 2014, after states submitted their nominees. CTB incorporated a multi-stage selection process to ensure state representation. All 504 panelists and 10 alternates were selected and confirmed by August 5; however a number of selected panelists dropped out and had to be replaced. Throughout the selection process states were notified of their nominees' selection status and when applicable were requested to provide additional nominees or support to obtain a representative panel. Table 2 provides the criteria used and notifications provided for each stage of the selection process.

Table 2
Panelist Selection Criteria

Selection Stage	Matching Criteria	Notifications
Stage 1: Selection of panel targets within state and role	Matched state representation, content expertise, role, and grade level	June 9, states were notified of their selected nominees and requested to provide additional nominees for any identified gaps.
Stage 2: Selection of panel targets within state and across roles	Matched sample targets for state representation, content expertise, and grade level across roles	June 30, states were notified of their nominees' selection status and asked to help obtain confirmation for any non-confirmed selections.
Stage 3: Selection of panel targets across states.	Matched sample targets across states for content expertise, role, and grade level	September 5, states were informed that all selected and non-selected panelists had been notified and Contractor had a complete panel confirmed to participate.
Stage 4: Replacement	Matched sample targets first within state and then across states for any vacated panel positions	October 30, states were provided a final in-person panel Report listing all panelists who attended from their state.

Once participants were selected, they received an invitation to confirm their intent to participate. On multiple occasions the contractor had secured a full panel of 504 confirmed panelists plus alternates; however, due to a variety of circumstances, some of these panelists later declined to participate. Therefore, CTB continued to fill newly vacant panel positions until September 29, 2014.

Cross-grade review panel. The task of the cross-grade review panel was to take all recommendations for all grades from both online and in-person panels and make final recommendations that would make cross-grade sense to policymakers. These panelists were selected from the 16 in-person panels. They were nominated by State Leads after those panelists had confirmed their participation in the in-person panel for achievement level setting. Thus, each panelist, knowing he or she would be serving on this panel, would have opportunity to hear discussion during the single-grade achievement level setting activity and bring those insights to the cross-grade review panel.

During the selection process, nominees indicated their interest in and availability for participating in the panel. On August 15, 2014, CTB staff provided each state with a list of confirmed nominees and requested the State Lead or designee to consider the following attributes in recommending cross-grade review panel members/table Leaders:

- Ability to facilitate discussions among small groups
- Ability to find diplomatic middle-ground solutions between panelists
- Distinguished service to education
- Superior knowledge of Smarter Balanced Achievement Levels and/or Common Core State Standards

CTB staff used the recommendations received from State Leads to create a cross-grade review panel as depicted in Table 3.

Table 3
Desired Composition of Cross-Grade Review Panel for Each Content Area

Group	Number Desired	
	Grades 3-8	Grade 11
On-Grade Special Education	1-2	1-2
On-Grade ELL	1-2	1-2
On-Grade General Education	6-8	6-8
Above-Grade General Education	2-4	0
Non-Teacher Educators	1-2	1-2
Higher Education Faculty	0	2-4
General Public	1-2	1-2
Total per Grade Span	16	16
Total per Content Area	32	

The goal was to create the two panels, as designed, with each Smarter Balanced Member State having at least one representative on either the ELA or Mathematics cross-grade review panel. This goal was completely met; each panel had 32 members distributed as shown in Table 3. The majority of committee members selected to participate in the cross-grade review panels also served as table leaders at the in-person panel. The exception was at the grade 11 level where there were more cross-grade review panel members than table leader positions available. Cross-grade review panel members

who had completed their on-grade panel work by October 15 or October 17 had the option to stay in Dallas to attend the meeting on October 20 or to leave and return to Dallas on October 19, if they wished.

Chiefs meeting. The ultimate goal of all this activity was to arrive at a set of cut scores that the chief state school officers of the Smarter Balanced Governing States could approve. To that end, these Chiefs were brought into the process early, kept informed, and given an opportunity to comment on and recommend changes to the process before being asked to approve the results. The complete plan – including a description of the Bookmark procedure, timetable, and in-person panel agenda – was presented to Chiefs on April 29 and approved on April 30, 2014.

Implementing

Field test. Prior to putting any of the components of the project in place, MI conducted a field test of the online, in-person, and cross-grade review panels in August. Eighty MI readers with characteristics similar to those of the panelists who would convene in October were recruited to participate in either the in-person/cross-grade or online panels. Forty readers served on the online panel, and the remaining 40 served on the in-person/cross-grade panel. MI and CTB staff prepared draft versions of all training materials and procedures. A beta version of the online achievement level setting software was used for both activities. Because the ETS data handoff had not yet taken place, field test panelists examined one of four ELA booklets comprised of field tested Smarter Balanced selected-response items. Within both the online and in-person panels, 10 panelists examined each of the four tests (grades 4, 6, 8, and 11).

Online panelists participated in a 45-minute webinar and then logged in to the online system to review one OIB and set one bookmark (10 panelists for each of the four tests). Some panelists completed the task fairly quickly; others took several hours. MI staff sent a Survey Monkey evaluation form to all online panelists the next day to obtain feedback.

In-person panelists received the initial training that was being planned for the October meeting. They then practiced using the software and the Bookmark procedure under the direction of experienced facilitators. After the first round, they discussed results at their respective tables. After the second round, there was roomwide discussion with impact data. After the third round, 20 of the in-person panelists reconvened as a cross-grade review panel, and the other 20 reconvened to debrief with one of the lead facilitators. Both groups provided feedback to their facilitators regarding the clarity of the process and their level of confidence in the cut scores they had set.

After all panel activities were completed, MI staff met with Smarter Balanced staff to review all processes, programs, and materials. Outcomes of that discussion are summarized below.

- Preparation for online panelists – Online panelists will feel time pressure even though they have two days to complete their task. MI staff need to review all directions and navigation aids to make sure online panelists understand their tasks on the first attempt and navigate without error.
- Scripts for facilitators and Smarter Balanced staff – Field test scripts are useful resource documents, but need to be recast in bite-size chunks for facilitators to use in October.
- Logistical issues – MI and CTB should revise the schedule to accommodate more instruction on Day 1 and time to address issues not technically central to achievement level setting but

important to panelists. Attention should be given to sound equipment (wireless headsets), dual computers for facilitators, and an expanded breakfast menu. Security issues should also be addressed, such as the use of notepaper that can be collected at the end of each session and a clear and enforceable cell phone policy.

- Software issues – Several small problems that arose during the field test were corrected within minutes. Other issues will need to be addressed up until October 6. For example, MI staff should work with other Smarter Balanced contractors to upgrade the appearance of items rendered in PDF format so that panelists will be able to identify the parts of the item that are in play (e.g., highlight, drag and drop) and which elements constitute a correct response. The software and directions for the online panel should also be tightened up to prevent the issues that arose during the online field test.
- Event management – The command center, pre-event briefings, and daily debriefings are critical parts of the event management plan. Clear lines of responsibility, communication, and issue escalation should be documented as well.

Using feedback from all panelists and this meeting, MI and CTB staff made several changes to prepare for October. Basically, the software proved to be quite robust, and all procedures were fundamentally sound, requiring minor adjustments. The most important outcome of the field test was the standardization of facilitator behavior. MI and CTB staff prepared a facilitator script and resource guide and additional training materials to make sure all facilitators presented the same message to panelists.

Online panel. Online panelists signed up for one of six 48-hour windows, the first of which started on October 6, 2014. Ultimately, all windows were extended, and the final closing date was moved to October 18. By October 6, 10,099 individuals had registered to participate. Of that number, 5,840 logged in, and 2,660 placed a bookmark. Results for online panelists entering a bookmark are presented in Table 4.

Table 4
Numbers of Online Panelists Entering a Bookmark, by Role, Grade, and Subject

	Teachers		Administrators		Higher Education		Other	
Grade	ELA	Math	ELA	Math	ELA	Math	ELA	Math
3	151	167	67	37	9	5	31	30
4	89	124	31	28	2	4	16	22
5	96	114	31	35	5	5	12	21
6	66	91	11	22	4	8	9	17
7	70	100	12	22	4	5	6	8
8	87	115	27	39	4	7	11	22
11	193	267	55	64	60	83	13	26
Total	1,730		481		205		244	
Grand Total	2,660							

The 2,660 bookmarks (for Level 3) were translated into cut scores, which were then translated into percentages of students scoring at or above those cut scores. Results are presented in Figure 1.

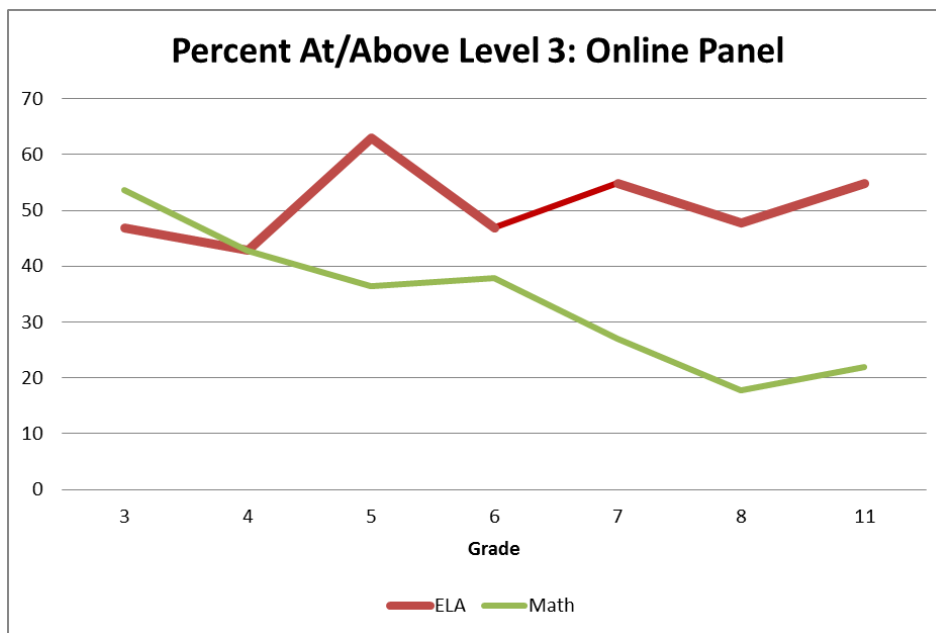


Figure 1. Results of Online Panel

In-person panel. A total of 601 nominees were invited to attend. This number included 504 confirmed panelists plus replacements and alternates. Shortly before the online panel event was to take place in Dallas, Texas, a case of the Ebola virus was reported there, and several confirmed panelists requested to be excused. Alternates took their place, but the requests to be excused continued. In the end, 482 panelists (96 percent of target) participated in the in-person panel for achievement level setting. Tables 5 and 6 present the numbers of panelists who served on each panel. Table 7 shows the actual numbers of panelists participating, by sex and race, along with a comparison to national and Consortium norms. Hispanic panelists were somewhat underrepresented, while Asian and multiracial panelists were somewhat overrepresented.

Table 5
In-Person Workshop Panelists by Subject and Grade

Grade	English Language Arts/Literacy		Mathematics	
	Planned	Obtained	Planned	Obtained
3	1 panel of 30	1 panel of 26	1 panel of 30	1 panel of 30
4	1 panel of 30	1 panel of 27	1 panel of 30	1 panel of 29
5	1 panel of 30	1 panel of 27	1 panel of 30	1 panel of 29
6	1 panel of 30	1 panel of 30	1 panel of 30	1 panel of 30
7	1 panel of 30	1 panel of 27	1 panel of 30	1 panel of 30
8	1 panel of 30	1 panel of 30	1 panel of 30	1 panel of 29
11	2 panels of 36	2 panels of 34	2 panels of 36	2 panels of 35
Total	252	235	252	247

Table 6
In-Person Panelist Representation by Role

Role	Grade							Total
	3	4	5	6	7	8	11	
Educator: ELL experience	3	5	7	5	4	5	7	36
Educator: General	33	34	32	31	31	33	48	242
Educator: Non-Teaching	12	8	12	16	11	14	17	90
Educator: SWD experience	3	5	3	4	5	4	7	31
General Public	5	4	2	4	6	3	4	28
Higher Education							55	55
Grand Total	56	56	56	60	57	59	138	482
% of Target	93	93	93	100	95	98	96	96

Table 7
Demographic Representation for All In-Person Panels

Group	Panels	Nation	Consortium
Female	76.6%	75.9%	---
Male	23.4%	24.1%	---
Hispanic	3.9%	7.8%	6.7%
Multiple Race, not Hispanic	3.5%	1.0%	1.2%
American Indian, not Hispanic	0.2%	0.5%	0.0%
Asian, not Hispanic	4.6%	1.8%	2.1%
Black, not Hispanic	6.4%	6.8%	3.3%
Native Hawaiian, not Hispanic	0.2%	0.1%	0.0%
White, not Hispanic	81.1%	81.8%	85.0%

The in-person panel meeting site was the Hilton Anatole in Dallas, Texas. MI's meeting planner visited the site, armed with the agenda and meeting room requirements, and worked out a contract, and Smarter Balanced approved the site. The site featured large meeting rooms for plenary sessions and multiple break-out rooms for the 30-person and 36-person panels. The meetings were set up in waves: Grade 11 on October 13-15, grades 6-8 on October 15-17, and grades 3-5 on October 17-19. All cross-grade review was done on October 20. There was never a time when all panels were meeting simultaneously, but there was one time (the morning of October 17) when 12 of the 16 panels were meeting, and all had to be accommodated in separate meeting rooms. Figure 2 shows a diagram of a sample breakout room.

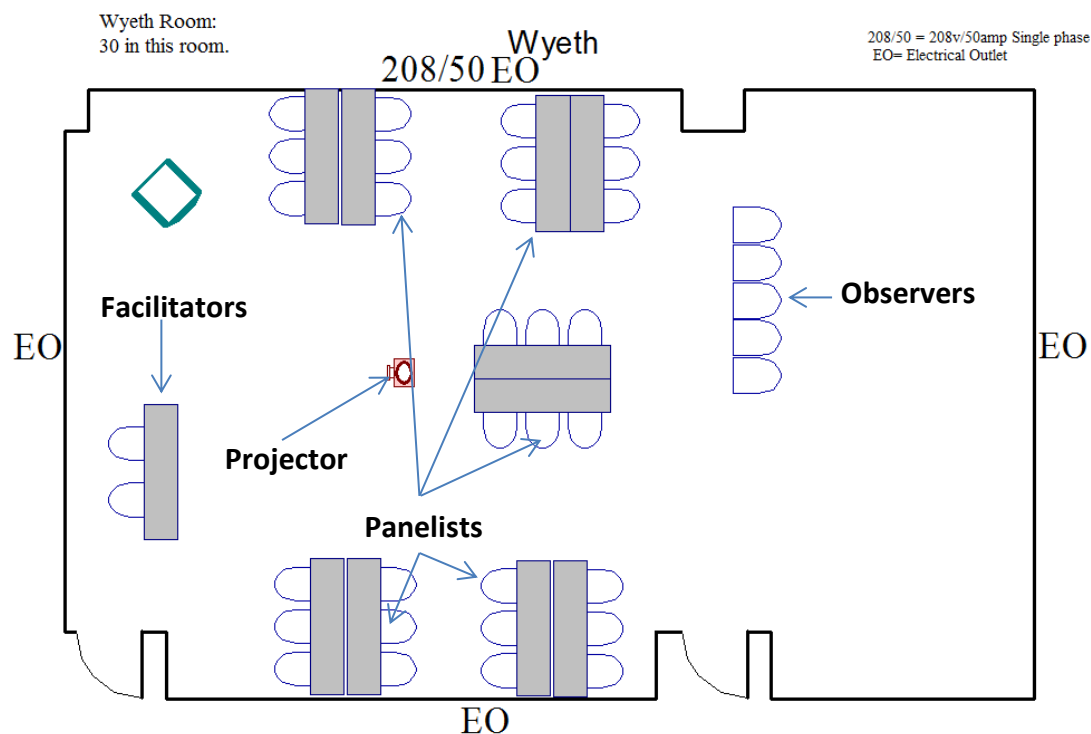


Figure 2. Layout of a Break-Out Room

- **Transportation** – Getting 500 people to Dallas from 22 states required the full-time services of one staff member. Some of those 500 people had to fly to Dallas twice, once for the in-person panel and once for cross-grade review. Once in Dallas, panelists needed to get from the airports (Dallas-Fort Worth International and Love Field) to the hotel and back as well as around Dallas. MI provided shuttle service to and from both airports and to nearby restaurants and shopping malls. This last measure contributed to the larger goal of making sure all panelists were comfortable, rested, and alert during achievement level setting. MI provided over 500 airline tickets and over 1,000 shuttle rides.
- **Lodging** – MI staff made hotel reservations for all panelists plus Smarter Balanced staff, MI staff, CTB staff, Hager Sharp staff, observers from Smarter Balanced states, and the auditors. Panelists returning for cross-grade review had two separate reservations. All reservations were for single rooms unless otherwise requested by the panelist. In all, MI made reservations for 2,005 room nights.
- **Meals** – Making panelists comfortable and attending to their basic needs are important aspects of successful management of an achievement level setting meeting, particularly one that stretches on for more than a week (cf. Bunch, 2012). MI staff worked closely with hotel staff to plan nutritious breakfasts and lunches as well as afternoon snacks and beverage breaks. Meals and snacks were varied, and individuals with specific dietary needs or restrictions received appropriate meals. All were on their own for dinner, but MI provided reimbursement for dinner meals on site and all meals en route to and from the meeting. In all, MI provided 4,338 meals and 2,169 snacks on site and reimbursed panelists for thousands more. Comments on the final evaluation forms about lodging and meals were uniformly very positive.
- **Materials and equipment** – All achievement level setting was web based, for both online and in-person panels. While online panelists used their own computing devices and internet connections, in-person panelists needed computers, monitors, and other equipment to do their jobs. MI transported most of this equipment from Durham, North Carolina to Dallas, Texas. The movement of equipment from Durham to Dallas was greatly facilitated by advance inspection of the site, conversations with hotel management and IT staff, and exchange of technical information between the two parties. IT staff at MI assembled and packed all materials, secured bonded transportation, and accompanied the equipment to Dallas. There, they worked with hotel staff and vendors to set up each room, run cables, and test all equipment two days prior to the start of the in-person workshop. Appendix B contains a list of all materials and equipment transported to Dallas as well as additional materials and equipment purchased or leased on site.
- **Flow** – As noted above, there were three waves of panelists for the in-person workshop in Dallas. Grade 11 panelists arrived on Sunday, October 12, started work the morning of October 13, and completed Round 3 by noon October 15. Panelists for grades 6-8 arrived on Tuesday, October 14, began work on the morning of October 15, and finished Round 3 by noon October 17. Panelists for grades 3-5 arrived on Thursday, October 16, began work on the morning of October 17, and finished Round 3 by noon on October 19. Table 8 shows the detailed agenda for each wave, including training of table leaders.

Table 8
Detailed Agenda for Each In-Person Workshop

Day	Time	Event
0	4:00 p.m.	Facilitator Training - October 12, 14, 16
	6:00 p.m.	Table Leader Training - October 12, 14, 16
1	7:30 a.m.	Badges, place cards, and room and table assignments
		Continental breakfast
	8:30 a.m.	Opening Session (Carpenter Ballroom)
		Keynote address and charge
		Q&A; introduction of the parking lot concept
	9:00 a.m.	Overview of schedule of events for the in-Person Panel
	9:30 a.m.	Panelists are dismissed to breakout rooms
	9:35 a.m.	Introductions
	9:45 a.m.	Orientation to hardware and software
	9:55 a.m.	Discussion of CCSS and ALDs
	10:30 a.m.	Break
	10:45 a.m.	Introduction to the test
		Overview of test development and scoring
	10:55 a.m.	Practice Test administration
	11:55 a.m.	Table-wide discussion of tests
	12:20 p.m.	Closing comments on test construction
	12:25 p.m.	Completion of Readiness Form Part 1
	12:30 p.m.	Lunch (Ballroom Lobby). Joe answers Parking Lot questions last 20 minutes.
	1:20 p.m.	Introduction to the Ordered Item Booklet (Carpenter Ballroom)
	1:50 p.m.	OIB training and review (breakout rooms)
	4:50 p.m.	Wrap-up for Day 1. Panelists log out; computers secured
	5:00 p.m.	Adjourn
	5:15 p.m.	Smarter Balanced, MI, CTB, and auditor debriefing
	6:00 p.m.	Secure meeting rooms and depart
2	7:30 a.m.	Registration opens (Ballroom Lobby)
		Badges, place cards, and room and table assignments
		Continental breakfast
	8:30 a.m.	Orientation to the Bookmark Procedure (Carpenter Ballroom)
	8:50 a.m.	Reconvene in breakout rooms
	9:00 a.m.	Orientation round
	9:40 a.m.	Completion of Readiness Form Part 3
	9:45 a.m.	Round 1
	11:45 a.m.	Reflection questionnaire
	12:00 Noon	Lunch (Ballroom Lobby). Joe answers Parking Lot questions last 20 minutes.

Day	Time	Event
2		Confer with Smarter Balanced staff to approve Round 1 analyses and reports
	1:00 p.m.	Review of Round 1
		Grade group facilitators review results of Round 1; lead discussion of bookmark placements, impact data, and rationales for bookmark placements
	2:30 p.m.	Completion of Readiness Form Part 4
	2:35 p.m.	Begin Round 2
		Panelists work in small groups as in Round 1
	4:40 p.m.	Reflection questionnaire
	4:50 p.m.	Wrap-up for Day 2
		Panelists log out; computers secured
	5:00 p.m.	Adjourn
		MI/CTB staff analyze data and prepare reports.
		Meet with Vertical Articulation Committee panelists.
	5:30 p.m.	Smarter Balanced, MI, CTB, and auditor debriefing
	6:00 p.m.	Secure meeting rooms and depart
	6:00 p.m.	Table Leader Training - Oct. 12th, 14th, 16th
3		Facilitator Training
	7:30 a.m.	Registration opens (Ballroom Lobby)
		Badges, place cards, and room and table assignments
		Continental breakfast
	8:30 a.m.	Convene in breakout rooms
		Review Round 2 results; grade group facilitators review results of Round 1; lead discussion of bookmark placements, impact data, and rationales for bookmark placements
	9:45 a.m.	Completion of Readiness Form Part 5
	10:00 a.m.	Begin Round 3
		Panelists work in small groups as in Rounds 1 and 2.
		Wrap-up for Day 3
	11:30 a.m.	MI/CTB staff process and share Round 3 results for this group plus others that have completed previously (not applicable to Grade 11; grade 6–8 groups will see only their results and those for Grade 11; grade 3–5 groups may see all results).
	11:45 a.m.	Completion of Evaluation Forms; log out; all computers secured
	12:00 Noon	Adjourn for lunch (Ballroom Lobby). On Days 3 and 5, outgoing panelists will be joined by incoming panelists. The Parking Lot concept should still work, but there will be different kinds of questions.
		Panelists may stay for lunch or leave if they have successfully logged out.
		MI/CTB staff discuss Round 3 data with Smarter Balanced staff and auditor, update achievement level setting report.

- On-site management** – Management of the in-person achievement level setting was a joint responsibility of Smarter Balanced and Measurement Incorporated. However, as prime contractor, MI was ultimately responsible for all activities. We carried out this responsibility through daily briefings with Smarter Balanced staff and the auditors, on-site training and retraining of facilitators and table leaders, and daily observation of panel activities, an activity Peters & Waterman (1982) refer to simply as “management by walking around.” Questions that came up in one panel were quickly discussed among facilitators and shared with other panels if they were relevant. The daily briefings identified problems not only with content but with schedule as well. Although we had extended training and practice time after the field test, we discovered that panelists still needed more time. MI and CTB staff modified the schedule of rounds after the first day of the grade 11 wave and allowed more time for the completion of Round 1 for all panels while still allowing them to complete the final round on time without being rushed. By clearly defining who was responsible for what, it was possible to respond quickly and effectively to anything that came up. A clear example of this role definition had to do with the ongoing Ebola situation. Hager Sharp was responsible for monitoring all media outlets and crafting messages. Dr. Willhoft was responsible for sharing those messages each morning or during lunch with panelists, observers, and the media. Everyone else supplied pertinent information related to their specific roles.

As Table 8 shows, the flow of activity was carefully orchestrated. Panels received task-focused training, followed by application of that training. Data flowed from panelists’ computers directly to the Bookmark Pro workstations and back again seamlessly. Panel discussions focused specifically on the previous round’s results, always within the context of the achievement level descriptors and what it means to be college/career ready or ready for the next grade’s work. Within panels, on-grade teachers, above-grade teachers, administrators, parents, business leaders, and others contributed their individual perspectives but always within the framework of the Common Core State Standards and ALDs.

In-person panelists were trained to apply the Bookmark procedure to online OIBs. They worked at tables of five or six computer stations (see Figure 2) and had access to the OIB, item map, ALDs, CCSS, and various Smarter Balanced resources during their review of the items. During Round 1, they worked independently. After Round 1, they discussed their results with other panelists at their own table and then received the results of the online panel for their grade and subject. After Round 2, they received their own results as well as impact data and engaged in a roomwide discussion. After Round 3, they were shown the final results and then completed an evaluation of the process. Figure 3 shows the results of Round 3 in terms of percentages of students who would be placed at or above Level 3. Figure 4 compares the Round 3 in-person panel impact with the online panel impact. The online and in-person trends were remarkably similar. Percentages of students at all levels are shown in Table 9. It should be noted that these percentages are calculated on field test performances.

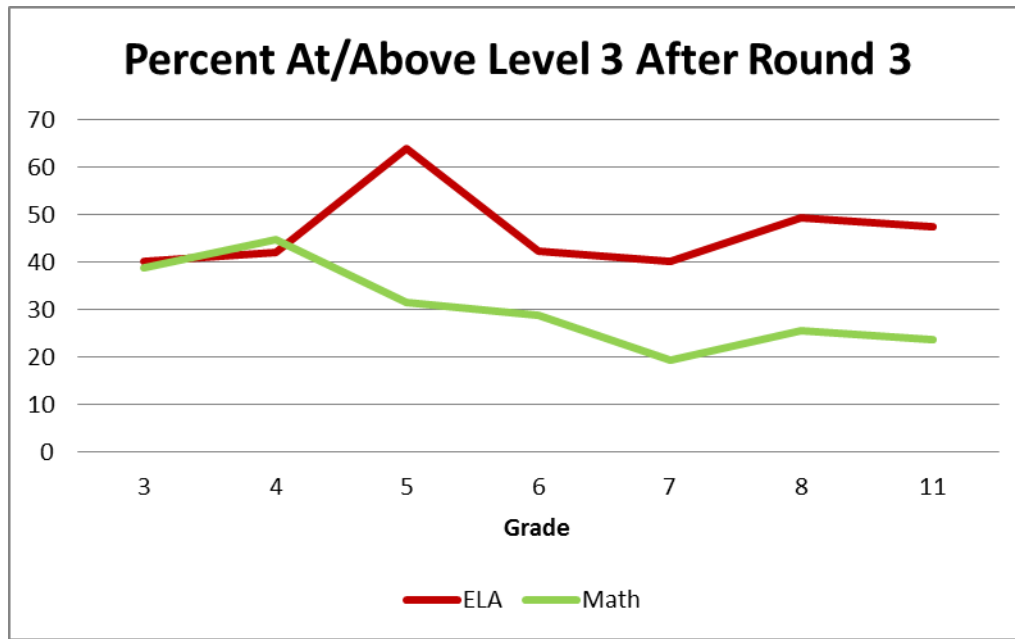


Figure 3. Round 3 Impact for In-Person Panel

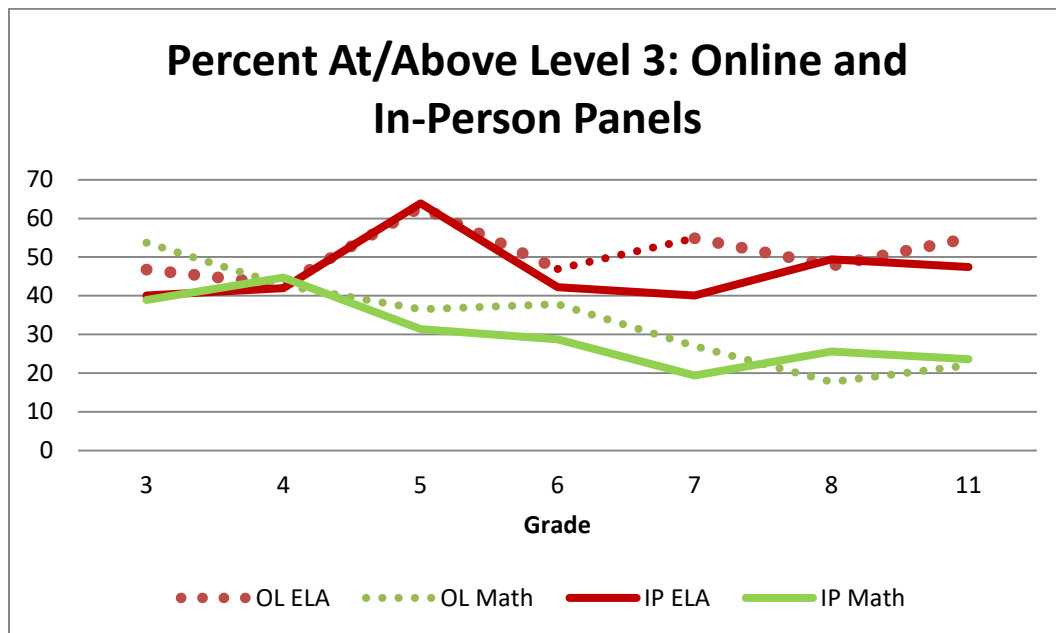


Figure 4. Impact Data for Online and In-Person Panels

Table 9
Percentages of Students at Each Level After Round 3

Test	Level 1	Level 2	Level 3	Level 4
Grade 3 ELA	33.5	26.4	21.0	19.1
Grade 4 ELA	35.6	22.4	23.1	18.9
Grade 5 ELA	21.4	14.7	47.0	16.9
Grade 6 ELA	33.8	24.0	30.0	12.2
Grade 7 ELA	31.7	28.2	33.5	6.6
Grade 8 ELA	23.6	27.0	39.2	10.2
Grade 11 ELA	27.3	25.3	35.8	11.6
Grade 3 Math	32.7	28.4	28.1	10.8
Grade 4 Math	22.4	32.9	29.1	15.6
Grade 5 Math	36.5	32.1	17.6	13.8
Grade 6 Math	41.7	29.6	13.4	15.3
Grade 7 Math	46.8	33.8	13.6	5.8
Grade 8 Math	48.7	25.7	18.2	7.4
Grade 11 Math	41.0	35.4	17.8	5.8

At the end of each wave, panelists were asked to complete evaluation forms. Responses are summarized in Table 10.

Table 10
In-Person Panelist Evaluations of Process

Statement	Agree or Strongly Agree
The orientation provided me with a clear understanding of the purpose of the meeting.	97%
The workshop leaders clearly explained the task.	93%
The training and practice exercises helped me understand how to perform the task.	96%
Taking the practice test helped me to understand the assessment.	97%
The Achievement Level Descriptions were clear and useful.	95%
The large and small group discussions aided my understanding of the process.	99%
The time provided for discussions was appropriate.	95%
There was an equal opportunity for everyone in my group to contribute his/her ideas and opinions.	97%
I was able to follow the instructions and complete the rating tasks accurately.	99%

Statement	Agree or Strongly Agree
The discussions after the first round of ratings were helpful to me.	97%
The discussions after the second round of ratings were helpful to me.	97%
The information showing the distribution of student scores was helpful to me.	96%
I am confident about the defensibility and appropriateness of the final recommended cut scores.	92%
The facilities and food service helped create a productive and efficient working environment.	98%

Cross-grade review panel. After Round 3, most panelists went home. Members of the grade 11 and grades 6-8 panels who had agreed to participate in cross-grade review returned on Sunday, October 19. All 64 panelists (32 per subject) described in Table 3 were present. Members of the grades 3-5 panels who had agreed to participate in cross-grade review simply stayed over Sunday night while the rest of their colleagues went home. The cross-grade review began with an introduction by Dr. Bunch, with comments from Dr. Willhoft and Dr. Alpert. The orientation focused on expectations, both within and across grades, and democratic principles. Groundrules were as follows:

- Moving bookmarks modestly up or down is reasonable
- Scale score cuts should increase across grades
- New recommendations should align with the ALDs

Panelists were instructed that any cut score could be changed within these guidelines and that the process would be carried out democratically; i.e., each change would require a motion, second, and 2/3 majority vote for passage, following these steps:

- Consider changes across grade-groups that may support articulation
- Give time for each grade to review OIBs and ALDs to consider changes
- Each grade reports to whole group on their findings
- Recommend actions based on findings and resulting articulation
- A motion for recommended actions
- Vote and note (2/3 majority)
- Secretary takes notes of each of the above steps, carefully records motions, votes, and results

After the general orientation, each panel started with a review of the grade 11 cut scores and then moved down to the cut scores for grades 3-8 in a systematic way. Panelists had access to all tests in their content area and could easily look up any item and its properties. The facilitators also had access to the items and could project them onto a screen to focus everyone's attention on a single item at one time. In addition, the facilitators had another projector to show Round 3 results in an interactive Excel spreadsheet that would change when any cut score would change, as shown in Figure 5. These panelists also had access to and discussed online panel data, disaggregated by group (teachers, administrators, higher education, general public).

Round 3 Bookmarks				% At or Above				% In Group				
Grade	Level 2	Level 3	Level 4	Grade	Level 2	Level 3	Level 4	Grade	Level 1	Level 2	Level 3	Level 4
3	12	29	69	3	79.7	63.7	12.1	3	20.3	16.0	51.6	12.1
4	15	28	77	4	77.6	58.7	8.8	4	22.4	18.9	49.9	8.8
5	14	26	68	5	75.1	54.5	6.6	5	24.9	20.6	47.9	6.6
6	13	24	67	6	74.4	49.5	7.8	6	25.6	24.9	41.7	7.8
7	9	29	60	7	69.7	34.7	6.5	7	30.3	35.0	28.2	6.5
8	8	25	60	8	63.7	41.4	7.4	8	36.3	22.3	34.0	7.4
11	15	28	65	11	64.0	42.2	7.2	11	36.0	21.8	35.0	7.2

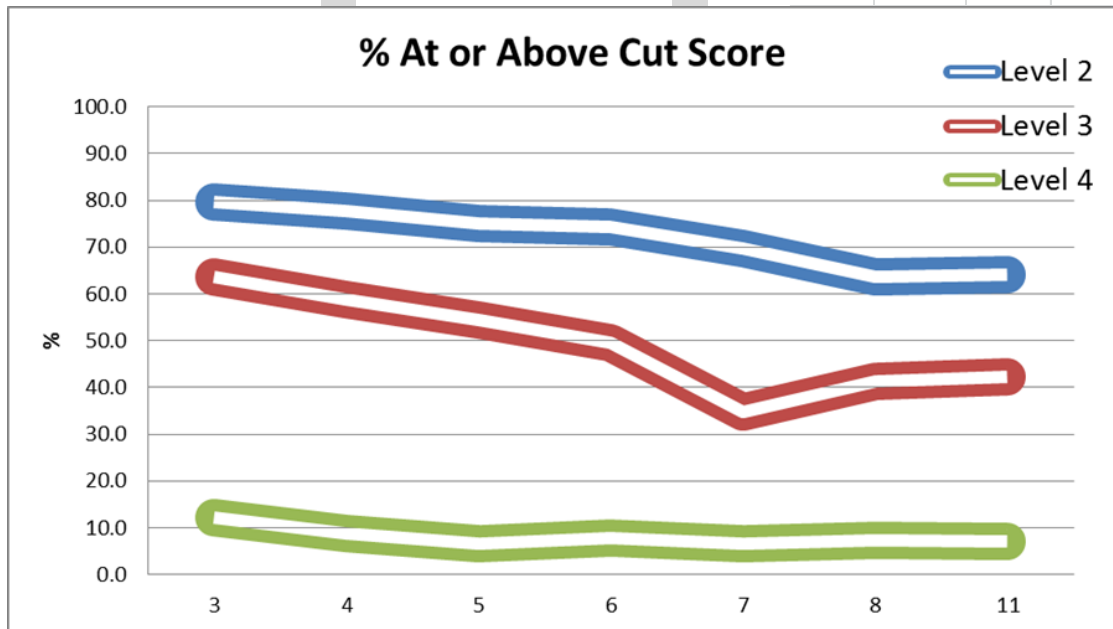


Figure 5. Sample Cross-Grade Review Interactive Spreadsheet

A form of the graphic shown in Figure 5 remained on one screen throughout cross-grade review. Panelists were free to suggest a change to any cut score, and the facilitator would make the appropriate change in the first table (Round 3 Bookmarks), at which time all tables and the graph would update. The ability to see quickly the impact of any change in bookmark placement made the process flow very smoothly. However, changes could not be made simply on the basis of smoothing of the lines in the graph; the person making the motion was required to defend the motion on the basis of the ALDs, the CCCS, and the contents of the test.

After orientation, ELA and math panels separated, each to review cut scores across grades for a single subject. Both panels started with a review of the grade 11 cut scores and then moved on to the cut scores for grades 3-8. The ELA panel made a total of 8 changes; the math panel, 11. Final results of the cross-grade review panel are shown in Table 11. Table 12 shows the magnitude of the changes, relative to Round 3 results. Changes with an absolute value of five percent or greater are highlighted.

Table 11
Percentages of Students at Each Level After Cross-Grade Review

Test	Level 1	Level 2	Level 3	Level 4
Grade 3 ELA	33.5	26.4	21.0	19.1
Grade 4 ELA	35.6	22.4	23.1	18.9
Grade 5 ELA	30.3	22.6	30.2	16.9
Grade 6 ELA	28.7	29.1	29.1	13.1
Grade 7 ELA	31.8	28.1	30.6	9.5
Grade 8 ELA	26.9	29.8	33.1	10.2
Grade 11 ELA	26.7	30.7	31.0	11.6
Grade 3 Math	31.7	29.4	26.8	12.1
Grade 4 Math	27.7	35.8	23.9	12.6
Grade 5 Math	36.5	32.1	17.6	13.8
Grade 6 Math	41.7	32.2	12.1	14.0
Grade 7 Math	46.9	29.9	17.4	5.8
Grade 8 Math	48.7	29.2	14.7	7.4
Grade 11 Math	51.7	26.3	16.2	5.8

Table 12
Magnitude of Changes from Round 3 to Cross Grade Review
(Round 3 Percent Minus Cross-Grade Review Percent)

	Level 1	Level 2	Level 3	Level 4
Grade 3 ELA	0.0	0.0	0.0	0.0
Grade 4 ELA	0.0	0.0	0.0	0.0
Grade 5 ELA	8.9	7.9	-16.8	0.0
Grade 6 ELA	-5.1	5.1	-0.9	0.9
Grade 7 ELA	0.1	-0.1	-2.9	2.9
Grade 8 ELA	3.3	2.8	-6.1	0.0
Grade 11 ELA	-0.6	5.4	-4.8	0.0
Grade 3 Math	-1.0	1.0	-1.3	1.3
Grade 4 Math	5.3	2.9	-5.2	-3.0
Grade 5 Math	0.0	0.0	0.0	0.0
Grade 6 Math	0.0	2.6	-1.3	-1.3
Grade 7 Math	0.0	-3.8	3.8	0.0
Grade 8 Math	0.0	3.5	-3.5	0.0
Grade 11 Math	10.7	-9.1	-1.6	0.0

Each change was brought about by a motion, second, discussion, review of the item on the OIB page associated with the original cut score and those on adjacent pages (including the page associated with the cut score being considered), and a 2/3 majority vote. At the end of the day, each panel passed a

final resolution adopting all cut scores, those revised as well as those not revised or even considered for revision. Results for Level 3 are summarized in Figure 6. In ELA, percentages of students at or above Level 3 were relatively constant – between 40 and 50 percent – with the fifth grade “bump” smoothed out but not eliminated. For math, there was a steady decline from grade 3 to grade 11, starting at about 40 percent at or above Level 3 in grade 3 and declining to about 20 percent in grade 11.

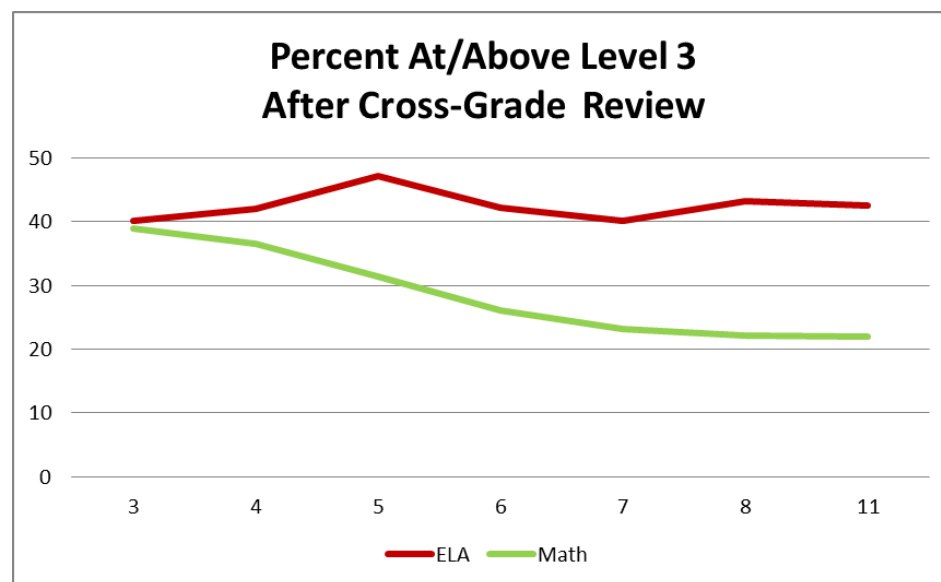


Figure 6. Impact After Cross-Grade Review

Chiefs meeting. As we moved from the October meetings in Dallas to the November meetings in Chicago and Portland, the delineation and clarification of roles became even more important. MI staff and our colleagues at CTB and Hager Sharp were responsible for processing all results from the in-person and online activities and preparing summaries for Smarter Balanced to present to the Chiefs. We prepared presentation materials and back-up data for these meetings, but Dr. Willhoft was the lead spokesperson throughout. Between October 20, the final day of cross-grade review, and November 14, the day of the Chiefs’ vote, there was considerable exchange of ideas and information between MI and Smarter Balanced, making sure that all data were correct, up to date, and clear.

When Chiefs met on November 6 in Chicago, some were unable to vote yea or nay because various groups in their home states were still reviewing the results. The discussion on November 6 therefore focused on how Smarter Balanced could assist the Chiefs in their communications with stakeholder groups.

When Chiefs met in San Diego on November 14, Dr. Willhoft first provided an overview of the process we had carried out, a process the Chiefs had approved on April 30. Dr. Willhoft then provided a context for adopting the recommended cut scores:

1. Honor the work of the panelists who contributed their content expertise to the process
2. Give consideration to external evidence of student readiness for credit-bearing college work:
Notably NAGB Grade 12 college preparedness data
3. Maintain a “system perspective” by using information from all grades (and both content areas) to make recommendations that support a coherent system

The Chiefs were then presented with three motions:

1. Smarter Balanced member states endorse the position paper “Interpretation and Use of Scores and Achievement Levels” dated November 14, 2014, as amended, and recommend that member states attend to it when considering the use of scores from Smarter Balanced assessments.
2. Smarter Balanced states approve the achievement levels for Grade 11 in Mathematics and English language arts/Literacy as shown in the accompanying table, displayed in Smarter Balanced scale scores. (staff recommendations)
3. Smarter Balanced states approve the achievement levels for Grades 3-8 in Mathematics and English language arts/Literacy as shown in the accompanying table, displayed in Smarter Balanced scale scores. (staff recommendations)

Accompanying these motions were tables and graphs depicting the recommended cut scores and impact. These are summarized in Table 13 and Figure 7. Following a brief discussion, Motion 1 (adoption of the position paper) passed by a vote of 20-0 with no abstentions. Motion 2 (adoption of grade 11 cut scores) passed by a vote of 18-0 with 2 abstentions. Motion 3 (adoption of grades 3-8 cut scores) also passed by a vote of 18-0 with 2 abstentions.

Table 13
Percentages of Students at Each Level After Chiefs Vote

Test	Level 1	Level 2	Level 3	Level 4
Grade 3 ELA	35%	27%	20%	18%
Grade 4 ELA	37%	22%	23%	18%
Grade 5 ELA	33%	23%	29%	15%
Grade 6 ELA	30%	29%	30%	11%
Grade 7 ELA	34%	28%	30%	8%
Grade 8 ELA	28%	31%	32%	9%
Grade 11 ELA	28%	31%	30%	11%
Grade 3 Math	32%	29%	27%	12%
Grade 4 Math	27%	36%	24%	13%
Grade 5 Math	35%	32%	18%	15%
Grade 6 Math	35%	32%	19%	14%
Grade 7 Math	36%	31%	20%	13%
Grade 8 Math	38%	30%	19%	13%
Grade 11 Math	40%	27%	22%	11%

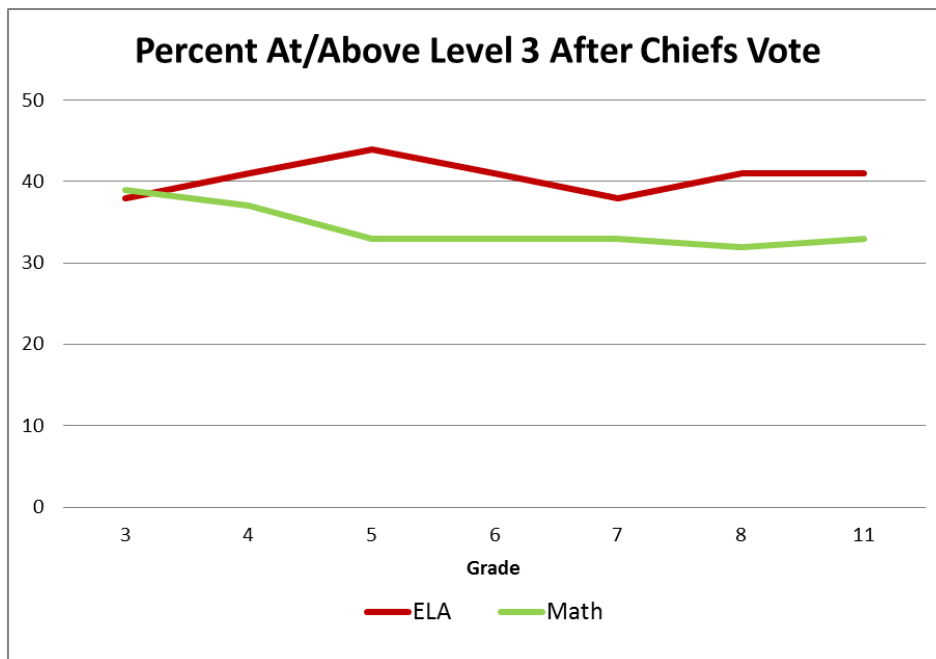


Figure 7. Impact After Chiefs Vote

Summary and Next Steps

By all objective measures, the Smarter Balanced achievement level setting was a great success. The plan approved by the Chiefs in April was faithfully executed in October, and the results were unanimously adopted. States using the Smarter Balanced assessments in the spring of 2015 were able to obtain results with both scale scores and achievement level designations. Those scores and levels can be compared reliably across states and reported with confidence to state and federal agencies. Additional evidence of the success of the project may be found in the Auditors' Report (for a link, see Appendix A, Auditors' Report) and in the endorsements of the TAC and Advisory Panel (for a link, see Appendix A, Statements of Support).

However successful the project may have been, the task of establishing and evaluating cut scores on the Smarter Balanced assessments has not been completed. We have moved into a new era of standard setting for K-12 assessments. The moment we applied the phrase "college and career ready" to the process, we moved from documentation of past accomplishments to certification of future expectations (cf. Bunch, 2013). Achievement level setting for the Smarter Balanced assessments occurred in that context. Inasmuch as scores and levels on Smarter Balanced assessments at least imply predictions of future performance, validation of those predictions is in order.

Sireci (2012) presented a comprehensive research plan and recommendations to validate Smarter Balanced assessments over a period of years. As part of the current project, MI staff also prepared a set of research recommendations specifically for validating cut scores. There will be opportunities following the 2015 operational test administration and during subsequent school years to examine the cut scores with respect to internal variables from the Smarter Balanced assessment system (e.g., interim assessments, formative assessments, digital library, scale scores and levels on subsequent tests) and

targeted external variables (e.g., teacher ratings, grades, scores on other tests, scores of college students on the grade 11 tests, college grades, employer evaluations).

Evaluating the cut scores from these various perspectives will yield important information regarding the appropriateness of the interpretations and uses of the scores and level designations. But it will likely be 2017 at the very least before a critical mass of validation information can be gathered and processed. In the meantime, it is worth noting that Smarter Balanced has met 10 of Hambleton's 11 criteria (Hambleton, 2001) for valid interpretations of results based on the 2014 achievement level setting. Establishment of predictive validity will satisfy #11.

References

- Bunch, M. B. (2012). Practical issues in standard setting. In G. J. Cizek (Ed.) *Setting Performance Standards: Foundations, Methods, and Innovations* (2nd Ed.). New York, NY: Routledge.
- Bunch, M. B. (2013). Setting Cut Scores on 21st Century Tests. Presession presented at the annual meeting of the National Council on Measurement in Education, San Francisco, California.
- Hager Sharp (2014). Key Insights from Informal Focus Group Among Parents. Internal Working Paper dated April 12, 2014.
- Hambleton, R. K. (2001). Setting performance standards on educational assessments and criteria for evaluating the process. In G. J. Cizek (Ed.), *Setting Performance Standards: Concepts, Methods, and Perspectives*. Mahwah, NJ: Lawrence J. Erlbaum.
- Lewis, D.M., Mitzel, H. C., Mercado, R. L. & Schulz, E. M. (2012). The Bookmark standard setting procedure. In G. J. Cizek (Ed.) *Setting Performance Standards: Foundations, Methods, and Innovations* (2nd Ed.). New York, NY: Routledge.
- Peters, T. & Waterman, R. H. (1982). *In Search of Excellence*. New York: Harper Collins.
- Sireci, S. G. (2012). Smarter Balanced Assessment Consortium: Comprehensive Research Agenda. Report presented to Smarter Balanced December 31, 2012 and available online at <http://www.smarterbalanced.org/wordpress/wp-content/uploads/2014/08/Smarter-Balanced-Research-Agenda-Recommendations-2012-12-31.pdf>
- State Of Washington Office Of Superintendent Of Public Instruction (2013), Request For Proposals (RFP): Smarter Balanced RFP No. 21. Olympia, Washington, Author.

Appendix A

Links to Smarter Balanced Achievement Level Setting Documents

About Achievement Level Setting

<http://www.smarterbalanced.org/achievement-levels/>

Achievement Level Setting Infographic

<http://www.smarterbalanced.org/wordpress/wp-content/uploads/2014/09/Smarter-Balanced-Achievement-Level-Setting-Overview-.pdf>

Auditors' Report

<http://www.smarterbalanced.org/wordpress/wp-content/uploads/2014/11/SBAC-Standard-Setting-Report-.pdf>

Ed Week Article (November 17, 2014)

<http://www.edweek.org/ew/articles/2014/11/17/13sbac.h34.html>

Frequently Asked Questions

<http://www.smarterbalanced.org/wordpress/wp-content/uploads/2014/11/Smarter-Balanced-Achievement-Levels-QA.pdf>

Interpretation and Use of Scores and Achievement Levels

<http://www.smarterbalanced.org/wordpress/wp-content/uploads/2014/11/Interpretation-and-Use-of-Scores.pdf>

Panelist Videos

https://www.youtube.com/watch?v=bW_yGf4BB1E

Reporting Achievement Level Descriptors

<http://www.smarterbalanced.org/wordpress/wp-content/uploads/2015/03/Reporting-ALDs-Final.pdf>.

Sireci Research Agenda

http://www.smarterbalanced.org/wordpress/wp-content/uploads/2014/08/Smarter-Balanced-Research-Agenda_Recommendations-2012-12-31.pdf

Smarter Balanced States Approve Achievement Level Recommendations Press Release

<http://www.smarterbalanced.org/news/smarter-balanced-states-approve-achievement-level-recommendations/>

Statements of Support (Auditors, Advisory Panel, and Technical Advisory Committee)

<http://www.smarterbalanced.org/wordpress/wp-content/uploads/2014/11/ALS-Statements-of-Support.pdf>

Appendix B

Materials and Equipment Used for the In-Person Panel Workshop

Table B-1
Training Materials

Product	Purpose	Brief Description
Amplified Agenda	Keep facilitators and panelists on schedule	Complete task-by-task agenda of the workshop
Facilitator Script/Supplement	Provide consistent instruction across panels	Detailed bulleted points for facilitators to make, accompanied by detailed script and reference materials
Table Leader Training PowerPoint	Orient table leaders to their task	34 slides describing ALDs, the Bookmark procedure, and round-by-round tasks
Observer Script	Orient observers to the process and groundrules	6-page description of the schedule and process
In-Person Intro PowerPoint	Orient panelists to the tasks to be performed	35 slides providing background information as well as detailed instructions regarding schedule and logistics
Achievement Level Descriptors	Orient panelists to ALDs	Specially formatted ALDs for each test highlighting expectations of students just barely at each cut score
OIB Training/Introduction to the Bookmark Procedure PowerPoint	Orient panelists to the Bookmark procedure	39 slides showing how the Bookmark procedure translates panelists' bookmarks into cut scores
Links to Smarter Balanced Materials	Orient panelists to ALDs, and Practice Test	All panelists took one Smarter Balanced Practice Test and reviewed other materials on the website.
Cross-Grade Review PowerPoint	Orient panelists to the cross-grade review process	25 slides showing purpose and process of cross-grade review

Table B2
Hardware and Equipment

Shipped from MI	Purchased/Leased On Site
<ul style="list-style-type: none"> • 400 Celeron Lenovo PCs running Windows 7 with wireless internet access • 400 flat-screen 21" monitors • 400 headsets • 2 servers • 12 LCD projectors (1 per break-out room) • 12 projector screens (1 per break-out room) • 3 portable hand sanitizer kiosks 	<ul style="list-style-type: none"> • 1 flat-screen 32" monitor (for a sight-impaired panelist) • 400 internet connections • 4 tripod speaker stands • 4 12" speakers • 2 4-channel mixers • 2 headset microphones • 4 handheld microphones • 1 LCD projector (6500 lumen XGA) and screen (7'6" x 13'4")for large-group instruction • Cabling for sound systems