Appendix C: Michigan Assessment System Participant Groups

This appendix provides more details on the stakeholders and participants involved in the Michigan Assessment System.

Appendix C.1 Michigan Educators

Michigan educators (including classroom teachers from K–12 and higher education, curriculum specialists, and administrators) play a vital role in all phases of the test development process. Committees of Michigan educators write MI-Access test items, review the test specifications, and provide advice on the model or structure for assessing each content area. They also work to ensure that test content and question types align closely with best practices in classroom instruction.

Appendix C.2 Technical Advisory Committee

Michigan's Technical Advisory Committee (TAC) serves as an advisory body to MDE. The TAC provides recommendations on technical aspects of large-scale assessments, including item development, test construction, administration procedures, scoring and equating methodologies, and standard-setting workshops. The TAC also provides guidance on other technical matters, such as practices not already described in the *Standards for Educational and Psychological Testing* (AERA, APA, & NCME, 2014), and continues to provide advice and consultation on the implementation of new assessments and adherence to the federal requirements set forth by the Every Student Succeeds Act. Table C-1 can be referenced for TAC member information.

Table C-1. Technical Advisory Committee

Name	Position	Organization
Dr. Mark Reckase, Chair	Distinguished Professor of Measurement and Quantitative Methods (retired)	Michigan State University
Dr. Damian Betebenner	Senior Associate	National Center for the Improvement of Educational Assessment
Dr. Gregory J. Cizek	Distinguished Professor of Educational Measurement and Evaluation	University of North Carolina, Chapel Hill
Dr. George E. Engelhard, Jr.	Professor Emeritus of Educational Measurement and Policy	University of Georgia
Dr. Christine Carrino Gorowara	Interim Director	Delaware Center for Teacher Education, University of Delaware
Dr. Joseph Martineau	Senior Associate	National Center for the Improvement of Educational Assessment
Dr. Dave Treder	Coordinator of Research, Evaluation, and Assessment	Genesee Intermediate School District, Flint, Michigan

Appendix C.3 Michigan's Division of Educator, Student, and School Supports (DESSS) Advisory Committee

The DESSS Advisory Committee meets quarterly to provide input, ideas, expert advice, and/or recommendations to MDE and DESSS on matters related to assessment and accountability, professional preparation, educator evaluations, assessment policy, and related communications to the field. The committee also meets to keep its respective organizations abreast of changes to the above areas that will affect Michigan's schools and students. The committee comprises representatives from educational agencies, organizations, and representatives from both two-year and four-year colleges and universities across the state. Table C-2 shows the members of the DESSS Advisory Committee.

Table C-2. Division of Educator, Student, and School Supports Advisory Committee

Last Name	First Name	Organization
Anand	Johanna	Michigan Department of Education/Low Incidence Outreach
Arnswald	Jennifer	Michigan Science Teachers Association
Berry	Kathy	Michigan Council of Teachers of Mathematics
Clingman	Cindy	Michigan Reading Association
Cox	Mary	Michigan Council of Teachers of English
Czerwinski	Harvey	Michigan Education Research Association
Dewsbury-White	Kathryn	Michigan Assessment Consortium
DeYoung	Ann	Michigan Elementary and Middle School Principals Association
Flukes	Jonathan	Michigan Education Research Association
Gordon	Casey	MI Council of Teachers of English to Speakers of Other Languages
Greer	Doug	Oakland Area Intermediate School District
Kher	Neelam	Michigan State University
Koekkoek	Matthew	Michigan Association of Administrators of Special Education
Langdon	Thomas	Michigan Association of School Administrators
Mastie	Marge	Washtenaw Intermediate School District - Retired
McIntyre	Rebecca	Michigan Association of Administrators of Special Education
Miller	Kathy	Michigan School Facilitators Network
Trout	Kelly	Ingham Intermediate School District
Vespremi	Stacy	Michigan Association of State and Federal Programs Specialists
Vorenkamp	Ellen	Wayne Regional Educational Services Agency
Zdeb	Wendy	Michigan Association of Secondary School Principals

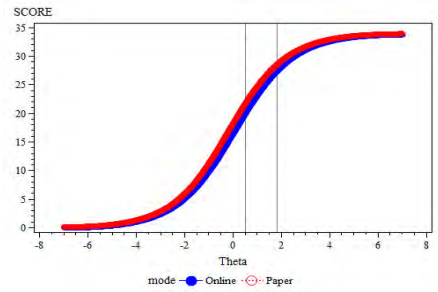
Committee Substitutes

Last Name	First Name	Organization
McGoran	Holly	Michigan Science Teachers Association
Musial	Joe	Wayne Regional Educational Services Agency
Ripmaster	Colin	Michigan Association of Secondary School Principals
Taraskiewicz	Cindy	Wayne Regional Educational Services Agency

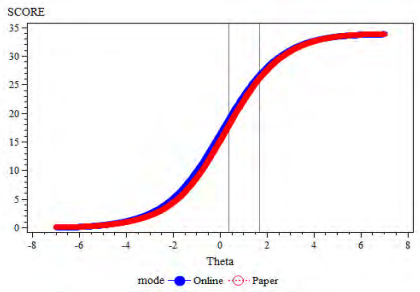
Appendix D: Test Characteristic Curves for MI-Access FI, Spring 2019

Figure D-1. IRT-based TCCs for FI English Language Arts by Grade

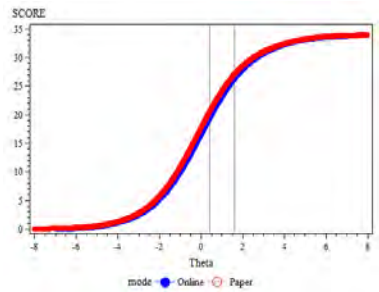
Test Characteristic Curve for FI Grade 3 English Language Arts



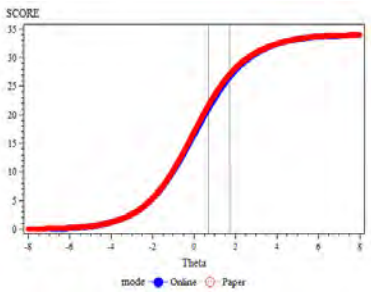
Test Characteristic Curve for FI Grade 4 English Language Arts



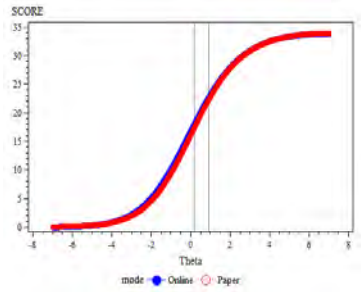
Test Characteristic Curve for FI Grade 5 English Language Arts



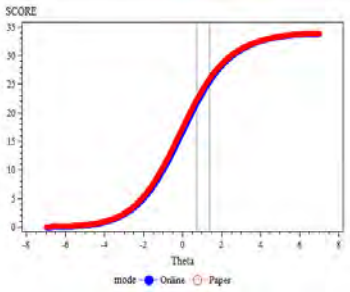
Test Characteristic Curve for FI Grade 6 English Language Arts



Test Characteristic Curve for FI Grade 7 English Language Arts



Test Characteristic Curve for FI Grade 8 English Language Arts



Test Characteristic Curve for FI Grade 11 English Language Arts

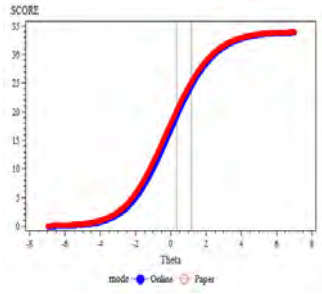
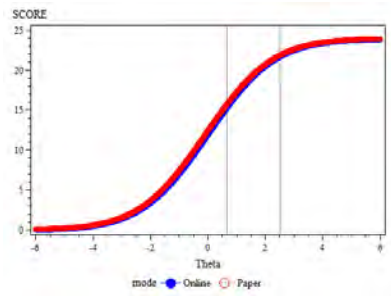
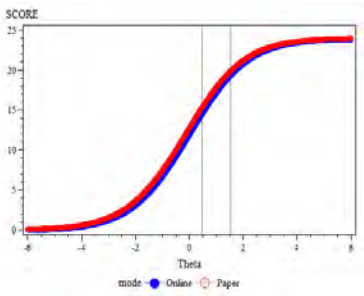


Figure D-2. IRT-based TCCs for FI Mathematics by Grade

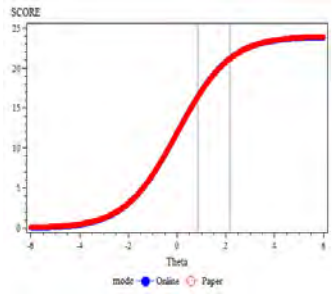
Test Characteristic Curve for FI Grade 3 Mathematics



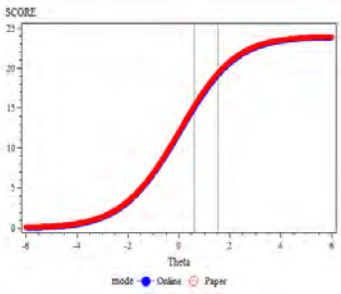
Test Characteristic Curve for FI Grade 4 Mathematics



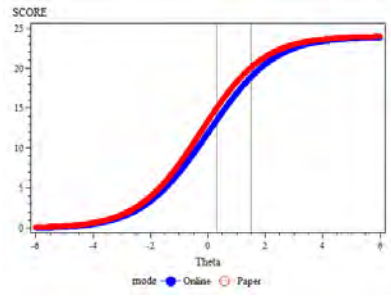
Test Characteristic Curve for FI Grade 5 Mathematics



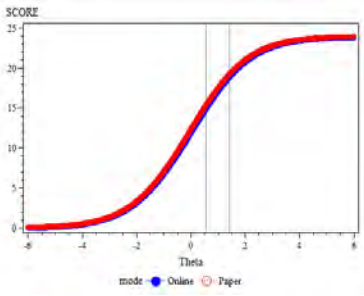
Test Characteristic Curve for FI Grade 6 Mathematics



Test Characteristic Curve for FI Grade 7 Mathematics



Test Characteristic Curve for FI Grade 8 Mathematics



Test Characteristic Curve for FI Grade 11 Mathematics

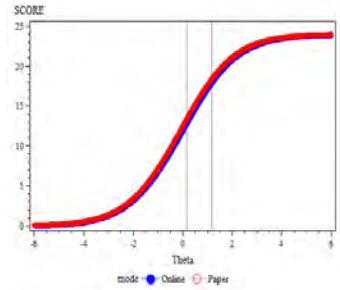
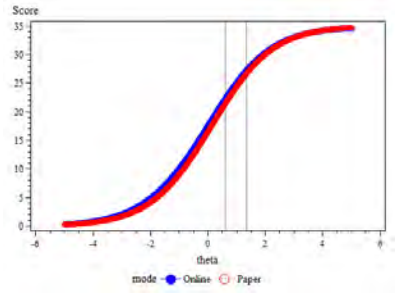
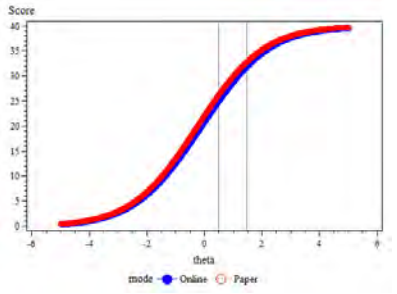


Figure D-3. IRT-based TCCs for FI Science Tests by Grade

Test Characteristic Curve for FI Grade 4 Science



Test Characteristic Curve for FI Grade 7 Science



Test Characteristic Curve for FI Grade 11 Science

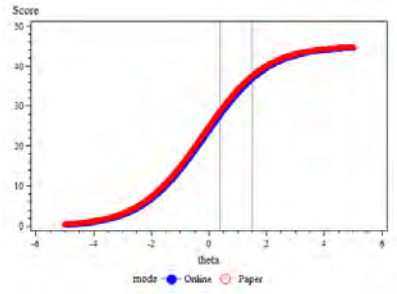
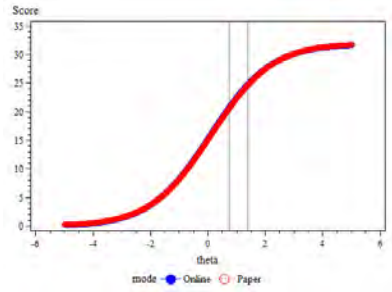
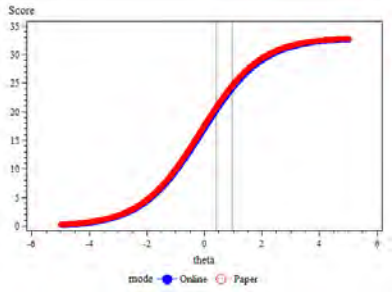


Figure D-4. IRT-Based TCCs for FI Social Studies Tests by Grade

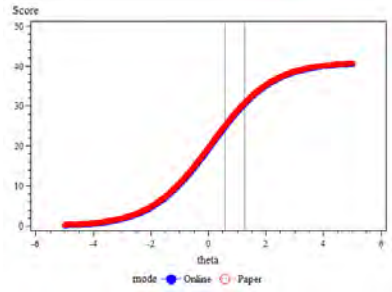
Test Characteristic Curve for FI Grade 5 Social Studies



Test Characteristic Curve for FI Grade 8 Social Studies



Test Characteristic Curve for FI Grade 11 Social Studies



Appendix E: MI-Access Standard Setting Reports

Appendix E.1 MI-Access Standard Setting (2015)

MI-Access Standard Setting
Participation and Supported Independence (June 15-18, 2015)
Functional Independence (June 29-July 2, 2015)

Measurement Incorporated
July 17, 2015

MICHIGAN Department of Education

MI Access™
Michigan's Alternate Assessment Program

Table of Contents

List of Tables and Figures ii
Executive Summary iii
Final Report 1
Planning and Implementation 3
Participation and Supported Independence 3
Functional Independence 12
Results 17
Vertical Articulation 23
Review and Critique 27
Sample Comments 29
Conclusions and Recommendations 31
References 32
Appendix A Participation and Supported Independence Training Materials 33
Appendix B Functional Independence Training Materials 46
Appendix C Evaluation Comments 59
Appendix D PowerPoint Presentations 66

List of Tables and Figures

Table Page

ES-1 Standard Setting Panels iii
ES-2 Round 3 Cut Scores and Impact for Participation v
ES-3 Round 3 Cut Scores and Impact for Supported Independence v
ES-4 Round 3 Cut Scores and Impact for Functional Independence vi
ES-5 Cut Scores and Impact for Participation - After Vertical Articulation vii

ES-6 Cut Scores and Impact for Supported Independence - After Vertical Articulation vii
ES-7 Cut Scores and Impact for Functional Independence - After Vertical Articulation viii
1 Standard Setting Panels 1
2 Demographic Summary: Participation and Supported Independence Panelists 2
3 Demographic Summary: Functional Independence Panelists (N=77) 2
4 MI-Access Participation and Supported Independence Standard Setting Agenda 8
5 MI-Access Functional Independence Agenda 13
6 Round 1 Results for Participation 17
7 Round 1 Results for Supported Independence 18
8 Round 1 Results for Functional Independence 18
9 Round 2 Results for Participation 19
10 Round 2 Results for Supported Independence 19
11 Round 2 Results for Functional Independence 20
12 Round 3 Results for Participation 20
13 Round 3 Results for Supported Independence 21
14 Round 3 Results for Functional Independence 21
15 Cut Scores and Impact for Participation – After Vertical Articulation 24
16 Cut Scores and Impact for Supported Independence – After Vertical Articulation 24
17 Cut Scores and Impact for Functional Independence – After Vertical Articulation 25
18 Summary of Evaluations for Participation and Supported Independence (N=62) 28
19 Summary of Evaluations for Functional Independence (N=76) 28

Figure

1 MI-Access Participation Scoring Document 4
2 MI-Access Supported Independence Scoring Document 5
3 Sample Body of Work for MI-Access Participation Student 6
4 Body of Work Data Entry Sheet 7
5 Bookmark Entry Form 16
6 Round 3 Impact: Participation 22

7 Round 3 Impact: Supported Independence 22
8 Round 3 Impact: Functional Independence 23
9 Impact After Vertical Articulation: Participation..... 26
10 Impact After Vertical Articulation: Supported Independence 26
11 Impact After Vertical Articulation: Functional Independence 27

MI-Access Standard Setting Report
Executive Summary

Measurement Incorporated
July 17, 2015

Measurement Incorporated (MI) assisted the Michigan Department of Education (MDE) in the conduct of standard setting for MI-Access Participation (P), Supported Independence (SI), and Functional Independence (FI) for grades 3-8 plus high school, providing a lead facilitator, panel facilitators, and sufficient psychometric and clerical staff to conduct 16 panel meetings the weeks of June 15-18 and June 29-July 2, 2015.

For all MI-Access assessments, the MDE provides three performance levels:

3 – Emerging toward the standard
2 – Attained the standard
1 – Surpassed the standard

For the P and SI standard-setting activities, MI staff conducted a Body of Work standard-setting procedure with one round of rangefinding and two rounds of pinpointing for eight panels. For the FI standard-setting activity, MI staff conducted a Bookmark standard-setting procedure with three rounds of bookmark placements. Panels are described in Table ES-1.

Table ES-1
Standard Setting Panels

Participation/Supported Independence		Functional Independence	
Panel	Members	Panel	Members
English Language Arts P/SI 3-5	9	English Language Arts FI 3-4	10
English Language Arts P/SI 6-8	6	English Language Arts FI 5-7	9
English Language Arts P/SI 11	7	English Language Arts FI 8, 11	10
Mathematics P/SI 3-5	8	Mathematics FI 3-5	10
Mathematics P/SI 6-8	8	Mathematics FI 6-7	10
Mathematics P/SI 11	7	Mathematics FI 8, 11	9
Science P/SI 4, 7	8	Science FI 4, 7, 11	10
Science P/SI 11	7	Social Studies FI 5, 8, 11	9

Panelists received general instruction in the purpose of the meeting, followed by specific instruction on the tests and the Performance Level Descriptors (PLDs). Instruction on the tests included review of tests and manuals and, for the P/SI panelists, a 45-minute video of test administration and scoring procedures. PLD review consisted of facilitator-led discussion of PLDs with questions and answers. MI staff then provided an overview of the standard-setting procedure, followed by a short practice round to give all panelists an opportunity to practice the method before applying it. After a brief question-and-answer session, panelists indicated their readiness to proceed with Round 1.

Participation/Supported Independence. Panelists followed a Body of Work procedure in which they reviewed sets of 30 student work samples arranged in packets from lowest to highest score. Their task was to compare each work sample with the appropriate PLD and assign that work sample to one of the three levels. In any given session, panelists reviewed two sets of work samples, one for Participation and one for Supported Independence. The panelists entered their ratings on scannable documents.

After Round 1, MI staff collected the scannable documents, scanned them, and calculated cut score regions using graphical methods. They then removed certain work samples that did not contribute to the determination of cut score regions and inserted additional work samples with scores in the region of the Round 1 preliminary cuts.

MI facilitators then shared Round 1 results with panelists, including the distribution of ratings, cut score regions, and preliminary impact data. After a discussion of the results, panelists indicated their readiness for Round 2 and commenced, again entering their ratings of the work samples on scannable forms. As in Round 1, panelists evaluated both P and SI work samples in a session. At the end of the session, MI facilitators collected all materials and dismissed the panels. MI psychometricians then analyzed Round 2 data using logistic regression. In some instances, logistical regression failed to produce a usable result for one of two possible reasons: lack of resolution due to poor model fit; or perfect agreement, which yields no data for logistic regression to process. In those instances, MI staff reverted to the original graphical method.

Facilitators presented Round 2 results, similar to those after Round 1 but with the addition of impact data from 2014 for comparison. Panelists reviewed the results and impact data, indicated their readiness to begin Round 3, and commenced. In Round 3, panelists reviewed the same work samples they had reviewed in Round 2. At the end of Round 3, facilitators collected all materials and dismissed the panel. MI psychometricians then calculated final cut scores using logistic regression as described above.

Functional Independence. Panelists engaged in a Bookmark procedure in which they reviewed ordered item booklets with items arranged in difficulty order from easiest to hardest. Their task was to place two bookmarks: one to note the location of the dividing line between Level 3 and Level 2 students and another to note the location of the dividing line between Level 2 and Level 1 students. Panelists used scannable documents to record their bookmark placements, and MI staff translated those scanned bookmarks into cut scores.

After Round 1, MI staff collected the scannable documents, scanned them, and calculated cut scores as well as distributions of bookmarks. MI facilitators then shared Round 1 results with panelists, including the distribution of bookmarks, cut scores, and preliminary impact data. After a discussion of the results, panelists indicated their readiness for Round 2 and commenced, again entering their bookmarks on scannable forms.

At the end of Round 2, MI facilitators collected all materials and dismissed the panels. MI psychometricians then analyzed Round 2 data as they had done for Round 1. Facilitators presented Round 2 results, similar to those after Round 1 but with the addition of impact data from 2013 for comparison. Panelists reviewed the results and impact data, indicated their readiness to begin Round 3, and commenced. In Round 3, panelists reviewed the items once more as in Round 2. At the end of Round 3, facilitators collected all materials and dismissed the panel. MI psychometricians then calculated final cut scores.

Results are presented in Tables ES-2,ES-3, and ES-4. Cut scores are expressed in terms of raw scores in Tables ES-2 and ES-3 and in scale score terms in Table ES-4.

Table ES-2
Round 3 Cut Scores and Impact for Participation

	Cut Scores		% in Level		
Test	Attained	Surpassed	Emerging	Attained	Surpassed
ELA 3-5	27	44	36.6	33.5	29.9
ELA 6-8	29	45	43.0	33.8	23.2
ELA 11	26	43	27.9	36.6	35.5
Math 3-5	33	46	44.8	23.9	31.3
Math 6-8	24	47	31.3	47.7	21.0
Math 11	27	47	29.9	28.2	41.9
Science 4	46	72	38.8	39.6	21.7
Science 7	44	72	43.2	34.6	22.2
Science 11	48	75	39.5	35.8	24.7

Table ES-3
Round 3 Cut Scores and Impact for Supported Independence

	Cut Scores		% in Level		
Test	Attained	Surpassed	Emerging	Attained	Surpassed
ELA 3-5	29	47	16.8	43.2	40.0
ELA 6-8	37	47	26.4	26.0	47.6
ELA 11	28	43	13.1	27.4	59.4
Math 3-5	32	48	21.7	42.4	35.9
Math 6-8	29	50	20.2	49.1	30.7
Math 11	24	43	15.1	46.8	38.1
Science 4	32	55	11.0	41.1	48.0
Science 7	33	55	16.0	42.6	41.3
Science 11	45	57	31.0	26.4	42.7

Table ES-4
Round 3 Cut Scores and Impact for Functional Independence

	Cut Scores		% in Level		
Test	Attained	Surpassed	Emerging	Attained	Surpassed
ELA 3	2297	2313	30.4	38.1	31.5
ELA 4	2406	2420	29.7	29.8	40.5
ELA 5	2506	2528	30.2	40.9	28.9
ELA 6	2611	2628	29.0	31.9	39.1
ELA 7	2716	2732	29.1	29.5	41.4
ELA 8	2803	2821	13.7	23.3	63.0
ELA 11	3152	3172	19.6	20.6	59.8
Math 3	2308	2323	28.7	21.6	49.7
Math 4	2411	2426	26.6	32.4	40.9
Math 5	2511	2523	32.2	20.6	47.1
Math 6	2607	2626	29.9	39.6	30.5
Math 7	2708	2723	42.1	30.1	27.8
Math 8	2799	2815	34.8	29.7	35.6
Math 11	3094	3116	18.3	36.6	45.2
Science 4	2383	2394	40.4	28.6	31.0
Science 7	2709	2728	44.9	35.3	19.8
Science 11	3160	3188	42.5	36.3	21.2
Social Studies 5	2515	2526	63.3	22.6	14.1
Social Studies 8	2805	2818	56.8	22.4	20.8
Social Studies 11	3164	3182	51.1	24.9	24.0

On the final day of both weeks of standard setting, English language arts (ELA) and Math panelists were regrouped as follows:

- ELA Vertical Articulation Committee (VAC)
- Math Vertical Articulation Committee
- ELA Review and Critique Committee
- Math Review and Critique Committee

The Science and Social Studies panels, given that their grades are noncontiguous, continued through Round 3 and final review on the final day and did not participate in vertical articulation. The P/SI ELA and Math panels had been given the opportunity to advise MDE as to whether articulate by grade span or by individual grade. For Participation, the panels unanimously endorsed articulation by grade span. For Supported Independence, there was some support for articulation by individual grade, but those in favor of articulation by grade span outnumbered those in support of individual-grade articulation by about three to one.

Vertical articulation began with an overview of the process, followed by a question-and-answer period. During this phase of the process, both ELA and Math VACs met together. After the question-and-answer session, ELA and Math VACs separated into different rooms. During the remainder of the day, each VAC reviewed results (cut scores and impact) across grade spans and recommended changes. Changes were

effected by a motion, second, discussion and vote. Given that the changes were to override cut scores set over three rounds of deliberation, a 2/3 majority was required to pass any motion. The two P/SI committees made a total of three changes (one for ELA and two for Math). The two FI committees made a total of five changes three for ELA and two for Math). Results are depicted in Tables ES-5, ES-6, and ES-7. Highlighted entries in these tables indicate changes, relative to Round 3 (Tables ES-2, ES-3, and ES-4). Science and Social Studies cut scores and impact are included even though they were not subject to vertical articulation.

Table ES-5
Cut Scores and Impact for Participation – After Vertical Articulation

	Cut Scores		% in Level		
Test	Attained	Surpassed	Emerging	Attained	Surpassed
ELA 3-5	27	44	36.6	33.5	29.9
ELA 6-8	29	45	43.0	33.8	23.2
ELA 11	26	43	27.9	36.6	35.5
Math 3-5	33	46	44.8	23.9	31.3
Math 6-8	26	47	35.9	43.1	21.0
Math 11	27	47	29.9	28.2	41.9
Science 4	46	72	38.8	39.6	21.7
Science 7	44	72	43.2	34.6	22.2
Science 11	48	75	39.5	35.8	24.7

Table ES-6
Cut Scores and Impact for Supported Independence – After Vertical Articulation

	Cut Scores		% in Level		
Test	Attained	Surpassed	Emerging	Attained	Surpassed
ELA 3-5	29	47	16.8	43.2	40.0
ELA 6-8	34	47	19.5	32.9	47.6
ELA 11	28	43	13.1	27.5	59.4
Math 3-5	32	49	21.7	48.5	29.8
Math 6-8	29	50	20.2	49.1	30.7
Math 11	24	43	15.1	46.8	38.1
Science 4	32	55	11.0	41.1	48.0
Science 7	33	55	16.0	42.6	41.3
Science 11	45	57	31.0	26.4	42.7

Table ES-7
Cut Scores and Impact for Functional Independence – After Vertical Articulation

	Cut Scores		% in Level		
Test	Attained	Surpassed	Emerging	Attained	Surpassed
ELA 3	2297	2312	30.4	35.7	33.9
ELA 4	2406	2420	29.7	29.8	40.5
ELA 5	2506	2523	30.2	32.0	37.8
ELA 6	2611	2628	29.0	31.9	39.1
ELA 7	2713	2732	26.0	32.6	41.4
ELA 8	2803	2821	13.7	23.3	63.0
ELA 11	3152	3172	19.6	20.6	59.8
Math 3	2308	2323	28.7	21.6	49.7
Math 4	2411	2426	26.6	32.4	40.9
Math 5	2511	2529	32.2	32.8	35.0
Math 6	2607	2626	29.9	39.6	30.5
Math 7	2704	2723	30.4	41.8	27.8
Math 8	2799	2815	34.8	29.7	35.6
Math 11	3094	3116	18.3	36.6	45.2
Science 4	2383	2394	40.4	28.6	31.0
Science 7	2709	2728	44.9	35.3	19.8
Science 11	3160	3188	42.5	36.3	21.2
Social Studies 5	2515	2526	63.3	22.6	14.1
Social Studies 8	2805	2818	56.8	22.4	20.8
Social Studies 11	3164	3182	51.1	24.9	24.0

The review and critique sessions occurred at the same time as the vertical articulations. In each panel, facilitators appointed roughly half the panelists to the VAC and the other half to the review/critique committee. The purpose of the review/critique committees was to evaluate the process and provide feedback to the Department regarding improvements in future standard-setting activities. While the overall tone of the sessions was very positive, several excellent suggestions were received.

Panelists evaluated the process and their facilitators on eight critical-incident factors, each on a 2-point scale (Agree/Disagree). With regard to facilitators and process, 98-100 percent of panelists agreed with each statement. With regard to facilities and food, reaction was mixed, with 43 percent of P/SI panelists and 53 percent of FI panelists agreeing that the facilities and food service helped to create a good working environment.

Conclusion and Recommendation

The process for arriving at cut scores for both standard-setting activities was rigorous and consistent with best practices and overseen by highly competent practitioners. The resulting cut scores and corresponding impacts were reasonably consistent across grade spans or individual grades as well as with historical trends in Michigan for these populations. It is our recommendation that the cut scores be adopted without modification or adjustment.

Introduction

Measurement Incorporated (MI) assisted the Michigan Department of Education (MDE) in the conduct of standard setting for MI-Access Functional Independence (FI), Supported Independence (SI), and Participation (P) for grades 3-8 plus high school. Specifically, MI provided a lead facilitator, panel facilitators, and sufficient, psychometric, and clerical staff to conduct eight panel meetings the week of June 15-18, 2015, and eight panel meetings the week of June 29-July 2, 2015.

For all MI-Access assessments, the MDE provides three performance levels:

3 – Emerging toward the standard
2 – Attained the standard
1 – Surpassed the standard

For the P and SI standard-setting activities, MI staff conducted a Body of Work procedure with one round of rangefinding and two rounds of pinpointing for eight panels. For the FI standard-setting activity, MI staff conducted a Bookmark procedure with three rounds of bookmark placements. Panels are described in Table 1. Their demographic characteristics are summarized in Tables 2 and 3.

Table 1
Standard Setting Panels

Participation/Supported Independence		Functional Independence	
Panel	**Members**	**Panel**	**Members**
English Language Arts P/SI 3-5	9	English Language Arts FI 3-4	10
English Language Arts P/SI 6-8	6	English Language Arts FI 5-7	9
English Language Arts P/SI 11	7	English Language Arts FI 8, 11	10
Mathematics P/SI 3-5	8	Mathematics FI 3-5	10
Mathematics P/SI 6-8	8	Mathematics FI 6-7	10
Mathematics P/SI 11	7	Mathematics FI 8, 11	9
Science P/SI 4, 7	8	Science FI 4, 7, 11	10
Science P/SI 11	7	Social Studies FI 5, 8, 11	9

Table 2
Demographic Summary: Participation and Supported Independence Panelists

Gender		Region	
Female	57	1	5
Male	5	2	18
		3	15
Ethnicity		4	5
Black Nonhispanic	5	5	16
White Nonhispanic	50	Missing	3
No Answer	7		
Other	0	**School Type**	
		Urban	13
Teaching Experience		Suburban	18
1-5 years	13	Rural	21
6-10 years	9	Other	5
11-20 years	23	Missing	5
20+ years	7		
Missing	10		

Table 3
Demographic Summary: Functional Independence Panelists (N=77)

Gender		Region	
Female	65	0	5
Male	12	1	4
		2	16
Ethnicity		3	14
Black Nonhispanic	11	4	15
Hispanic	2	5	23
White Nonhispanic	54		
No Answer	7	**School Type**	
Other	3	Urban	23
		Suburban	24
Teaching Experience		Rural	25
1-5 years	11	Other	5
6-10 years	18		
11-20 years	35		
20+ years	13		

Planning and Implementation

MI submitted a detailed plan to MDE and modified it in response to comments from the Technical Advisory Committee (TAC). The plan called for two four-day meetings, one the week of June 15-18, 2015 for Participation and Supported Independence, and one the week of June 29-July 2, 2015 for Functional Independence. The plan called for application of a Body of Work procedure (Kingston & Tiemann, 2012) for the P and SI event, given that tests were composed primarily of performance tasks, and a Bookmark procedure (Lewis, Mitzel, Mercado, & Schulz, 2012) for the FI event, given the fact that the tests were almost entirely selected response and the items were scaled with the Rasch model. Details of the plan and its execution are provided below.

Participation and Supported Independence

The nature of the assessments for P and SI (portfolio) lends itself to the body of work procedure (Kingston & Tiemann, 2012). This procedure requires panelists to sort work samples into categories based on performance levels. Panelists sort a preliminary collection of student work samples, ordered by total score, to identify regions in which cut scores might be located in a process known as rangefinding. After rangefinding, some of the original work samples may be removed and replaced by different work samples with scores within the regions identified during the rangefinding round. This subsequent round is often referred to as pinpointing. One or more pinpoint rounds may be employed. For this activity, one round of rangefinding and two rounds of pinpointing were applied. After the final round of item review, MI conducted a vertical articulation, engaging representatives of each grade level for each content area to examine all cut scores across all grades and recommend changes.

Planning. In planning for this set of panel activities, we made the following assumptions:

1. Teachers score the P and SI assessments as students respond to them; therefore, there would be no need for additional scoring after online and paper documents are collected. However, there was a need for MDE verification of samples of teacher-rendered scores as a validity check. Time for that activity was built into the overall project schedule.

2. Because P and SI assessments are administered to grade bands, rather than to single grades, a single cut score per grade band was considered sufficient.

3. Science panels would begin reviewing and revising cut scores within and across grades on the afternoon of June 17. They would continue on June 18 and wrap up and evaluate the process on June 18. Math and ELA panels would meet on June 18 to review cross-grade cuts and impact and make adjustments as necessary.

Bodies of work. The MI-Access Participation exam consists of a series of activities administered by a trained educator (primary administrator) with the assistance of a "shadow administrator." For Participation students, each task is scored on a 0-3 scale, as shown in Figure 1. For Supported Independence students, each task is scored on a 0-2 scale, as shown in Figure 2. In all instances, the total score for any student on any item is the sum of the scores entered by the two administrators. Thus, for Participation students, scores for each item can range from 0 to 6, while for Supported Independence students, scores for each item can range from 0 to 4.

MI-Access Participation Scoring Document – SPRING 2015

Combined Primary and Shadow Assessment Administrator
Content Area: Mathematics Form A Grades 6-8

Item	**3** Responds correctly with no administrator assistance	**2** Responds correctly after assessment administrator provides verbal/physical cues	**1** Responds correctly after assessment administrator provides modeling, short of hand-over-hand assistance	**A** Incorrect Response	**B** Resists/ Refuses	**C** Assessment administrator provides hand-over-hand assistance and/or step-by-step directions
1						
2						
3						
4						
5						
6						
7						
8						
9						
10						

Figure 1. MI-Access Participation Scoring Document

MI-Access Supported Independence Scoring Document- SPRING 2015

☐ **Primary Assessment Administrator** ☐ **Shadow Assessment Administrator**

Student Name_____

Optional: print and place the student's bar code label here.
Note: this document will **not** be returned with secure materials.

Content Area: English Language Arts Form:_____ **Grade:** _____

Directions: Use this guide to score the student using the MI-Access rubric for each item of the assessment using a check mark or other mark in the corresponding box for each item. Please be careful that your scores from this sheet are transferred to the correct numbered item on the student online answer document.

Item	**2** – Responds correctly with no assessment administrator assistance	**1** – Responds correctly after assessment administrator provides verbal/physical cues	**A** – Incorrect Response	**B** – Resists/ Refuses	**C** – Assessment administrator provides hand-over-hand assistance and/or step-by-step directions
1					
2					
3					

Figure 2. MI-Access Supported Independence Scoring Document

MI staff worked closely with MDE staff to identify scored documents from the spring 2015 administration such that scores from 0 to perfect or very nearly so for each grade or grade span were included among the work samples. MI staff then translated those scored documents into worksheets panelists used to evaluate the performance levels of the students whose work was represented thereon. Each worksheet included not only the scores for each item but the total score and the average score for each item. The purpose of the total score was to give panelists a clear indication of the total performance of the student; i.e., the body of work for that student. The purpose of the average score for each item was to help panelists place item-level performance for each student in a larger context of how students statewide had performed on that item. A sample body of work is shown in Figure 3. After reviewing each work sample, panelists entered their evaluation on a form similar to that shown in Figure 4.

2015 Standard Setting

Work Sample: 1

Total Score: 2

Item Number	Student Score	Average Item Score
1	0	3.15
2	0	3.85
4	0	2.18
5	0	3.44
6	0	3.25
8	0	3.24
10	0	3.02
12	0	3.76
14	2	4.50
15	0	3.80

Barcode: 3288562600

Program: MI-Access P

Subject: ELA

Grade: 11

Figure 3. Sample Body of Work for MI-Access Participation Student

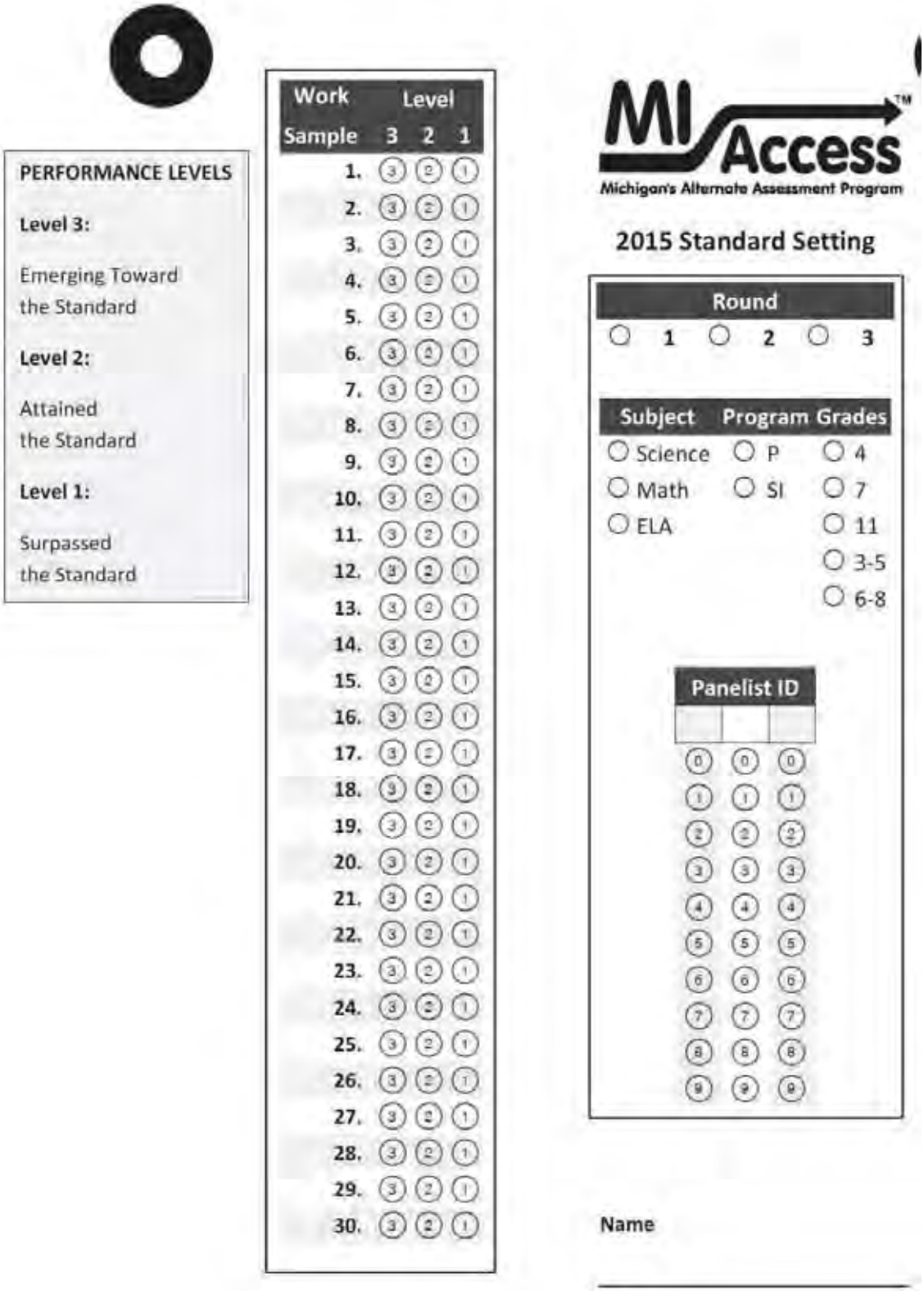


Figure 4. Body of Work Data Entry Sheet

Training materials. MI prepared materials for an opening session that included the goals and tasks of the session as well as a PowerPoint presentation on the body of work procedure. In addition, panel facilitators prepared grade/subject-specific materials that helped panelists understand the nature of the tests and factors affecting performance. Performance level descriptors (PLDs) were developed by MDE test development and curriculum teams. All training materials and forms were submitted to the MDE for review and approval prior to implementation. These materials are listed below and included in full in Appendix A. PowerPoint presentations are included in Appendix D.

- Overview (PowerPoint)
- Body of Work (PowerPoint)
- Facilitator Script
- Body of Work Refresher Notes
- Body of Work Practice Round Form
- Body of Work Entry Form – Round 1
- Body of Work Entry Form – Rounds 2 and 3
- Readiness Form
- Process Evaluation Form

Agenda. Table 4 shows the day-by-day agenda for the four-day event. Facilitator names are included in **bold type**.

Table 4
MI-Access Participation and Supported Independence Standard Setting

Day/ Session	Panel 1: Science P 4, 7; Science SI 4, 7 **Corey Palermo**	Panel 2: Science P 11; Science SI 11 **Karen Kemp**	Panel 3: Math P 3-5; Math SI 3-5 **Winnie Reid**	Panel 4: Math P 6-8; Math SI 6-8 **Christina Luke**	Panel 5: Math 11 P, SI **Dan Bowen**	Panel 6: ELA P 3-5; ELA SI 3-5 **Tom Kelsh**	Panel 7: ELA P 6-8; ELA SI 6-8 **Craig Deville**	Panel 8: ELA 11 P, SI **Tracy Robertson**
June 15								
• **7:30 a.m.**	**Breakfast/Registration**							
• 8:30	Overview and Charge (large group) **Bunch**							
• 9:15	Test Review and PLDs (by Panel in breakout rooms)							
• **10:30**	**Break**							
• 10:45	P/SI Scoring (large group) **Video**							
• 11:30	Introduction to the Body of Work Procedure (large group) **Bunch**							
12:15 p.m.	**Lunch**							

• 1:00 p.m.	BoW Practice	BoW Practice	BoW Practice	BoW Practice	BoW Practice	BoW Practice	BoW Practice	BoW Practice
• 1:45	S4P R1	S11P R1	M3-5P R1	M6-8P R1	M11P R1	E3-5P R1	E6-8P R1	E11P R1
• 3:15	S7P R1							
• 4:45	Wrap Up	Wrap Up	Wrap Up	Wrap Up	Wrap Up	Wrap Up	Wrap Up	Wrap Up
• 5:00	Dismiss	Dismiss	Dismiss	Dismiss	Dismiss	Dismiss	Dismiss	Dismiss
June 16								
• **7:30 a.m.**	**Breakfast/Registration**							
• 8:30	Brief review	Brief review	Brief review	Brief review	Brief review	Brief review	Brief review	Brief review
• 9:00	S4SI R1	S11SI R1	M3-5SI R1	M6-8SI R1	M11SI R1	E3-5SI R1	E6-8SI R1	E11SI R1
• 10:30	S7SI R1							
Noon	**Lunch**							
• 12:45 p.m.	R1 Results	R1 Results	R1 Results	R1 Results	R1 Results	R1 Results	R1 Results	R1 Results
• 1:30	S4P R2	S11P R2	M3-5P R2	M6-8P R2	M11P R2	E3-5P R2	E6-8P R2	E11P R2
• 3:00	S4SI R2							
•4:45	Wrap Up	Wrap Up	Wrap Up	Wrap Up	Wrap Up	Wrap Up	Wrap Up	Wrap Up
•5:00	Dismiss	Dismiss	Dismiss	Dismiss	Dismiss	Dismiss	Dismiss	Dismiss
June 17								
•**7:30 a.m.**	**Breakfast/Registration**							
•8:30	R1 Results	R1 Results	R1 Results	R1 Results	R1 Results	R1 Results	R1 Results	R1 Results
•9:15	S7P R2	S11SI R2	M3-5SI R2	M6-8SI R2	M11SI R2	E3-5SI R2	E6-8SI R2	E11SI R2
•10:30	S7SI R2							
• **Noon**	**Lunch**							
•12:45 p.m.	Review; Revise P	Review; Revise P	Review; Revise	Review; Revise	Review; Revise	Review; Revise	Review; Revise	Review; Revise
•4:45	Wrap Up[1]	Wrap Up[1]	Wrap Up[1]	Wrap Up[1]	Wrap Up[1]	Wrap Up[1]	Wrap Up[1]	Wrap Up[1]
•5:00	Dismiss	Dismiss	Dismiss	Dismiss	Dismiss	Dismiss	Dismiss	Dismiss
June 18								
•**7:30 a.m.**	**Breakfast/Registration**							
•8:30	Review; Revise SI	Review; Revise SI	Vertical Articulation Training for half of Panels 3-8 **Bunch**					
			Math Debrief/Critique **Bowen/Reid**			ELA Debrief/Critique **Kelsh**		

• 9:15			Math Vertical Articulation **Deville/Luke**	ELA Vertical Articulation **Bunch/Robertson**
			Math Debrief/Critique	ELA Debrief/Critique
• **Noon**	**Lunch**			
• 12:45 p.m.	Wrap Up; Evaluate	Wrap Up; Evaluate	Math Vertical Articulation	ELA Vertical Articulation
			Math Debrief/Critique	ELA Debrief/Critique
• 4:45			Wrap Up; Evaluate	Wrap Up; Evaluate
• 5:00	Dismiss	Dismiss	Dismiss	Dismiss

[1] During Wrap-Up on June 17, panelists received information regarding their June 18 room assignments. The two Science panels (1 and 2) returned to their same breakout rooms. For Math and ELA panels (3-8), half the panelists were selected to participate in vertical articulation, and the other half were selected to participate in a final critique of the process. Within each panel, the facilitator selected half the panelists at each table to go to vertical articulation, and the other half to go to the critique.

Conduct of the meeting. Dr. Bunch provided an overview of the four days and gave the panels their charge (see Appendix A). Afterwards, panelists dispersed to their breakout rooms to review the P and SI tests under the direction of the facilitators listed in Table 2. These same facilitators also led the panelists in a review of the PLDs. After lunch on June 15, Dr. Bunch provided an overview of the Body of Work procedure (see Appendix A). Panelists then dispersed to their breakout rooms for Body of Work practice with a small set of P or SI Scoring Documents. Following this practice round, facilitators answered questions, and determined readiness to begin Round 1 by administering and reviewing the Round 1 Readiness Form (see Appendix A). Dr. Bunch and MDE staff circulated among the eight panel rooms throughout each day to observe and answer questions. At the end of each day, MI and MDE staff met for approximately one hour to debrief the day's activities and outline the next day's activities.

Panelists worked in small groups of 3-4 within a room of 7-9. They consulted with others at their table during each round. One panel (Science grades 4 and 7) had two different sets of tests to review (four tests in all); therefore, their schedule was a bit different from those of the remaining panels. Although most other panels had multiple grades to consider (e.g., ELA P 3-5/ELA SI 3-5), those panelists actually had only two tests to consider for a single grade band. For example, the ELA 3-5 panel had to set two cut scores for the Participation test that would apply to all three grades.

Each panel completed Round 1 for all tests before beginning Round 2 for any test. Review materials consisted of a packet of 30 completed Scoring Documents (see Figures 1 and 2) arranged from lowest to highest score. Their task was to assign each completed Scoring Document to one of the following three levels using the PLDs:

3 – Emerging
2 – Attained
1 - Surpassed

Panelists were free to discuss any Scoring Document with others at their tables, but the entries

they made had to be their own, not that of the table. They entered their ratings on forms similar to that shown in Figure 4.

After Round 1, MI staff analyzed the ratings and identified regions where cut scores might be, using the standard rangefinding procedure associated with Body of Work (cf. Cizek & Bunch, 2007, Ch. 9). Scoring Documents that did not contribute to the identification of a cut score were eliminated from the set, and additional Scoring Documents with scores in the regions identified in Round 1 as possible cut scores were inserted to make up the Round 2 packets.

In Round 2, panelists rated the Scoring Documents as in Round 1, assigning each to one of the three performance levels, using the PLDs. As they completed Round 2, they turned in their completed rating sheets, and MI staff calculated cut scores for Levels 2 and 1, using logistic regression as described in Cizek & Bunch (2007, Ch. 9).

On the final day of the meeting, the Science panels continued to review work samples. The ELA and Math panels were divided into two groups: one for vertical articulation and another for overall critique and evaluation of the process. Assignment to these groups was done by the panel facilitators, who took demographics, overall participation, and other factors into consideration. Dr. Bunch provided an introduction to vertical articulation (see Appendix A) and gave the panelists their charge. They then divided by subject and conducted separate reviews for ELA and Math.

The vertical articulation facilitators (Drs. Bunch and Deville) presented displays of data depicting the Round 3 results in terms of cut scores, percent of students at or above each cut score, and percent of students in each category (Emerging, Attained, and Surpassed). Panelists also had access to all test materials they had used during the three rounds of standard setting. The process for changing any cut score were as follows:

1. Motion from the floor to make a specific cut score change (e.g., change the Math 6-8 Participation Level 2 cut score from 24 to 26)
2. Second to the motion
3. Discussion
4. Vote

For the vote, a 2/3 majority was required for passage inasmuch as the action effectively overrode the work of three rounds of panel activity. As panelists recommended changes, the facilitator would enter the new cut score, and the remaining tables and graphic on the display would update so that panelists could see the immediate impact of the change. The two facilitators kept the discussion focused on the PLDs and the relationship between the new cut score and the performance level.

Meanwhile, the remaining ELA and Math panelists gathered in two separate rooms to critique the process and provide feedback to the MDE regarding the manner in which the standard setting was conducted. MDE plans to use this feedback not only in documenting this standard

setting but in planning for future events.

Functional Independence

Test booklets for Functional Independence are similar to those for MI-STEP (i.e., a combination of selected and constructed response items for FI ELA and selected response items only for FI Math, Science and Social Studies). The numbers of students taking the FI assessments are sufficient to calibrate the items using item response theory. Therefore, the Bookmark procedure (Lewis, Mitzel, Mercado, & Schulz, 2012) was not only appropriate but clearly indicated. In this procedure, panelists review test items from easiest to most difficult and identify points in the ordered item booklet where students at the threshold of a given performance level would cease to have a reasonable chance of answering correctly. This level is typically 50 or 67 percent, depending on the nature of the test and the judgment of the technical advisory committee (TAC). For this particular application, the threshold level was set at 67 percent. After the third round of item review, MI conducted a vertical articulation, engaging representatives of each grade level for each content area to examine all cut scores across all grades and recommend changes.

MDE constructed and administered the tests and carried out item calibration. It was then necessary to calibrate for each selected-response item and each score point for each constructed-response item a theta value associated with a fixed probability (.67) of answering each selected-response item correctly or achieving that particular score or better on each constructed-response item. These theta estimates were then used to order selected-response items and score points of constructed-response items from easiest to most difficult in order to construct an ordered item booklet (OIB) for each assessment. MDE conducted all necessary item calibrations and construct all OIBs based on input and requirements developed with MI. MI staff reviewed the item calibrations and the OIBs prior to on-site standard setting. MI staff prepared training materials and made copies of the OIBs.

Training materials. MI prepared materials for an opening session that included the goals and tasks of the session as well as a PowerPoint presentation on the Bookmark procedure. In addition, panel facilitators prepared grade/subject-specific materials that helped panelists understand the nature of the tests and factors affecting performance. All training materials were submitted to the MDE for review and approval prior to implementation. Training materials are listed below and included in Appendix B. PowerPoint presentations are included in Appendix D.

- Overview (PowerPoint)
- Bookmark Training (PowerPoint)
- Facilitator Script
- Body of Work Refresher Notes
- Bookmark Practice Round Form

- Bookmark Entry Form – Round 1
- Bookmark Entry Form – Rounds 2 and 3
- Readiness Form
- Process Evaluation Form

Agenda. Table 5 shows the day-by-day agenda for the four-day event.

Table 5
MI-Access Functional Independence Agenda

Day/Session	Panel 9: Math Grades 3, 4, 5 **Craig Deville**	Panel 10: Math Grades 6, 7 **Jennifer Bowen**	Panel 11: Math Grades 8, High School **Lidia Martinez**	Panel 12: ELA Grades 3-4 **Dan Bowen**	Panel 13: ELA Grades 5, 6, 7 **Jeff Barker**	Panel 14: ELA Grades 8, High School **Job Thomas**	Panel 15: Science Grades 4, 7, 11 **Steve Cramer**	Panel 16: Social Studies Grades 5, 8, 11 **Stephanie Lai**
June 29								
• 7:30 a.m.	**Breakfast/Registration**							
• 8:30	Overview and Charge (large room) **Bunch**							
• 9:15	Test Review (by Panel in breakout rooms)							
• 10:30	**Break**							
• 10:45	PLD Review (by Panel in breakout rooms)							
• 11:30	Introduction to the Bookmark Procedure **Bunch**							
• 12:15 p.m.	**Lunch**							
1:00	Bookmark Practice (by Panel in breakout rooms							
1:45	M3R1	M6R1	M8R1	E3R1	E5R1	E8R1	S4R1	SS5R1
4:45	Wrap-Up	Wrap-Up	Wrap-Up	Wrap-Up	Wrap-Up	Wrap-Up	Wrap-Up	Wrap-Up
5:00	Dismiss	Dismiss	Dismiss	Dismiss	Dismiss	Dismiss	Dismiss	Dismiss
June 30								
• 7:30 a.m.	**Breakfast/Registration**							
• 8:30	Brief Review	Brief Review	Brief Review	Brief Review	Brief Review	Brief Review	Brief Review	Brief Review
• 9:00	M4R1	M7R1	MHSR1	E4R1	E6R1	EHSR1	S7R1	SS8R1
• 10:30	M5R1				E7R1			
• Noon	**Lunch**							
• 12:45 p.m.	M3R1 Results	M6R1 Results	M8R1 Results	E3R1 Results	E5R1 Results	E8R1 Results	S11R1	SS11R1
• 1:30	M3R2	M6R2	M8R2	E3R2	E5R2	E8R2		
• 4:45	Wrap-Up	Wrap-Up	Wrap-Up	Wrap-Up	Wrap-Up	Wrap-Up	Wrap-Up	Wrap-Up
• 5:00	Dismiss	Dismiss	Dismiss	Dismiss	Dismiss	Dismiss	Dismiss	Dismiss

Day/Session	Panel 9: Math Grades 3, 4, 5 **Craig Deville**	Panel 10: Math Grades 6, 7 **Jennifer Bowen**	Panel 11: Math Grades 8, High School **Lidia Martinez**	Panel 12: ELA Grades 3-4 **Dan Bowen**	Panel 13: ELA Grades 5, 6, 7 **Jeff Barker**	Panel 14: ELA Grades 8, High School **Job Thomas**	Panel 15: Science Grades 4, 7, 11 **Steve Cramer**	Panel 16: Social Studies Grades 5, 8, 11 **Stephanie Lai**
July 1								
•7:30a.m.	**Breakfast/Registration**							
•8:30	MR4R1 Results	M7R1 Results	MHSR1 Results	E4R1 Results	E6R1 Results	E8R1 Results	S4R1 Results S7R1 Results	SS5R1 Results SS8R1 Results
•9:15	M4R2	M7R2	MHSR2	E4R2	E6R2	EHSR2	S4R2 S7R2	SS5R2 SS8R2
•10:30	M5R1 Results				E7R1 Results			
•10:30	M5R2				E7R2			
• Noon	**Lunch**							
•12:45 p.m.	M3R2 Results M4R2 Results M5R2 Results	M6R2 Results M7R2 Results	M8R2 Results MHSR 2 Results	E3R2 Results E4R2 Results	E5R2 Results E6R2 Results E7R2 Results	E8R2 Results EHSR2 Results	S11R1 Results	SS11R1 Results
•2:15	M3R3 M4R3 M5R3	M6R3 M7R3	M8R3 MHSR3	E3R3 E4R3	E5R3 E6R3 E7R3	E8R3 EHSR3	S11R2	SS11R2
•4:45	Wrap-Up[1]	Wrap-Up[1]	Wrap-Up[1]	Wrap-Up[1]	Wrap-Up[1]	Wrap-Up[1]	Wrap-Up[1]	Wrap-Up[1]
•5:00	Dismiss	Dismiss	Dismiss	Dismiss	Dismiss	Dismiss	Dismiss	Dismiss
•								
July 2								
•7:30a.m.	**Breakfast/Registration**							
•8:30a.m.	Vertical articulation Training for half of Panels 9-14 **Bunch**						S4R2 Results S7R2 Results	SS5R2 Results SS8R2 Results
	Math Debrief/Critique **Bowen/Martinez**			ELA Debrief/Critique **McClintock**				
•9:15	Math Vertical articulation **Deville/Barker**			ELA Vertical articulation **Bunch/Thomas**			S4R3 S7R3	SS5R3 SS8R3
	Math Debrief/Critique			ELA Debrief/Critique				
• Noon	**Lunch**							
•12:45 p.m.	Math Vertical articulation			ELA Vertical articulation			S11R2 Results	SS11R2 Results
	Math Debrief/Critique			ELA Debrief/Critique				
•2:00	Math Vertical articulation			ELA Vertical articulation			S11R3 Wrap-	SS11R3 Wrap-Up;
	Math Debrief/Critique			ELA Debrief/Critique				

Day/Session	Panel 9: Math Grades 3, 4, 5 **Craig Deville**	Panel 10: Math Grades 6, 7 **Jennifer Bowen**	Panel 11: Math Grades 8, High School **Lidia Martinez**	Panel 12: ELA Grades 3-4 **Dan Bowen**	Panel 13: ELA Grades 5, 6, 7 **Jeff Barker**	Panel 14: ELA Grades 8, High School **Job Thomas**	Panel 15: Science Grades 4, 7, 11 **Steve Cramer**	Panel 16: Social Studies Grades 5, 8, 11 **Stephanie Lai**
• 4:45	Wrap-Up; Evaluate			Wrap-Up; Evaluate			Up; Eval	Eval
• 5:00	Dismiss			Dismiss			Dismiss	Dismiss

[1] During Wrap-Up on July 1, panelists received information regarding their July 2 room assignments. The Science and Social Studies panels returned to their same breakout rooms. For Math and ELA panels, half the panelists were selected to participate in vertical articulation, and the other half were selected to participate in a final critique of the process. Within each panel, the facilitator selected half the panelists at each table to go to vertical articulation, and the other half to go to the critique.

Conduct of the meeting. Dr. Bunch provided an overview of the four days and gave the panels their charge (see Appendix B). Afterwards, panelists dispersed to their breakout rooms to review the FI tests under the direction of the facilitators listed in Table 3. These same facilitators also led the panelists in a review of the PLDs. After lunch on June 29, Dr. Bunch provided an overview of the Bookmark procedure (see Appendix B). Panelists then dispersed to their breakout rooms for Bookmark practice with a small set of items. Following this practice round, facilitators answered questions, and determined readiness to begin Round 1 by administering and reviewing the Round 1 Readiness Form (see Appendix B). Dr. Bunch and MDE staff circulated among the eight panel rooms throughout each day to observe and answer questions. At the end of each day, MI and MDE staff met for approximately one hour to debrief the day's activities and outline the next day's activities.

Panelists worked in small groups of 3-5 within a room of 9-10. They consulted with others at their table during each round. The Science and Social Studies panels followed a slightly different schedule than the other panels, partly due to the noncontiguous nature of their tests and partly due to the number of tests they had to review.

Panelists proceeded through three rounds of Bookmark item rating with feedback and discussion between rounds. As they worked their way through their ordered item booklets, they entered their bookmarks on scannable documents like the one shown in Figure 5.

As panelists completed their Bookmark item ratings, MI staff gathered them and processed the results. Although tests were administered in grade bands, panels had a unique form for each grade. Thus, for example, the panel recommending cut scores for Mathematics grades 3-5 evaluated items in three separate ordered item booklets, one each for grade 3, 4, and 5.

As panelists completed a round of Bookmark rating, MI staff collected the forms and processed them as described in Cizek & Bunch (2007, Ch. 10). Prior to Round 2, facilitators shared results

of Round 1, facilitated a discussion of results in terms of dispersion of bookmarks as well as median cut score, and shared impact data.

After discussion of Round 1 results, panelists completed the Readiness Form, indicating readiness to begin Round 2. They completed Round 2 as they had completed Round 1, working in small groups and entering two bookmarks. At the end of the round, MI staff collected the scannable forms, processed them as in Round 1, and prepared results to present to panelists. During the discussion of Round 2 results, facilitators shared the same types of information they had shared after Round 1 but also revealed impact data from previous years as additional context.

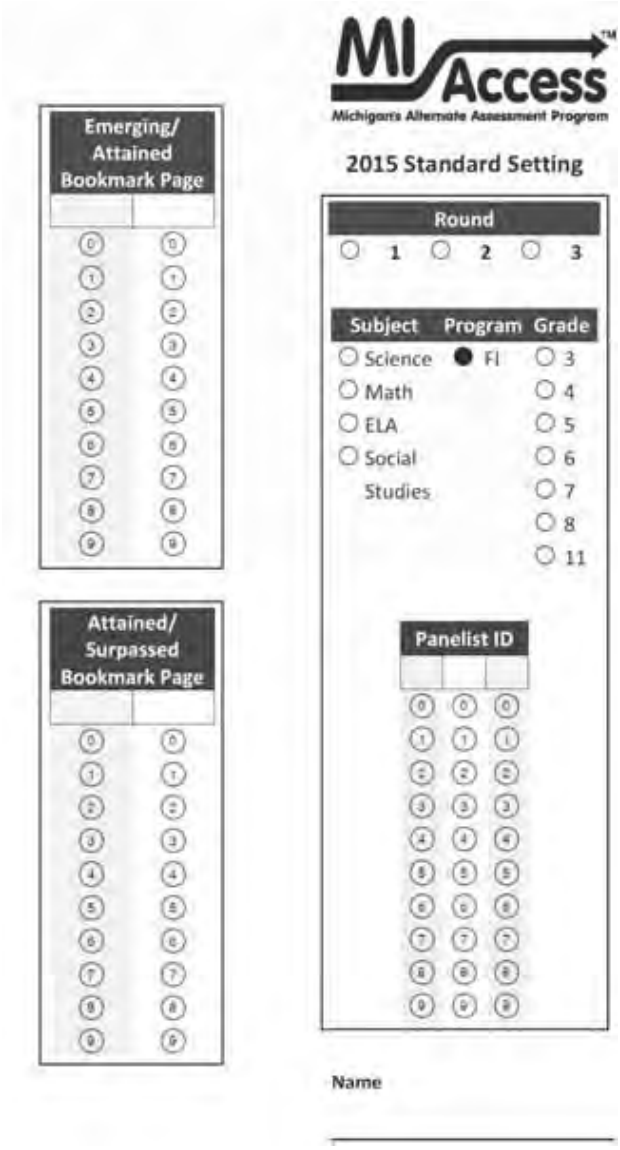


Figure 5. Bookmark Entry Form

At the close of the Round 2 discussion, panelists indicated their readiness to begin Round 3. They proceeded through Round 3 as they had in Rounds 1 and 2, entering two bookmarks on their scannable forms. MI facilitators collected the forms and processed them as in Rounds 1 and 2.

After Round 3, ELA and Math panels divided into four groups: ELA vertical articulation, ELA critique, Math vertical articulation, and Math critique. Procedures for forming and leading the groups were the same as that described above for the P/SI panels.

Results

Round-by-Round Results

Tables 6-14 show the round-by-round results of the Body of Work and Bookmark activities. Figures 6-8 show the impacts of the Round 3 cut scores.

Table 6
Round 1 Results for Participation

	Cut Scores		% in Level		
Test	**Attained**	**Surpassed**	**Emerging**	**Attained**	**Surpassed**
ELA 3-5	27	47	36.6	40.5	23.0
ELA 6-8	32	47	47.6	31.6	20.9
ELA 11	30	44	34.9	29.9	35.2
Math 3-5	26	45	31.6	36.6	31.8
Math 6-8	20	38	26.1	34.0	39.9
Math 11	33	48	36.9	21.8	41.3
Science 4	36	68	28.2	45.0	26.8
Science 7	40	70	38.4	38.1	23.5
Science 11	48	77	39.5	39.5	20.9

Table 7
Round 1 Results for Supported Independence

	Cut Scores		% in Level		
Test	**Attained**	**Surpassed**	**Emerging**	**Attained**	**Surpassed**
ELA 3-5	22	49	8.2	58.8	33.0
ELA 6-8	30	44	14.1	27.1	58.9
ELA 11	27	47	13.1	37.6	49.3
Math 3-5	33	48	25.3	38.8	35.9
Math 6-8	25	46	14.1	41.2	44.7
Math 11	22	43	10.5	51.4	38.1
Science 4	30	52	10.3	32.5	57.2
Science 7	27	52	7.8	42.0	50.2
Science 11	39	58	20.8	37.3	41.9

Table 8
Round 1 Results for Functional Independence

	Cut Scores		% in Level		
Test	**Attained**	**Surpassed**	**Emerging**	**Attained**	**Surpassed**
ELA 3	2308	2316	57.6	14.2	28.2
ELA 4	2409	2421	35.6	23.9	40.5
ELA 5	2506	2529	30.2	40.9	28.9
ELA 6	2611	2628	29.0	31.9	39.1
ELA 7	2714	2736	26.0	42.8	31.2
ELA 8	2799	2815	11.2	18.0	70.8
ELA 11	3155	3167	21.3	10.9	67.8
Math 3	2307	2323	28.7	21.6	49.7
Math 4	2411	2423	26.6	26.3	47.1
Math 5	2505	2521	19.1	27.2	53.7
Math 6	2600	2616	18.9	35.4	45.7
Math 7	2709	2722	42.1	30.1	27.8
Math 8	2803	2815	40.0	24.4	35.6
Math 11	3093	3105	14.7	20.5	64.8
Science 4	2383	2394	40.4	28.6	31.0
Science 7	2708	2728	40.3	39.9	19.8
Science 11	3160	3189	42.5	36.3	21.2
Social Studies 5	2515	2529	63.3	22.7	14.1
Social Studies 8	2805	2819	56.8	26.1	17.1
Social Studies 11	3162	3183	51.1	24.9	24.0

Table 9
Round 2 Results for Participation

	Cut Scores		% in Level		
Test	**Attained**	**Surpassed**	**Emerging**	**Attained**	**Surpassed**
ELA 3-5	27	44	36.6	33.5	29.9
ELA 6-8	29	45	43.0	33.8	23.2
ELA 11	30	46	34.9	32.6	32.6
Math 3-5	34	47	45.3	27.4	27.3
Math 6-8	25	46	35.5	39.7	24.8
Math 11	29	48	32.9	25.9	41.3
Science 4	46	70	38.8	36.9	24.4
Science 7	46	72	46.0	31.9	22.2
Science 11	47	76	39.0	36.6	24.4

Table 10
Round 2 Results for Supported Independence

	Cut Scores		% in Level		
Test	**Attained**	**Surpassed**	**Emerging**	**Attained**	**Surpassed**
ELA 3-5	28	47	14.1	45.8	40.0
ELA 6-8	35	47	23.2	29.3	47.6
ELA 11	30	41	15.9	20.1	64.0
Math 3-5	36	50	29.5	41.2	29.3
Math 6-8	29	50	20.2	49.1	30.7
Math 11	23	44	14.9	47.0	38.1
Science 4	31	54	11.0	37.0	52.0
Science 7	31	55	11.7	47.0	41.3
Science 11	44	58	26.8	31.4	41.9

Table 11
Round 2 Results for Functional Independence

	Cut Scores		% in Level		
Test	**Attained**	**Surpassed**	**Emerging**	**Attained**	**Surpassed**
ELA 3	2297	2313	30.4	38.1	31.5
ELA 4	2406	2420	29.7	29.8	40.5
ELA 5	2506	2528	30.2	40.9	28.9
ELA 6	2611	2628	29.0	31.9	39.1
ELA 7	2716	2732	29.1	29.5	41.4
ELA 8	2804	2821	13.7	23.3	63.0
ELA 11	3153	3174	19.6	20.6	59.8
Math 3	2308	2323	28.7	21.6	49.7
Math 4	2411	2423	26.6	26.3	47.1
Math 5	2511	2523	32.2	20.7	47.1
Math 6	2607	2616	29.9	24.4	45.7
Math 7	2710	2723	48.7	23.5	27.8
Math 8	2803	2815	40.0	24.4	35.6
Math 11	3095	3117	18.3	40.9	40.8
Science 4	2383	2394	40.4	28.6	31.0
Science 7	2709	2731	44.9	39.1	16.0
Science 11	3160	3189	42.5	36.3	21.2
Social Studies 5	2515	2526	63.3	22.6	14.1
Social Studies 8	2805	2819	56.8	26.1	17.1
Social Studies 11	3163	3182	51.1	24.9	24.0

Table 12
Round 3 Results for Participation

	Cut Scores		% in Level		
Test	**Attained**	**Surpassed**	**Emerging**	**Attained**	**Surpassed**
ELA 3-5	27	44	36.6	33.5	29.9
ELA 6-8	29	45	43.0	33.8	23.2
ELA 11	26	43	27.9	36.6	35.5
Math 3-5	33	46	44.8	24.0	31.3
Math 6-8	24	47	31.3	47.7	21.0
Math 11	27	47	29.9	28.2	41.9
Science 4	46	72	38.8	39.6	21.7
Science 7	44	72	43.2	34.6	22.2
Science 11	48	75	39.5	35.8	24.7

Table 13
Round 3 Results for Supported Independence

	Cut Scores		% in Level		
Test	**Attained**	**Surpassed**	**Emerging**	**Attained**	**Surpassed**
ELA 3-5	29	47	16.8	43.2	40.0
ELA 6-8	37	47	26.4	26.0	47.6
ELA 11	28	43	13.1	27.4	59.4
Math 3-5	32	48	21.7	42.4	35.9
Math 6-8	29	50	20.2	49.1	30.7
Math 11	24	43	15.1	46.8	38.1
Science 4	32	55	11.0	41.1	48.0
Science 7	33	55	16.0	42.6	41.3
Science 11	45	57	31.0	26.4	42.7

Table 14
Round 3 Results for Functional Independence

	Cut Scores		% in Level		
Test	**Attained**	**Surpassed**	**Emerging**	**Attained**	**Surpassed**
ELA 3	2297	2313	30.4	38.1	31.5
ELA 4	2406	2420	29.7	29.8	40.5
ELA 5	2506	2528	30.2	40.9	28.9
ELA 6	2611	2628	29.0	31.9	39.1
ELA 7	2716	2732	29.1	29.5	41.4
ELA 8	2803	2821	13.7	23.3	63.0
ELA 11	3152	3172	19.6	20.6	59.8
Math 3	2308	2323	28.7	21.6	49.7
Math 4	2411	2426	26.6	32.5	40.9
Math 5	2511	2523	32.2	20.7	47.1
Math 6	2607	2626	29.9	39.6	30.5
Math 7	2708	2723	42.1	30.1	27.8
Math 8	2799	2815	34.8	29.6	35.6
Math 11	3094	3116	18.3	36.5	45.2
Science 4	2383	2394	40.4	28.6	31.0
Science 7	2709	2728	44.9	35.3	19.8
Science 11	3160	3188	42.5	36.3	21.2
Social Studies 5	2515	2526	63.3	22.6	14.1
Social Studies 8	2805	2818	56.8	22.4	20.8
Social Studies 11	3164	3182	51.1	24.9	24.0

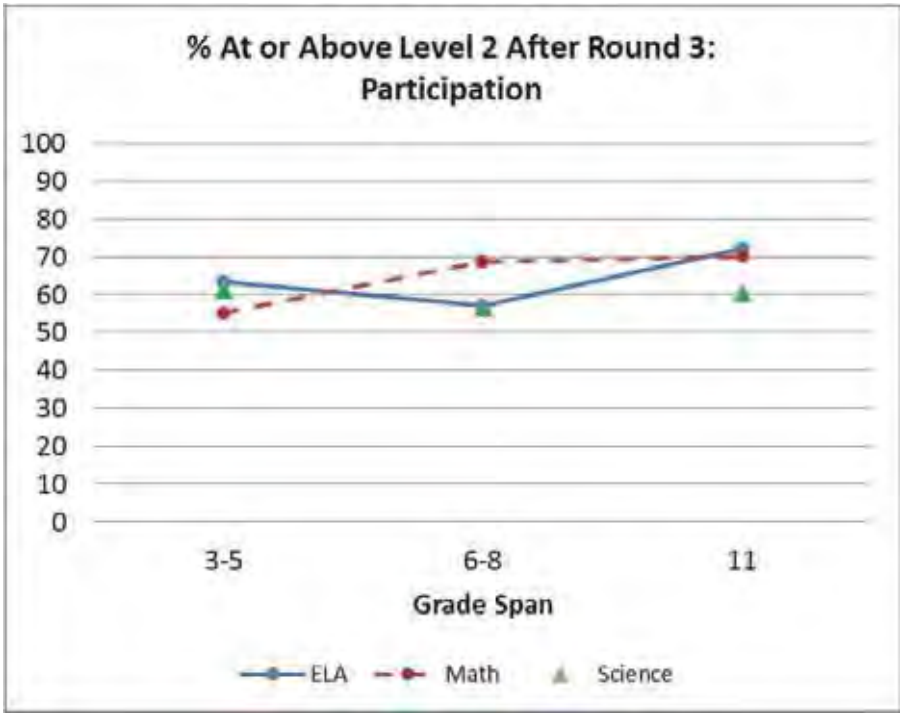


Figure 6. Round 3 Impact: Participation

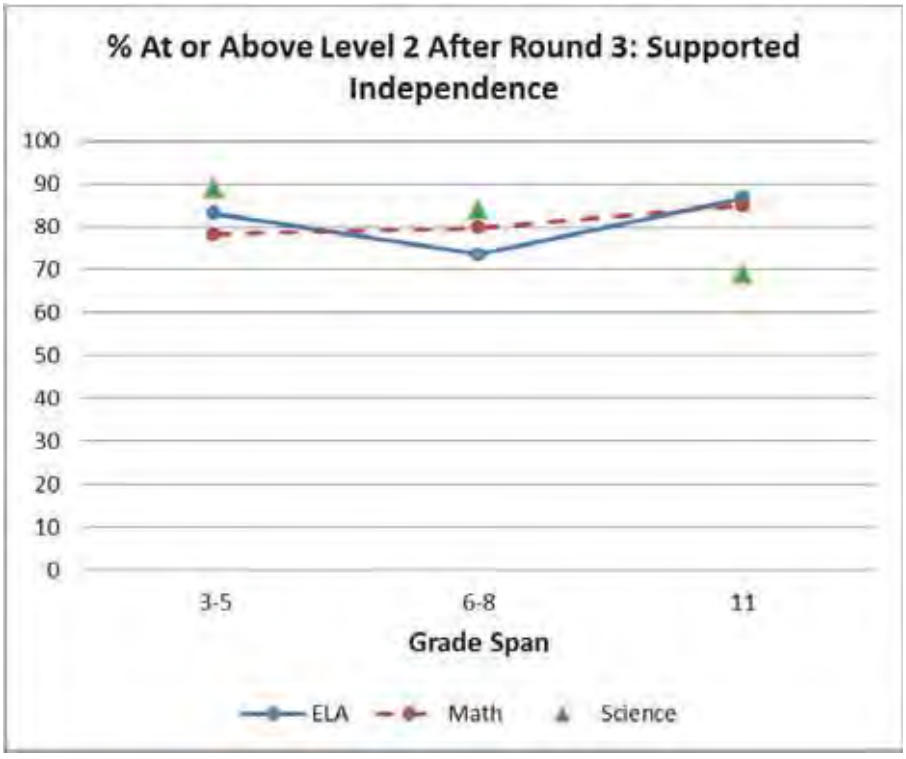


Figure 7. Round 3 Impact: Supported Independence

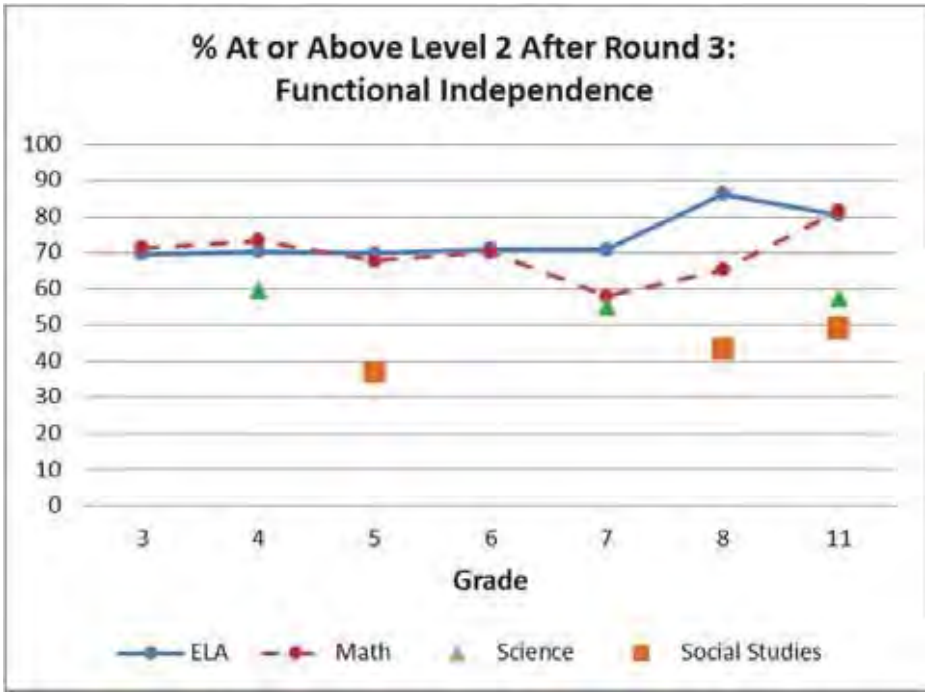


Figure 8. Round 3 Impact: Functional Independence

Vertical Articulation

On the final day of both weeks of standard setting, English language arts (ELA) and Math panelists were regrouped as follows:

- ELA Vertical Articulation Committee (VAC)
- Math Vertical Articulation Committee
- ELA Review and Critique Committee
- Math Review and Critique Committee

The Science and Social Studies panels, given that their grades are noncontiguous, continued through Round 3 and final review on the final day and did not participate in vertical articulation. The P/SI ELA and Math panels had been given the opportunity to advise MDE as to whether to articulate by grade span or by individual grade. For Participation, the panels unanimously endorsed articulation by grade span. For Supported Independence, there was some support for articulation by individual grade, but those in favor of articulation by grade span outnumbered those in support of individual-grade articulation by about three to one.

Vertical articulation began with an overview of the process, followed by a question-and-answer period. During this phase of the process, both ELA and Math VACs met together. After the question-and-answer session, ELA and Math VACs separated into different rooms. During the remainder of the day, each VAC reviewed results (cut scores and impact) across grade spans

and recommended changes. Changes were effected by a motion, second, discussion and vote. Given that the changes were to override cut scores set over three rounds of deliberation, a 2/3 majority was required to pass any motion. The two P/SI committees made a total of three changes (one for ELA and two for Math). The two FI committees made a total of five changes three for ELA and two for Math). Results are depicted in Tables 15-17. Highlighted entries in these tables indicate changes, relative to Round 3 (Tables 12-14). Science and Social Studies cut scores and impact are included even though they were not subject to vertical articulation. Figures 9-11 show the impacts after vertical articulation.

Table 15
Cut Scores and Impact for Participation – After Vertical Articulation

	Cut Scores		% in Level		
Test	**Attained**	**Surpassed**	**Emerging**	**Attained**	**Surpassed**
ELA 3-5	27	44	36.6	33.5	29.9
ELA 6-8	29	45	43.0	33.8	23.2
ELA 11	26	43	27.9	36.6	35.5
Math 3-5	33	46	44.8	23.9	31.3
Math 6-8	26	47	35.9	43.1	21.0
Math 11	27	47	29.9	28.2	41.9
Science 4	46	72	38.8	39.6	21.7
Science 7	44	72	43.2	34.6	22.2
Science 11	48	75	39.5	35.8	24.7

Table 16
Cut Scores and Impact for Supported Independence – After Vertical Articulation

	Cut Scores		% in Level		
Test	**Attained**	**Surpassed**	**Emerging**	**Attained**	**Surpassed**
ELA 3-5	29	47	16.8	43.2	40.0
ELA 6-8	34	47	19.5	32.9	47.6
ELA 11	28	43	13.1	27.5	59.4
Math 3-5	32	49	21.7	48.5	29.8
Math 6-8	29	50	20.2	49.1	30.7
Math 11	24	43	15.1	46.8	38.1
Science 4	32	55	11.0	41.1	48.0
Science 7	33	55	16.0	42.6	41.3
Science 11	45	57	31.0	26.4	42.7

Table 17
Cut Scores and Impact for Functional Independence – After Vertical Articulation

	Cut Scores		% in Level		
Test	**Attained**	**Surpassed**	**Emerging**	**Attained**	**Surpassed**
ELA 3	2297	2312	30.4	35.7	33.9
ELA 4	2406	2420	29.7	29.8	40.5
ELA 5	2506	2523	30.2	32.0	37.8
ELA 6	2611	2628	29.0	31.9	39.1
ELA 7	2713	2732	26.0	32.6	41.4
ELA 8	2803	2821	13.7	23.3	63.0
ELA 11	3152	3172	19.6	20.6	59.8
Math 3	2308	2323	28.7	21.6	49.7
Math 4	2411	2426	26.6	32.5	40.9
Math 5	2511	2529	32.2	32.8	35.0
Math 6	2607	2626	29.9	39.6	30.5
Math 7	2704	2723	30.4	41.8	27.8
Math 8	2799	2815	34.8	29.6	35.6
Math 11	3094	3116	18.3	36.5	45.2
Science 4	2383	2394	40.4	28.6	31.0
Science 7	2709	2728	44.9	35.3	19.8
Science 11	3160	3188	42.5	36.3	21.2
Social Studies 5	2515	2526	63.3	22.6	14.1
Social Studies 8	2805	2818	56.8	22.4	20.8
Social Studies 11	3164	3182	51.1	24.9	24.0

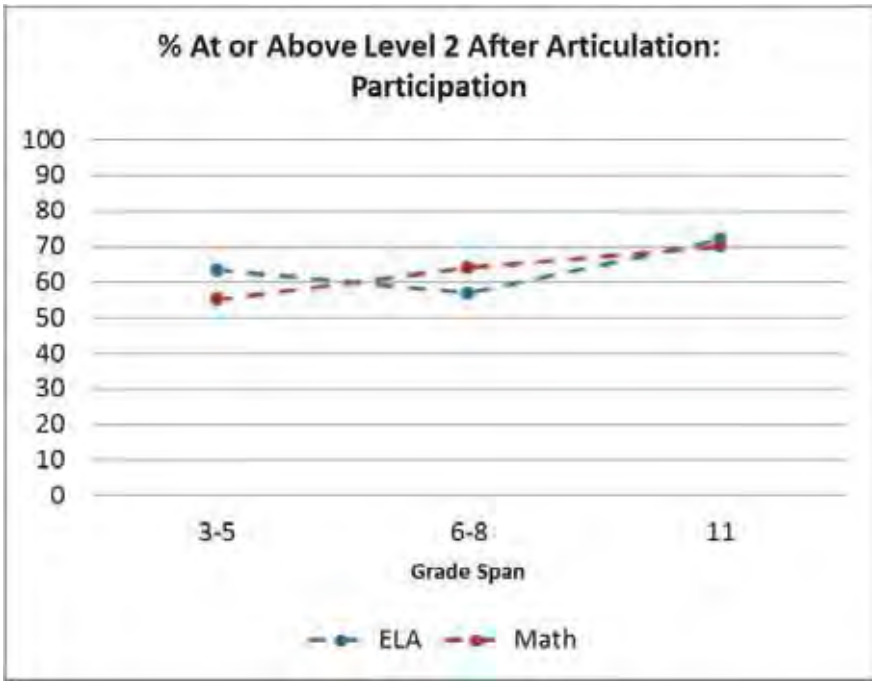


Figure 9. Impact After Vertical Articulation: Participation

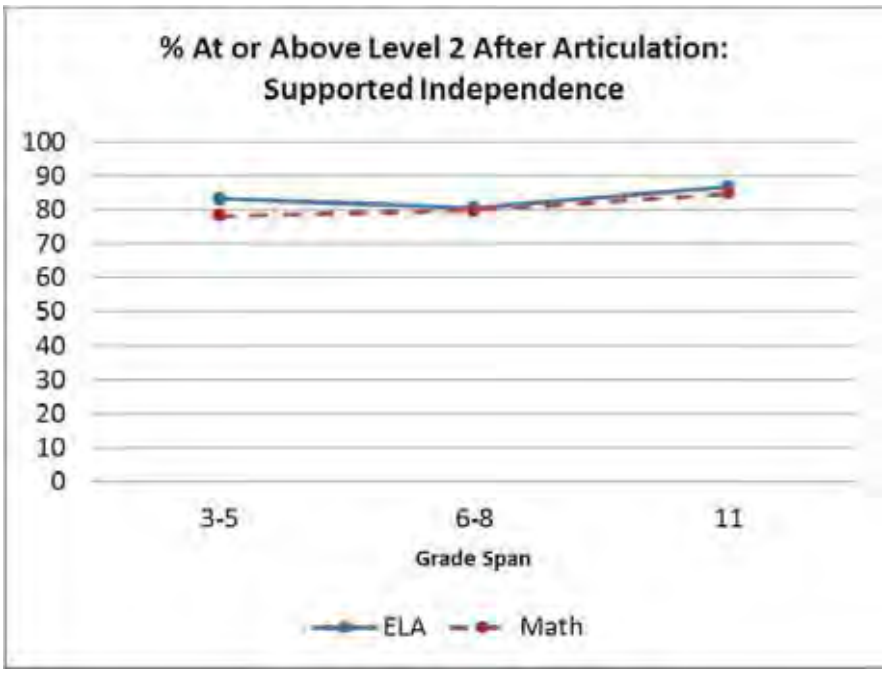


Figure 10. Impact After Vertical Articulation: Supported Independence

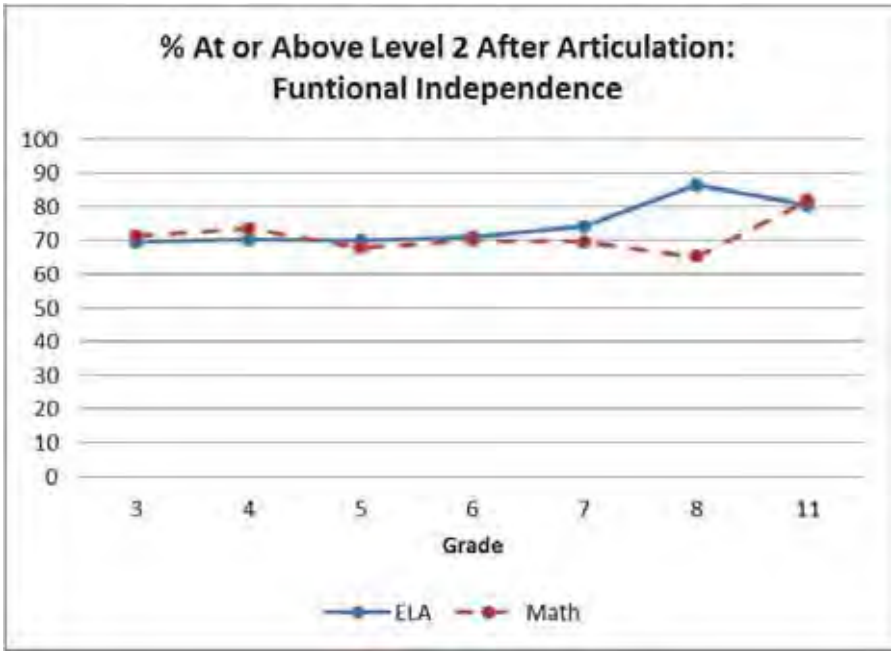


Figure 11. Impact After Vertical Articulation: Functional Independence

Review and Critique

The review and critique sessions occurred at the same time as the vertical articulations. In each panel, facilitators appointed roughly half the panelists to the VAC and the other half to the review/critique committee. The purpose of the review/critique committees was to evaluate the process and provide feedback to the Department regarding improvements in future standard-setting activities. Facilitators employed a common checklist (See Appendix A) to elicit comments from panelists. While the overall tone of the sessions was very positive, several excellent suggestions were received.

Panelists also evaluated the process and their facilitators on eight critical-incident factors, each on a 2-point scale (Agree/Disagree). With regard to facilitators and process, 98-100 percent of panelists agreed with each statement. With regard to facilities and food, reaction was mixed, with 43 percent of P/SI panelists and 53 percent of FI panelists agreeing that the facilities and food service helped to create a good working environment. Results are summarized in Tables 18 and 19. Sample comments follow Table 19. All comments are included in Appendix C.

Table 18
Summary of Evaluations for Participation and Supported Independence (N=62)

	Statement	Agree	Disagree
1	Overall, the facilities and food service helped to create a good working environment.	43%	57%
2	Overall, the training in the standard-setting purpose and methods was clear.	98%	2%
3	Overall, I am confident that I was able to apply the standard setting methods appropriately.	100%	0%
4	Overall, the standard setting procedures allowed me to use my experience and expertise to recommend cut scores for the tests.	100%	0%
5	Overall, the facilitator helped to ensure that everyone was able to contribute to the group discussions and that no one unfairly dominated the discussions.	100%	0%
6	Overall, I was able to understand and use the feedback provided (e.g., other participants' ratings, impact data).	100%	0%
7	I believe that the final group-recommended cut score fairly represents the minimal level of performance for students at the Attained level.	100%	0%
8	I believe that the final group-recommended cut score fairly represents the minimal level of performance for students at the Surpassed level.	100%	0%

Table 19
Summary of Evaluations for Functional Independence (N=76)

	Statement	Agree	Disagree
1	Overall, the facilities and food service helped to create a good working environment.	53%	47%
2	Overall, the training in the standard-setting purpose and methods was clear.	99%	1%
3	Overall, I am confident that I was able to apply the standard setting methods appropriately.	100%	0%
4	Overall, the standard setting procedures allowed me to use my experience and expertise to recommend cut scores for the tests.	100%	0%
5	Overall, the facilitator helped to ensure that everyone was able to contribute to the group discussions and that no one unfairly dominated the discussions.	100%	0%
6	Overall, I was able to understand and use the feedback provided (e.g., other participants' ratings, impact data).	100%	0%
7	I believe that the final group-recommended cut score fairly represents the minimal level of performance for students at the Attained level.	100%	0%
8	I believe that the final group-recommended cut score fairly represents the minimal level of performance for students at the Surpassed level.	100%	0%

Sample comments. In general, panelists were enthusiastic about the facilities and the facilitators but less so about the meals choices. Most comments on the Final Evaluation forms were about food. That issue aside, panelists were very favorably disposed toward all aspects of the experience.

A nearly-universal comment from the P/SI panelists concerned the sequence of tests to be reviewed. Each panel reviewed both the Participation and the Supported Independence tests for a given grade span. In every case, panelists went through Round 1 for Participation and then Round 1 for Supported Independence. Most would have preferred to go through all three rounds of one test and then all three rounds for the other. We will consider those suggestions in preparation for the next P/SI standard setting or any activity that requires review of both sets of exams.

There were many expressions of thanks to MDE for listening and for paying attention to the needs of this population of students. Panelists found the experience very rewarding and expressed their gratitude for the support they receive from MDE.

The following comments are taken from the debriefings conducted on June 18 and July 2. These comments and suggestions are representative of all panels both weeks.

What did you think about the process you went through?

- I loved it. It wasn't what I was expecting and I just loved it. I learned a lot.
- I was glad we did three rounds because I felt it helped me make better decisions.

What helped?

- The facilitator reminded everyone about the PLD. We joked about not answering a question with a question.
- Feeling that we could agree to disagree. Everyone was very professional. We were not ostracized about feeling differently.

What wasn't helpful?

- Flopping back and forth from P to SI. If we could have gone through participation and continued those discussions about participation before shifting our mindset. We did it okay, but it was harder.
- Color code the forms so we can separate out the different rounds of work samples.

When did things click?

- Round 2 because we got round 1 impact data and we realized we kind of got there but we didn't really get it.
- Half way through round 1 it clicked better. Then we all got there. For round 1 for supported independence it was better. We left and we got it and then we talked to him the next day and we didn't get it.

Was there anything we might have done to make it click sooner?

- I think you have to go through it. I think you have to do it. On the job training. You learn by doing.

What was the most difficult part for you?

- Switching back and forth between participation and supported independence.
- I think making decisions about the work samples right on the edge. Having to look at those and re-look at those.
- Putting myself in the place of kids who I don't teach.

When you needed help, did you get it?

- The facilitator and anyone doing this process was available to answer questions. Other people at our table, our peers, were available.
- We definitely had sufficient materials.

What advice would you like us to take back to MDE?

- Thank you for listening to our opinions.
- Thank you for paying attention to this population.
- Thank you for inviting our input.

Conclusions and Recommendations

The two sessions went remarkably well. Panels were large enough to yield reliable results and diverse enough to yield generalizable results. Panelists expressed satisfaction with their training and confidence in their final recommendations. Observations during each day of standard setting revealed that facilitators were following their scripts and keeping panelists focused on the test contents and performance level descriptors at all times.

The process for arriving at cut scores for both standard-setting activities was rigorous and consistent with best practices and overseen by highly competent practitioners. The resulting cut scores and corresponding impacts were reasonably consistent across grade spans or individual grades as well as with historical trends in Michigan for these populations. It is our recommendation that the cut scores recommended by these panels be adopted without modification or adjustment.

References

Cizek, G. J. & Bunch, M. B. (2007). *Standard Setting, Establishing and Evaluating Performance Standards on Tests*. Thousand Oaks, CA: Sage.

Kingston, N. M. & 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, NY: Routledge.

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.

Appendix A
Participation and Supported Independence Training Materials

- Body of Work Facilitator Script
- Body of Work Refresher Notes
- Body of Work Practice Round Form
- Readiness Form
- Process Evaluation Form
- Debriefing Script

Body of Work Facilitator Script

Day 1 A.M. (9:15am – 11:30am) Post-Overview Presentation

Goals: Introduce panelists to the Participation and Supported Independence (P/SI) tests and to the PLDs.

Materials of Importance: P/SI Tests; P/SI PLDs; non-disclosure agreements.

Facilitator Outline:

1. Conduct group introduction (15-30 seconds per person).
2. Have panelists sign non-disclosure agreements.
3. Allow panelists to become familiar with the P/SI tests.
4. Dismiss panelists for 15-minute break at 10:30.
5. Lead panelists in a discussion of the PLDs.
6. Dismiss panelists for 11:30 presentation of the Body of Work procedure.

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.
- Remind panelists that they are reviewing the tests so that they can have first-hand experience of the types of items that students will be charged with completing, not so that they can critique the test development process.
- Ask panelists to discuss their impressions of the test content. What did they think would have been easy or difficult for MI-Access students? What types of skills did they notice would be needed to successfully answer the items on the test?
- Briefly remind panelists that PLDs are simply descriptions of students' abilities at a particular performance level; also point out that all their decisions concerning setting cut scores must be firmly grounded in the PLDs.
- Ask panelists to read the PLDs carefully and to contemplate what it means to be Emerging, Attained, or Surpassed.
- Encourage panelists to imagine students they have known who might have fit the Emerging, Attained, or Surpassed 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.

Day 1 P.M. (1:00pm – 5:00pm) Post-Body of Work Orientation Presentation

Goals: Introduce panelists to student work samples by leading them through the practice

Body of Work (BoW) samples. Ensure all panelists are prepared to begin the Standard Setting process. Complete BoW Round 1 for the Participation test.

Materials of Importance: BoW practice samples; Readiness Form; Round 1 BoW samples for the Participation test.

Logistics

- Table Assignments
- Distribution of Materials
- Groundrules – Discussion by Table

Facilitator Outline:

1. Assist panelists through BoW Practice Round
2. Have panelists complete Round 1 Readiness Form
3. Begin Round 1 – Participation
4. Monitor Round 1 - Participation
5. End Round 1 – Participation (key points)
6. Dismiss panelists for the evening (Collect secure materials)

Facilitator Talking Points:

- Conduct a room-wide discussion concerning the first two work samples in the BoW Practice samples.
- Allow panelists to complete the remaining four practice work samples with their tables.
- Encourage panelists to consult with the other people sitting at their tables during each round.
- Remind panelists that all their decisions concerning their placement of work samples into performance categories must be firmly grounded in the PLDs.
- Remind panelists to consider the following questions for each work sample:
    - What types of skills and abilities must a student possess to be capable of each work sample?
    - What skills and abilities make work samples progressively more challenging?
    - What performance level does each work sample best represent?
- Explain to panelists that it is ok for them to have reversals (e.g. work sample #4 is place in Level 2 and work sample #5 is placed in Level 3) as they are sorting the work samples into categories. However, if they are consistently having an inordinate number of reversals encourage them to talk to you or to revisit their PLDs.

- Remind panelists to pace themselves. They have three hours to sort all of their work samples into one of three performance categories.

Day 2 A.M. (8:30am – 12:00pm) Review of Day 1; Finalize Round 1

Goals: Complete BoW Round 1 for the Supported Independence test.

Materials of Importance: Round 1 BoW samples for the Supported Independence test.

Facilitator Outline:

1. Briefly review essential topics covered in Day 1.
2. Begin Round 1 – Supported Independence
3. End Round 1 – Supported Independence
4. Dismiss panelists for lunch

Facilitator Talking Points:

- Briefly review the following topics with the panelists:
    - PLDs:
        - PLDs are simply descriptions of students' abilities at a particular performance level.
        - All decisions in sorting student work samples must be firmly grounded in the PLDs.
        - The differences among Emerging, Attained, and Surpassed that the panel noted in the PLDs.
    - Body of Work procedure:
        - Each work sample represents the complete body of work for one student on either the Participation or Supported independence test.
        - Work samples are ordered by total score with students receiving the highest score appearing at the beginning of the set and students receiving the lowest score appearing at the end of the set.
        - BoW procedure is characterized by utilizing the PLDs to place each student work sample into a performance category.
- Before beginning Round 1 – Supported Independence, reiterate to panelists that they should consider the following questions for each work sample:
    - What types of skills and abilities must a student possess to be capable of each work sample?
    - What skills and abilities make work samples progressively more challenging?
    - What performance level does each work sample best represent?

Day 2 P.M. (12:45pm – 5:00pm) Round 1 Discussion; Begin Round 2

Goals: Review Round 1 – Participation results. Ensure all panelists are prepared to begin Round 2. Complete BoW Round 2 for the Participation test.

Materials of Importance: Round 1 – Participation results (tables, graphs, and impact data); Round 2 Readiness form; Round 2 – Participation work samples

Facilitator Outline:

1. Discuss Round 1 results for the Participation test
2. Have panelists complete Round 2 Readiness Form
3. Begin Round 2 – Participation
4. End Round 2 – Participation
5. Dismiss panelists for the evening

Facilitator Talking Points:

- Conduct a room-wide discussion concerning the Round 1 results for the Participation test. Highlight the following topics:
    - What were the challenges panelists faced in Round 1?
    - What factors influenced panelists' decisions to rate certain work samples?
    - 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 work samples where the room was evenly divided in opinion (i.e. a work sample that half the room rated as Attained and the other half rated as Emerging).
- Review the Round 1 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 three performance categories seem realistic?
- Explain to panelists the Round 2 process:
    - Work samples will not be exactly the same as they were for Round 1. A targeted sample will be employed that focuses on work samples in the relative vicinity of the Round 1 cut score.
    - Panelists should rate each work sample using the same process used in Round 1.
    - Panelists are free to discuss the work samples and PLDs with their tablemates, but not across tables.
- Ask if the panelists have any questions and ensure they are ready to begin Round 2.

Day 3 A.M. (8:30am – 12:00pm) Round 1 Discussion Part II; Finalize Round 2

Goals: Review Round 1 – Supported Independence results. Complete BoW Round 2 for the Supported Independence test.

Materials of Importance: Round 1 – Supported Independence results (tables, graphs, and impact data); Round 2 Readiness form; Round 2 – Supported Independence work samples

Facilitator Outline:

1. Discuss Round 1 results for the Supported Independence test
2. Begin Round 2 – Supported Independence
3. End Round 2 – Supported Independence
4. Dismiss panelists for lunch

Facilitator Talking Points:

- Conduct a room-wide discussion concerning the Round 1 results for the Supported Independence test. Highlight the following topics:
    - What were the challenges panelists faced in Round 1?
    - What factors influenced panelists' decisions to rate certain work samples?
    - 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 work samples where the room was evenly divided in opinion (i.e. a work sample that half the room rated as Attained and the other half rated as Emerging).
- Review the Round 1 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 three performance categories seem realistic?
- Reiterate to panelists the Round 2 process:
    - Work samples will not be exactly the same as they were for Round 1. A targeted sample will be employed that focuses on work samples in the relative vicinity of the Round 1 cut score.
    - Panelists should rate each work sample using the same process used in Round 1.
    - Panelists are free to discuss the work samples and PLDs with their tablemates, but not across tables.
- Ask if the panelists have any questions and ensure they are ready to begin Round 2.

Day 3 P.M. (12:45pm – 4:45 pm) Review; Wrap-up

Goals: Review and Revise Round 2 results for both the Participation and Supported Independence tests. Ensure panelists complete the Final Evaluation Form. Inform panelists of Day 4 assignments.

Materials of Importance: Round 2 results (tables, graphs, and impact data) for both the P/SI tests; Final Evaluation Form.

Facilitator Outline:

1. Review Round 2 results for P/SI tests.
2. Revise Round 2 results for P/SI tests.
3. Have panelists complete Final Evaluation Form.
4. Inform panelists of Day 4 room assignments.
5. Dismiss panelists for evening.

Facilitator Talking Points:

- Conduct a room-wide discussion concerning the Round 2 results for both the P/SI tests. Highlight the following topics:
    - What were the challenges panelists faced in Round 2?
    - What factors influenced panelists' decisions to rate certain work samples?
    - 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 work samples where the room was evenly divided in opinion (i.e. a work sample that half the room rated as Attained and the other half rated as Emerging).
- Review the Round 2 impact data. Focus their attention on whether the percentages of students in the three performance categories seem realistic.
- The revision process will entail giving the panelists the opportunity to change the impact of the cut scores for all combinations of grade and test (P/SI) within their own respective grade band. Highlight the following topics:
    - Are the cut scores reasonably intuitive or are there major dips or peaks in the percentage of students within performance levels at certain grade levels? For example, if 50% of G3 students and 55% of G5 students are Attained or above would it make sense for only 45% of G4 students to be Attained or above?
    - Ask panelists to discuss possible work samples that it would be justified to move from one performance level to another.
    - Encourage panelists to make small changes across all grade levels as opposed to one large change at one grade level.

- If the reclassification of a work sample has been justified using the PLDs call for a motion and a second. Then allow the room to vote democratically on whether to change the cut score. A 2/3 majority is needed for the motion to pass.
- Remind panelists that recommendations must be firmly grounded in the PLDs.

Inform panelists of their Day 4 room assignments. Half of the Math and ELA panelists will participate in cross-grade review; the other half will participate in a final critique of the standard setting activity

Body of Work Refresher Notes for Facilitators

Facilitators should keep the following talking points in mind:

1. Remind panelists that work samples are ordered from lowest to highest scoring students.
2. Panelists will sort the work samples into one of three groups: "Emerging toward the standard," "Attaining the standard," or "Surpassed the standard."
3. It is OK—even expected—that panelists do not follow strict ordering, that is, that they place initial work samples only under "Emerging," followed by a group of samples only assigned to "Attained," with the final samples placed only under "Surpassed."
4. Panelists should have their PLDs beside them and refer to them for all decisions.
5. Allow table talk during Round 1. If an issue comes up at one table that should be brought to the attention of all panelists, do so.
6. Present impact data after Round 1. When panelists realize impact, some will want to immediately change their ratings. Remind them that any changes should be grounded in the panelists' consideration of the PLDs.
7. Remember to have panelists complete Readiness and Evaluation forms.

MI-Access Participation and Supported Independence
Standard Setting
June 15-18, 2015

Body of Work Rating Form: Practice Round

Content Area	Grade(s)	
English Language Arts	3-5	4, 7
Math	6-8	11
Science		

Panelist _____

For each Sample enter
3 for Emerging, 2 for Attained, or 1 for Surpassed

Sample	Level	Comment
1		
2		
3		
4		
5		
6		

Standard Setting
Readiness Form

Panelist Number_____

Ready for Round 1: I have completed the training, and I understand what I need to do to complete Round 1.

(Circle one): **Yes** **No**

Ready for Round 2: I have completed the discussion of Round 1, and I understand what I need to do to complete Round 2.

(Circle one): **Yes** **No**

Ready for Round 3: I have completed the discussion of Round 1, and I understand what I need to do to complete Round 3.

(Circle one): **Yes** **No** **N/A**

Ready for Articulation: I have completed the discussion of Round 2 and the articulation training, and I understand what I need to do to complete vertical articulation.

(Circle one): **Yes** **No**

Final: I have completed vertical articulation and discussed results, and I believe that the cut scores recommended by this panel are reasonable and fair.

(Circle one): **Yes** **No**

Comments (on back)

Final Evaluation

Facilitator_____

Directions: Check one box for each of the following statements by placing an "X" in the box corresponding to your opinion. If you have any additional comments, please write them in the space provided at the end of this form.

	Statement	Agree	Disagree
1	Overall, the facilities and food service helped to create a good working environment.		
2	Overall, the training in the standard-setting purpose and methods was clear.		
3	Overall, I am confident that I was able to apply the standard setting methods appropriately.		
4	Overall, the standard setting procedures allowed me to use my experience and expertise to recommend cut scores for the tests.		
5	Overall, the facilitator helped to ensure that everyone was able to contribute to the group discussions and that no one unfairly dominated the discussions.		
6	Overall, I was able to understand and use the feedback provided (e.g., other participants' ratings, impact data).		
7	I believe that the final group-recommended cut score fairly represents the minimal level of performance for students at the Attained level.		
(If you answered Disagree to Question 7, do you believe the final group-recommended cut score for Attained is:____too high or____too low (check one).			
8	I believe that the final group-recommended cut score fairly represents the minimal level of performance for students at the Surpassed level.		
(If you answered Disagree to Question 8, do you believe the final group-recommended cut score for Advanced Surpassed is:____too high or____too low (check one).			

Comments:

_____

_____

_____

_____

_____

_____

Thank you! When you have completed this form, please return it to your facilitator.

Debriefing Script for June 18

Show Round 3 Results/Impact

- How reasonable do these seem?

- Other comments about the results.

Comments on the Process

- In general, what did you think about the process?

- What helped?

- What didn't help?

- When did it "click?"

- How might we have made it click sooner?

- What was the most difficult part?

- Did you get the help you needed? [Explain]

- What advice would you like us to take back to MDE?

Appendix B
Functional Independence Training Materials

- Bookmark Facilitator Script
- Bookmark Refresher Notes
- Bookmark Practice Round Form
- Readiness Form
- Process Evaluation Form
- Debriefing Script

Facilitator Script – Post-Overview Presentation

Goals: Introduce panelists to the Functional Independence (FI) tests and to the PLDs.

Materials of Importance: FI Tests; FI PLDs; non-disclosure agreements; demographics form.

Facilitator Outline:

7. Assist panelists with their table assignment
8. Conduct group introduction (15-30 seconds per person).
9. Have panelists sign non-disclosure agreements and demographics form.
10. Allow panelists to become familiar with the FI tests.
11. Dismiss panelists for 15-minute break at 10:30. [Stagger across the 8 rooms by 2-3 minutes]
12. Lead panelists in a discussion of the PLDs.
13. Dismiss panelists for 11:30 presentation of the Bookmark procedure.

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.
- Remind panelists that they are taking the tests so that they can have first-hand experience of the types of items that students will be charged with completing, not so that they can critique the test development process.
- Ask panelists to discuss their impressions of the test content. What did they think would have been easy or difficult for MI-Access students? What types of skills did they notice would be needed to successfully answer the items on the test?
- Briefly remind panelists that PLDs are simply descriptions of students' abilities at a particular performance level; also point out that all their decisions concerning setting cut scores must be firmly grounded in the PLDs.
- Ask panelists to read the PLDs carefully and to contemplate what it means to be Emerging, Attained, or Surpassed.
- Encourage panelists to imagine students they have known who might have fit the Emerging, Attained, or Surpassed 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*** Attained students and what differentiates them from the Emerging performance level. Then, ask the panelists to focus on the ***Just Barely*** Surpassed students and what differentiates them from the Attained performance level.

Facilitator Script: Post-Bookmark Orientation Presentation

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

Materials of Importance: Bookmark practice OIB; Round 1 Readiness Form; Round 1 OIB for the Functional Independence test; Round 1 Bookmark Rating Forms.

Facilitator Outline:

7. Assist panelists through Bookmark Practice Round.
8. Have panelists complete Round 1 Readiness Form and begin Round 1.
9. Monitor Round 1 and be available to answer panelists' questions.
10. Dismiss panelists for the evening and collect their secure materials.
11. Bring completed Bookmark forms to data analysts.

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 ***Just Barely*** Attained students would have a less than 67% 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 Emerging and Attained performance levels.
- Discuss the results of the Practice Round with the group. Note the range of pages where panelists set their bookmarks.
- 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 ***Just Barely*** Attained students have a 67% chance of answering the item correctly?
    - After placing the Attained bookmark, would ***Just Barely*** Surpassed students have a 67% chance of answering the item correctly?
- Explain to panelists that once they identify an item that they think ***Just Barely*** Attained or Surpassed students have a less than 67% chance of answering correctly that they should take a

look at the next few items in the Order-Item Booklet to confirm that they have reached the best page 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.
- 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. They have three hours to place their bookmarks.

Facilitator Script: Review of Day 1; Finalize Round 1

Goals: Complete Bookmark Round 1.

Materials of Importance: Round 1 Ordered-Item Booklet; Round 1 Bookmark Rating Forms.

Facilitator Outline:

5. Briefly review essential topics covered in Day 1.
    a. Performance Level Descriptors (PLDs).
    b. Bookmark procedure.
6. Monitor Round 1 and be available to answer panelists' questions.
7. End Round 1 and collect panelists' secure materials.
8. Bring completed Bookmark forms to data analysts.

Facilitator Talking Points:

- Briefly review the following topics with the panelists:
    - PLDs:
        - PLDs are simply descriptions of students' abilities at a particular performance level.
        - All decisions in bookmark placement must be firmly grounded in the PLDs.
        - The differences among Emerging, Attained, and Surpassed that the panel noted in the PLDs.
    - Bookmark procedure:
        - 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.
        - The bookmark procedure we will use is characterized by the placement of two bookmarks on the first items in the OIB that the ***Just Barely*** Attained or ***Just Barely*** Surpassed students would have a less than 67% chance of answering correctly.

- Before beginning Round 1 for the next grade level, reiterate to panelists that they should consider the following questions for each item in the OIB:
    - What do you know about students who correctly answer this item?
    - What makes items progressively more challenging?
    - Would ***Just Barely*** Attained students have a 67% chance of answering the item correctly?
    - After placing the Attained bookmark, would ***Just Barely*** Surpassed students have a 67% chance of answering the item correctly?
- **[This bullet only pertains to Math 3-5, Math 6-7, ELA 3-4, and ELA 5-7]** Remind panelists that students in grade clusters take the same test. Panelists will start at the lowest grade in their cluster. After the lowest grade is completed they will consider how students in the next grade up should perform on the same test. When examining the higher grade levels, it may be more efficient to start at or near the bookmarked pages for the previous grade. After all, if a G3 ***Just Barely*** Attained student has a 67% chance to answer an item correctly then it can be surmised that a G4 ***Just Barely*** Attained student would have at least a 67% chance as well. **[NOTE: it will be necessary for the facilitators to return panelists' R1 rating forms so they may know they exact page they placed their bookmark for the previous grade level.]**

Facilitator Script: Round 1 Discussion; Begin Round 2

Goals: Review Round 1 results. Ensure all panelists are prepared to begin Round 2.

Materials of Importance: Round 1 results (tables, graphs, and impact data); Round 2 Readiness Form; Round 2 Ordered-Item Booklet; Round 2 Bookmark Rating Form.

Facilitator Outline:

6. Discuss Round 1 results.
7. Have panelists complete Round 2 Readiness Form and begin Round 2.
8. Monitor Round 2 and be available to answer panelists' questions.
9. End Round 2 and collect panelists' secure materials.
10. Bring completed Bookmark forms to data analysts.

Facilitator Talking Points:

- 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. If there was a wide divergence of opinions 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 Attained bookmark recommended in Round 1. Similarly, the last page in the OIB that they will review for Round 2 will be the highest recommended Surpassed bookmark.
    - Panelists should place their 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*** Attained students have a 67% chance of answering the item correctly?
    - After placing the Attained bookmark, would ***Just Barely*** Surpassed students have a 67% chance of answering the item correctly?
- Ask if the panelists have any questions and ensure they are ready to begin Round 2.

Facilitator Script: Round 2 Discussion; Begin Round 3

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

Materials of Importance: Round 2 results (tables, graphs, and impact data); Round 3 Readiness form; Ordered-Item Booklet.

Facilitator Outline:

1. Distribute then discuss Round 2 results.
2. Have panelists complete Round 3 Readiness Form and begin Round 3.
3. Monitor Round 3 and be available to answer panelists' questions.
4. End Round 3 and collect panelists' secure materials.
5. Bring completed Bookmark forms to data analysts.

Facilitator Talking Points:

- 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. If there was a wide divergence of opinions specifically ask panelists from both ends of the spectrum to explain their reasoning.
- 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 three performance categories seem realistic?
- Explain to panelists the Round 3 process:
    - Round 3 will be more targeted. Panelists will start Round 3 on the lowest recommended Attained bookmark recommended in Round 2. Similarly, the last page in the OIB that they will review for Round 3 will be the highest recommended Surpassed bookmark from Round 2.
    - Panelists should place their bookmarks using the same process employed in Rounds 1 and 2.
    - 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*** Attained students have a 67% chance of answering the item correctly?
    - After placing the Attained bookmark, would ***Just Barely*** Surpassed students have a 67% chance of answering the item correctly?
- Ask if the panelists have any questions and ensure they are ready to begin Round 3.

Facilitator Script: Review; Wrap-up

Goals: Review and Revise Round 3 results for Functional Independence tests. Ensure panelists complete the Final Evaluation Form. Inform panelists of Day 4 assignments.

Materials of Importance: Round 3 results (tables, graphs, and impact data) for the FI test; Final Evaluation Form.

Facilitator Outline:

6. Review Round 3 results for FI tests.
7. Revise Round 3 results for FI tests.
8. Have panelists complete Final Evaluation Form.
9. Inform panelists of Day 4 room assignments.
10. Dismiss panelists for the evening.
11. Bring completed Bookmark forms to data analysts.

Facilitator Talking Points:

- Conduct a room-wide discussion concerning the Round 3 results for both the FI tests. Highlight the following topics:
    - What were the challenges panelists faced in Round 3?
    - 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. If there was a wide divergence of opinions specifically ask panelists from both ends of the spectrum to explain their reasoning.
- Review the Round 3 impact data. Focus panelists' attention on whether the percentages of students in the three performance categories seem realistic.
- The revision process will entail giving the panelists the opportunity to change the impact of the cut scores for all grades within their own respective grade band. Highlight the following topics:
    - Are the cut scores reasonably intuitive or are there major dips or peaks in the percentage of students within performance levels at certain grade levels? For example, if 67% of G3 students and 55% of G5 students are Attained or above would it make sense for only 45% of G4 students to be Attained or above?
    - Ask panelists to discuss possible pages in the Ordered-Item Booklet that it would be justified to move from one performance level to another.
    - Encourage panelists to make small changes across all grade levels as opposed to one large change at one grade level.
    - If the reclassification of an item in the OIB has been justified using the PLDs call for a motion and a second. Then allow the room to vote democratically on whether to change the cut score. A 2/3 majority is needed for the motion to pass.
    - Remind panelists that recommendations must be firmly grounded in the PLDs.
- **[Note: On Day 4 All Science and Social Students panelists will remain in their breakout rooms and continue to place Bookmarks.]**Inform panelists of their Day 4 room assignments. Half of the Math and ELA panelists will participate in cross-grade review; the other half will participate in a final critique of the standard setting process.

Bookmark Refresher Notes
for Facilitators

Facilitators should keep the following talking points in mind:

1. Remind panelists that items are arranged from easiest to hardest based on actual student responses, not on our or somebody's perception of item difficulty.
2. They will proceed through booklet and set two bookmarks, the first separating "Emerging toward the standard" and "Attained the standard," the second separating "Attained the standard" from "Surpassed the standard."
3. Panelists should have their PLDs beside them and refer to them for all decisions.
4. When examining an item, panelists should ask themselves and one another the following two questions: **What skills must a student have in order to know the correct answer?** and **What makes this item more difficult than preceding items?**
5. Next, would a student JUST BARELY entering Level X (e.g., Attaining the standard) have a 2/3 chance or better of answering the question? If yes, move on. If no, set the bookmark on that page.
6. Allow table talk during Round 1. If an issue comes up at one table that should be brought to the attention of all panelists, do so.
7. Present impact data after Round 1. When panelists realize impact, some will want to immediately change their bookmarks. Remind them that any changes should be grounded in the panelists' consideration of the PLDs.
8. Remember to have panelists complete Readiness and Evaluation forms.

MI-Access: Functional Independence Standard Setting
June 29 - July 2, 2015

Bookmark Rating Form: Practice Round

Content Area	Grade(s)	
English Language Arts	3	7
Math	4	8
Science	5	11
Social Studies	6	

Panelist _____

Emerging/Attained Bookmark: ☐

Standard Setting
Readiness Form

Panelist Number_____

Ready for Round 1: I have completed the training, and I understand what I need to do to complete Round 1.

(Circle one): **Yes** **No**

Ready for Round 2: I have completed the discussion of Round 1, and I understand what I need to do to complete Round 2.

(Circle one): **Yes** **No**

Ready for Round 3: I have completed the discussion of Round 1, and I understand what I need to do to complete Round 3.

(Circle one): **Yes** **No** **N/A**

Ready for Articulation: I have completed the discussion of Round 2 and the articulation training, and I understand what I need to do to complete vertical articulation.

(Circle one): **Yes** **No**

Final: I have completed vertical articulation and discussed results, and I believe that the cut scores recommended by this panel are reasonable and fair.

(Circle one): **Yes** **No**

Comments (on back)

Final Evaluation

Facilitator_____

Directions: Check one box for each of the following statements by placing an "X" in the box corresponding to your opinion. If you have any additional comments, please write them in the space provided at the end of this form.

	Statement	Agree	Disagree
1	Overall, the facilities and food service helped to create a good working environment.		
2	Overall, the training in the standard-setting purpose and methods was clear.		
3	Overall, I am confident that I was able to apply the standard setting methods appropriately.		
4	Overall, the standard setting procedures allowed me to use my experience and expertise to recommend cut scores for the tests.		
5	Overall, the facilitator helped to ensure that everyone was able to contribute to the group discussions and that no one unfairly dominated the discussions.		
6	Overall, I was able to understand and use the feedback provided (e.g., other participants' ratings, impact data).		
7	I believe that the final group-recommended cut score fairly represents the minimal level of performance for students at the Attained level.		
(If you answered Disagree to Question 7, do you believe the final group-recommended cut score for Attained is:____too high or____too low (check one).			
8	I believe that the final group-recommended cut score fairly represents the minimal level of performance for students at the Surpassed level.		
(If you answered Disagree to Question 8, do you believe the final group-recommended cut score for Advanced Surpassed is:____too high or____too low (check one).			

Comments:

_____

_____

_____

_____

_____

_____

Thank you! When you have completed this form, please return it to your facilitator.

Debriefing Script for July 2

Show Round 3 Results/Impact

- How reasonable do these seem?

- Other comments about the results.

Comments on the Process

- In general, what did you think about the process?

- What helped?

- What didn't help?

- When did it "click?"

- How might we have made it click sooner?

- What was the most difficult part?

- Did you get the help you needed? [Explain]

- What advice would you like us to take back to MDE?

Appendix C
Evaluation Comments

- Evaluation Comments from Participation and Supported Independence Panelists
- Evaluation Comments from Functional Independence Panelists

Evaluation Comments from Participated and Supported Independence Panelists

food was horrible

Karen was a great facilitator who made the process easy. The lunch left much to be desired (it was mostly yucky) I didn't know Danish was considered breakfast, I had to go in search of protein or fruit. I would have brought my own breakfast, had I known we would not get reimbursed for the daily expense, I am sure the Danish did not cost $8.50 per person.

It would have been nice to have a better breakfast and not such a heavy carb lunch.

Breakfast lacked fruit or yogurt, only high carb choices. I did not stay in hotel or collect stipend. We should be reimbursed for a healthy breakfast. Food was horrible, high grease and no choice. Overall process was good. Not an easy job but worth it. Should have done Participation all the way through before starting SI.

The food options were not good. We should have had 2 options for lunch. Breakfast had no options for those that could not eat sweets. We should have covered P/SI without a switch between days.

Poor food, participants should have been reimbursed for breakfast due to lack of choice. Lunch should have had more options. Corey was a great facilitator! He kept conversation going and asked the right questions to keep us thinking.

Lansing Ctr. Is a nice facility but the meals were too heavy with carbs, no fruit, no water in work rooms, not enough protein, need more veggies. Difficult to work with so much carbs in the system. Need more diet pop.

Smart boards would speed up process or minimize "down time". Corey was excellent! Very neutral and professional. It may be more consistent to stick w/entire grade level, difficult to switch thinking. P & SI scoring is very different.

The food was disgusting but that was not the fault of Corey. :) Did not appreciate staying 2 hours more than other groups because we had Sci 4 & 7, they should have been divided.

Facilities great, unhealthy breakfast and lunch choices. I feel the group made informed decisions based on discussions. It was difficult to go between two types of tests (P/SI) Corey did an excellent job!

Facilities were nice. Food was not good.

more fruit and vegetables

lunch was terrible, facilitators were very competent and knowledgeable

A breakfast of sweets is not what most eat for breakfast. The lunch did not facilitate healthy eating habits.

food was horrible and terrible

Wonderful experience!!

Thank you!

Everyone did a nice job of keeping discussions going and open for growth on the testing critique and evaluation rating.

Less fattening lunches, better breakfasts (fruit, protein). Great facilitator! Great interactions with peers.

The breakfast was very poor.

I enjoyed the discussions and appreciated being able to hash out differences.

Christina did a wonderful job as facilitator, giving us instructions, encouraging other perspectives and keeping it positive.

First time I ever sat on a panel where disagreement was demanded- consensus was accomplished. Agree and disagree is not a good way to set this up.(re: rating form) (Use number scale)

Round one explanations could have been more clear with color coding P/SI. Would have loved fresh fruit for breakfast and a choice at lunch.

This rating sheet should be numbered for more choice not agree/disagree

Great!

Great job posing questions to the group. :)

This was a wonderful experience! :)

Just 1 little suggestion - fruit and yogurt at breakfast. Loved the process and my privilege to participate.

Being part of this process has been very beneficial to me as a teacher. I will have lots of info to take back to my district about reviewing the test taking process. Thanks for allowing me to participate. The facilitator was excellent.

Food service and quality needs to be addressed.

Food service was not great, facility worked well. Tom was wonderful. Very calming presence, very thorough.

I feel all facilitators should know and understand our assessments in order to facilitate the discussions. Tom was one that understood. Tom was great!

Tom did a great job! Food provided was not well-balanced or nutritious. Climate was either very cold or too hot.

facilities were good, food unhealthy. I highly enjoyed my group discussions and the role that I played. I would be interested in getting in touch with the Essential Elements group as we further the process for our students. Tom was a great facilitator.

Facilities fine, food was not good-only sugar and carbs for breakfast. Good process with very skilled facilitator.

Loved the experience-it was beneficial for me as a professional, and we really worked hard on what our students need regarding assessment. A better lunch would be appreciated.

I feel each group should have received the same directions and/or worksheets.

This was a wonderful learning experience. Craig was extremely helpful with explaining and directing our group. Thanks.

Need to provide healthier choices for breakfast and lunch.

Facility was fine, hotel was great, food could use some work! Very disappointed that breakfast didn't include fruit. We also weren't told you weren't going to cover breakfast to buy something healthy. Meal choices were OK but soup and sandwiches would have been good. Facilitator was wonderful! They were all very pleasant. Look forward to other opportunities.

I think this was a great process. I learned a lot.

I really enjoyed the experience. I learned so much about the assessment process and I'll take it back to my district and classroom.

Increased knowledge about assessment. Great conversations, a lot of food for thought

Food service was terrible.

Food choices were minimal. Need to give more variety and healthier choices. Tracey was awesome! She is a great facilitator.

We did not even have water provided during sessions. Tracey did a fantastic job facilitating our group and helping us through this process.

I feel this was a very encompassing project and I learned a lot about the test, administration and what goes into interpretation.

Evaluations from Functional Independence Panelists

Overall, I enjoyed this experience. Next time please provide the team w/healthy meal options. Teachers love fruit & veggies.

Did not like breakfast choices. I thought the process was fair. Although we finished early, I feel 4 days is appropriate so that groups have time to process between rounds.

Loved everything about the process except the food. Need healthy choices, variety or let us buy our own food. The process was very enlightening and beneficial to me and my students. The facilitator was respectful and fair.

Craig was an awesome facilitator! However, I would like to start Round 1 off with individual time to bookmark before discussing as a group. Healthier food options and/or "open" lunch option for reimbursement needed. Need chocolate at tables. :)

Breakfast is all sugar. Serve protein.

He did a great job asking leading questions to direct us in our discussion making and encouraging us to look at different variables. He was consistent in reminding us of what we were looking for in terms of PLD's, questions, and "barely attained" or "surpassed" 2/3 of the time.

Craig kept us productive and on task. His directions and feedback was clear to understand.

The overall experience was rewarding. This was my first opportunity to participate in an event so I feel there is a learning curve, but that being said, the guidance and interaction w/my peers made the decisions I reached mostly comfortable.

Lunch should be on our own w/reimbursement (food was terrible!) Offer way better breakfast. Jennie was fantastic as a facilitator - very personable, helpful & great to work with.

Facilitator was supportive and offered guidance - did not try to sway us. The process was very interesting. Thank you for allowing me to participate.

Facilitator was wonderful, very helpful! Lunch & breakfast needed better/healthier options!

Facilities were great, food left a lot to be desired.

Our facilitator was great- encouraging us to look beyond the first no, discussing each item & letting our voices be heard.

The breakfast and lunch would be better if there was some fruit/yogurt available or a lighter fare. Too carb heavy, which is not conducive to having an alert group.

Food service-Horrible. Bowen, great facilitator! Encourage group to think on a higher level. Great personality.

Jennie did an awesome job facilitating. She encouraged everyone to speak and share. She engaged us in discussions and asked probing questions. I love her excitement and true appreciation for teachers and the education process. Although the food wasn't always great my vegan diet was definitely accommodated and I appreciate that! :)

Lidia did a nice job leading the group through discussions, making sure everyone was heard & finding a place of agreement.

Down time is nice but it could be used more productively to shorten a day. Refocus all the time is difficult.

PLD terms need to be measurable. Limited, basic and consistent are not helpful.

The experience was awesome, I learned so much about testing, learning and process of the different students in different grades.

Food and computer facilities were better when provided by Montan-sp? Catering and working at a facility like DRC where computers are readily available.

Enjoyed being a part of the experience.

Facilities-yes, food- no. Very insightful experience which I really enjoyed.

I very much enjoyed participating in this process as it was my first time on a committee. I will most definitely repeat the experience again! D. Bowen a.k.a. (Ron)!

Dan was an excellent facilitator!! Overall, the process was excellent. I would love to come back to this committee.

I honestly feel this was an excellent well thought out process. I truly have an appreciation for the test items and the categories. Great to establish barely attained and barely surpassed.

Food too carb related & sugary for breakfast, so hard to concentrate without protein. Process excellent!

Food was good, however more fresh vegetable/fruit options would be wonderful. Also coffee/tea available all day.

There were no healthy options for food (including the salad dressing). Breakfast and lunch should be reimbursed for those who need different options.

Breakfast & lunch options were disappointing. The group process worked well and allowed for valuable discussion that contributed to the final recommendations and our level of confidence w/ those results.

It would have been nice to have more choices for breakfast and lunch - four days of empty carbs was too much. Two choices at lunch would be preferable. Break stations on both levels would have been nice. Would have been nice to have whiteboard in room. Also, what are the frequently encountered and functional word lists? We were to make decisions based on info we didn't have.

Very informative & thorough process! Dan is a highly effective facilitator.

Jeff was an excellent facilitator. Just a little cold in the conference rooms.

Excellent process- fair, clear, informative. I would love to see teachers (educators) work on clear PLDs.

Breakfast was not gluten free friendly. PLDs were so vague it caused anxiety.

I was disappointed in the food selection. I would have liked to see healthier options. Unfortunately, we were not reimbursed if we chose healthier options.

Food service not the best. Jeff Barker was a great facilitator! Better food options would have been better.

More healthy choices for breakfast- yogurt, fruit, bagels. Overall this was a very enlightening experience.

Jeff Barker was wonderful to work with. He had a great way of facilitating. I am hoping they improve the temperature and food. The breakfast was awful (full of carbs/sugar) with no healthy options. Lunch as well. It would have been great in ELA to have the "word lists" (functional/context-specific)

Jeff B. was a great facilitator and really helped us through the process. The temp in the room could be a little warmer. Better food choices in the future.

Job Thomas was very helpful & encouraging. His easy & gentle manner allowed participants to feel at ease & allowed the process to go much smoother.

Job did a great job facilitating & leading our group. He ensured that everyone was heard & we understood everyone's thought process. Very kind, approachable, & a good facilitator.

Job did an excellent job facilitating the discussion. The process was also excellent.

Job did a great job making sure everyone's voice was heard.

Everything was great & I am happy that I was able to be a part of this process & would definitely do it again! Job was a great facilitator! If there was one improvement it would be healthier food options for breakfast.

The entire experience was enlightening & beneficial. Job did an excellent job in guiding the process and keeping our panel on task and focused. I really appreciate the opportunity to participate.

Overall a great experience. Job did a great job! Maybe healthier options for breakfast.

I felt that this was a very beneficial experience! Bravo to the MDE for allowing us, those in the "trenches," to make these important decisions.

I think the initial PowerPoint shown by Mike Bunch, and the practice bookmarking session caused a lot of confusion for people but once things were clarified, the whole process of creating bookmarks seemed to flow more naturally. Also, it would have been nice if the facilitators had been better exposed to the process as well as they seemed confused in the beginning as well.

Food was not good.

It would be nice to have some fruit at breakfast with less sugary food. Maybe one dark roast coffee.

The standard setting process was very effective. I appreciated having the opportunity to make recommendations for cut off scores. As educators in the classroom we see what students are learning and what they should be capable of doing.

Temp of room cool. Protein at breakfast would help.

Several participants did previous scoring & repeatedly kept saying "this is what we did the last time" - little irritating. Mr. Cramer listened and handled well.

Thank you for the opportunity to participate in this important activity. Great meeting place.

Excellent facilitator Management of topics - wonderful.

food was substandard especially breakfast

The instructor was great! She kept you engaged! Great group of educators!

I appreciate the opportunity to help in the assessment process. If I had any complaint it could be the lack of a better breakfast. I would prefer yogurt or some type of protein. All another arrangements were great. :)

Really enjoyed the opportunity to participate. Great job ensuring group diversity and maintaining positive group dynamics. Excellent job with the facilitating.

She was an awesome facilitator - she pushed at you to double check our angles & perspectives but never pushed us to expect other opinions as our own - she expected that we were thoroughly informed before deciding.

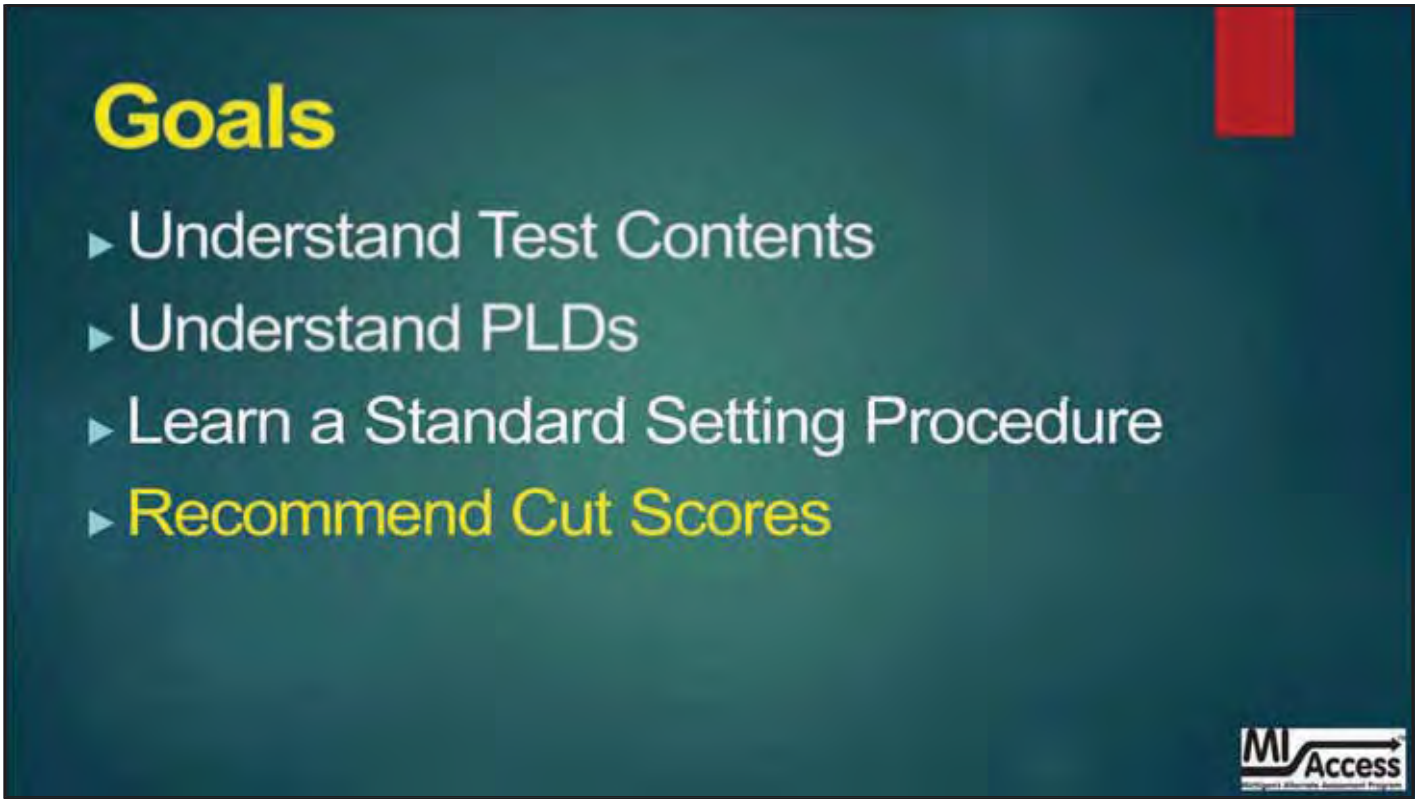
Breakfast is not a donut.

Appendix D
PowerPoint Presentations

- Overview June 15
- The Body of Work Procedure
- Vertical Articulation Training June 18
- Overview June 29
- The Bookmark Procedure
- Vertical Articulation Training July 2

MI-Access Standard Setting

June 15-18, 2015



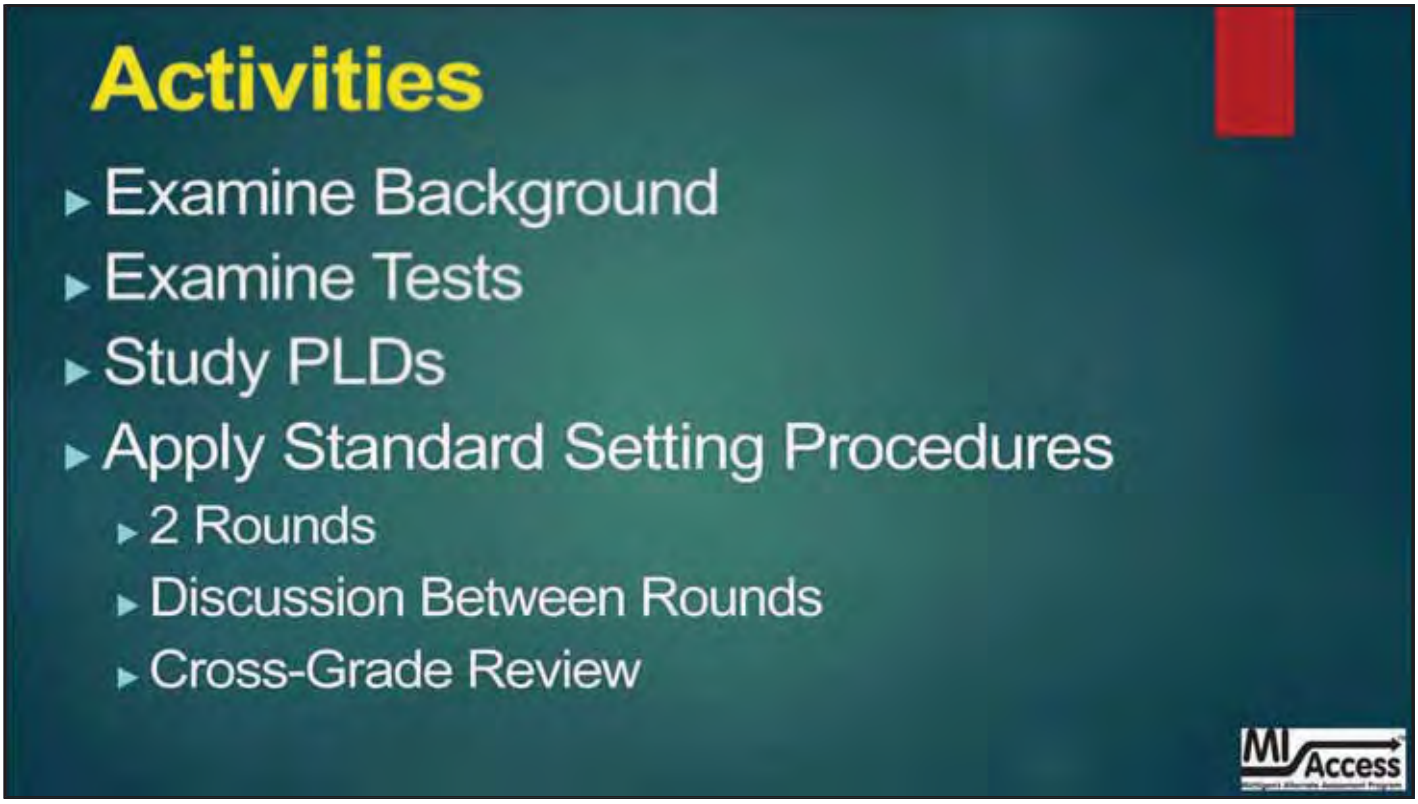
We are here to consider the tests students took this spring under MI-Access: Participation and Supported Independence; and to recommend performance standards – cut scores – to the State Superintendent. With that in mind, here are our goals for the next four days:

Understand Test Contents – In order to recommend meaningfully what scores a student should earn on a given test in order to be considered performing at the Emerging, Attained, or Surpassed levels, you should be very familiar with the contents of those tests, starting with the state content standards and ending with the individual items and their scoring rubrics.

Understand PLDs – We want you to be very familiar with the Performance Level Descriptors (PLDs) that describe what students at each performance level know and can do. Your recommended cut scores will translate those descriptions into numerical goals for students.

Learn Standard Setting Procedures – You will be using a process known as the Body of Work Procedure.

Recommend Cut scores – When all is said and done, the main thing you do this week will be recommend two cut scores for each test, one to separate Emerging from Attained performance and one to separate Attained from Surpassed performance. Everything else you will do will be primarily to prepare you to meet this specific goal.



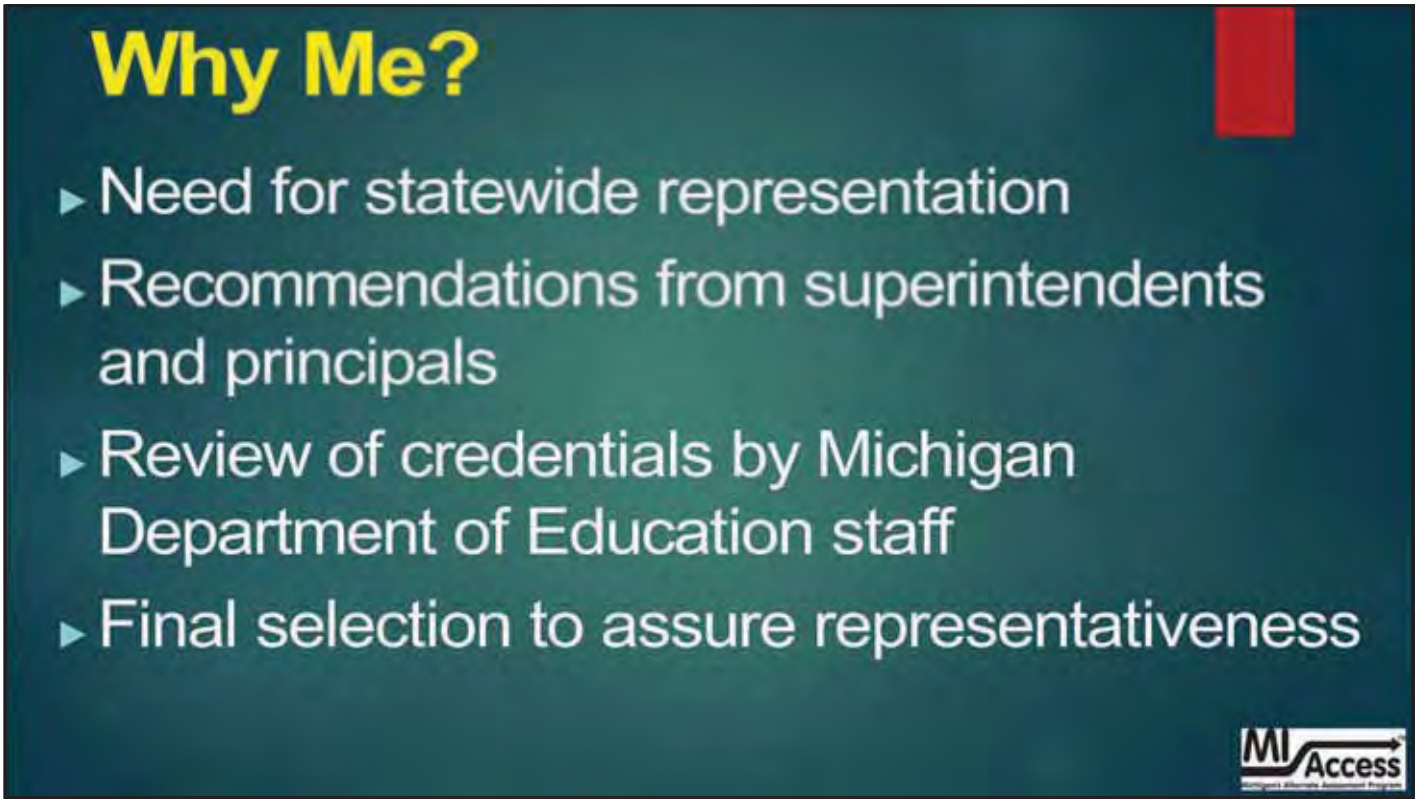
To accomplish these goals, we have developed a series of activities that will lead to the development of defensible cut scores. If you would take out your agenda, we will look at the activities we have planned for the next four days.

Later this morning, you will examine the tests in some detail. Afterwards, you are going to study the Performance Level Descriptors or PLDs in detail. Every recommendation we make this week must be firmly grounded in the Performance Level Descriptors. When we submit your recommendations to the Superintendent and when he acts on them, every action needs to be based on the PLDs because at some point, these performance standards will be reviewed by outside agencies, and the first question they will ask is whether or not we set our cut scores on the basis of clearly worded PLDs.

This afternoon, you are going to learn a specific way to use the information you have to make cut score recommendations. Let me state now, and we will no doubt remind you frequently later, that your job is to recommend cut scores. The State Superintendent will make the final decisions about cut scores, based on your recommendations and other considerations.

After you have learned the standard-setting procedure and had a chance to practice, each of you will then examine the test for your subject/grade band and recommend cut scores. We will tally results for your panel and share the results with you, after which time, you will discuss those results in your panel and do the same thing again. Between rounds of examining the tests, we will give you additional information to consider.

At the end of the week, some of you will participate in a cross-grade review. You will examine recommended cut scores for all grades or grade bands and consider whether some of them ought to be adjusted so that the overall impact of the cut scores would seem reasonable to parents, teachers, school administrators, and the general public. We will provide specific instructions on how that will work. The rest of you will participate in a debriefing about the process we are using this week. We will use the feedback you give us in our presentation to the Board and to help us improve the process for future standard-setting activities.



Having heard all this, you may be wondering how you happened to be chosen for this singular honor.

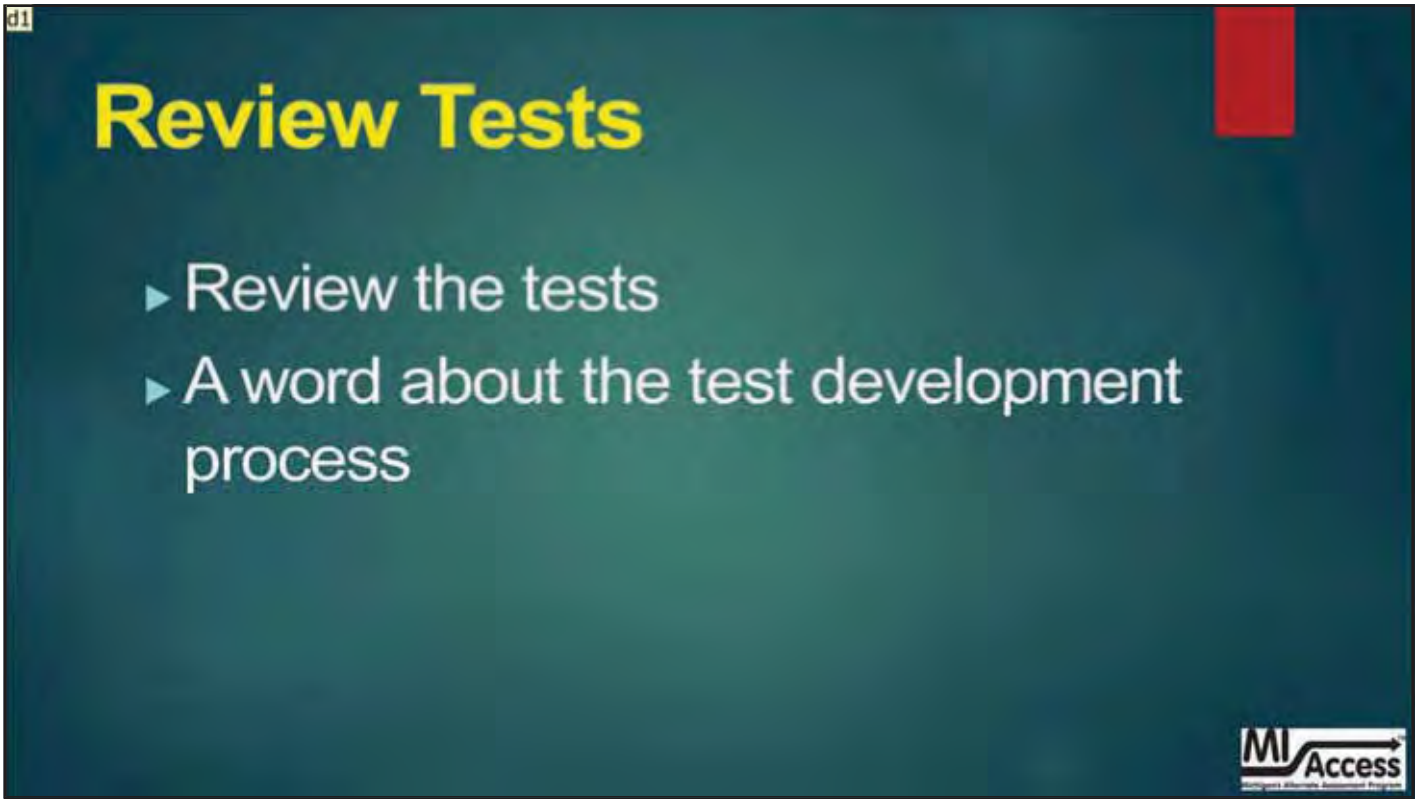
Standard setting is a high-profile activity, and we want as many people as possible involved in it. More importantly, we want those people to be representative of the State of Michigan as a whole. We have sent invitations to all parts of the state in order to find panelists who could fairly represent the state in terms of gender, ethnicity, length of service, and type of student population served. Staff of the Michigan Department of Education reviewed credentials of many people and chose you as the most representative and best qualified to carry out this important task. They put a lot of thought into selecting you, and we trust that you will put a lot of thought into what you do here this week. The performance standards we recommend this week, once approved or modified by the Superintendent, will be applied to all MI-Access Participation and Supported Independence students in Michigan not only this year but for years to come.

This may be a good time to provide a little background about standard setting and clarify some terms.
First, there are many different kinds of standards. Many people, when we refer to standards, automatically think of content standards. Actually, we couldn't set cut scores unless we had content standards that tell us what we are trying to teach and therefore what we are trying to test. But performance standards are numerical standards that specify how much we expect students to learn.

Part of the process of establishing performance standards is establishing performance levels. In Michigan, for the MI-Access test we have three:
3 – Emerging toward the standard
2 – Attained the standard
1 – Surpassed the standard

Each level has its own detailed description of what students at that level know and are able to do. No Child Left Behind requires states to spell out these levels with Performance Level Descriptors – PLDs. The law requires at least three levels; MEAP has four and MI-Access has three. Some states have even more. Groups of Michigan educators have drafted these descriptors, Michigan Department of Education staff have worked with those groups to refine and polish them, and the Superintendent ultimately has the responsibility for implementing them.

Again, the three performance levels are
1 – Surpassed the standard
2 – Attained the standard
3 – Emerging toward the standard



As I mentioned earlier, you will review the tests before you start recommending cut scores. A note about the tests: These tests were developed over a period of 2-3 years and have had considerable input already from groups of Michigan educators, for both content and fairness/sensitivity issues. You may not be thrilled with each and every item or how it is scored. That's OK. This is not the time to critique the tests because these are the tests we gave to students this year, and these are the tests from which they will receive their scores. If you see something you think is not as it should be, we would certainly like to know about it.

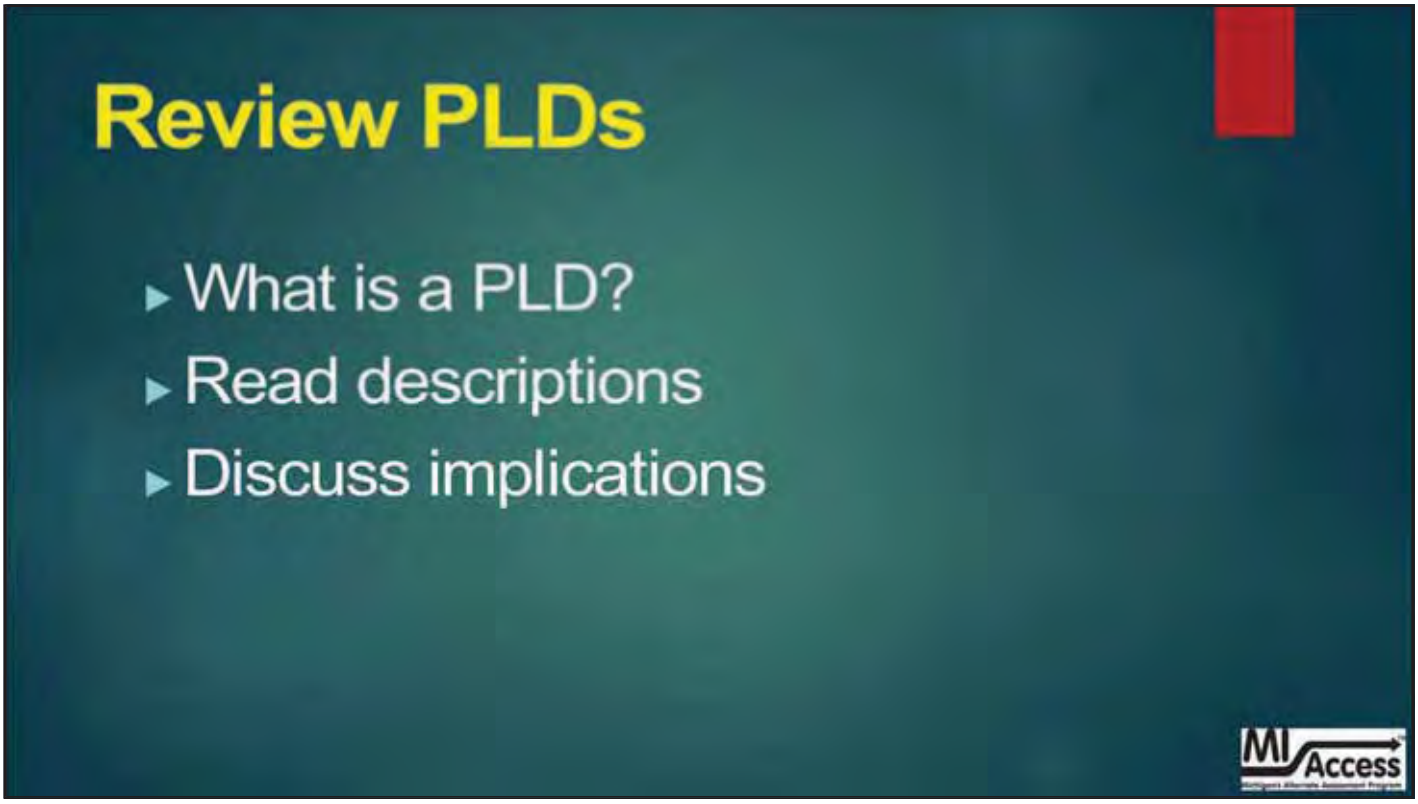
Please share your comments with us at the breaks or at other times, but we will not be conducting a test or item review during this workshop. It is appropriate, however, to keep those concerns in mind as you recommend cut scores. For example, if you see something that you believe would hinder an otherwise proficient student from answering an item correctly, you may take that into consideration when you make your cut score recommendation.

After you review the tests and receive instruction in the standard-setting procedure, you will be ready to start recommending cut scores. As you look at each test item , you may find something that you disagree with—again please feel free to mention it to one of us, and take that into consideration when you make your cut score recommendation.

After you review the tests, we will conduct a brief discussion about your reactions to them. Primarily, we will be interested in what you think it takes to answer particular items

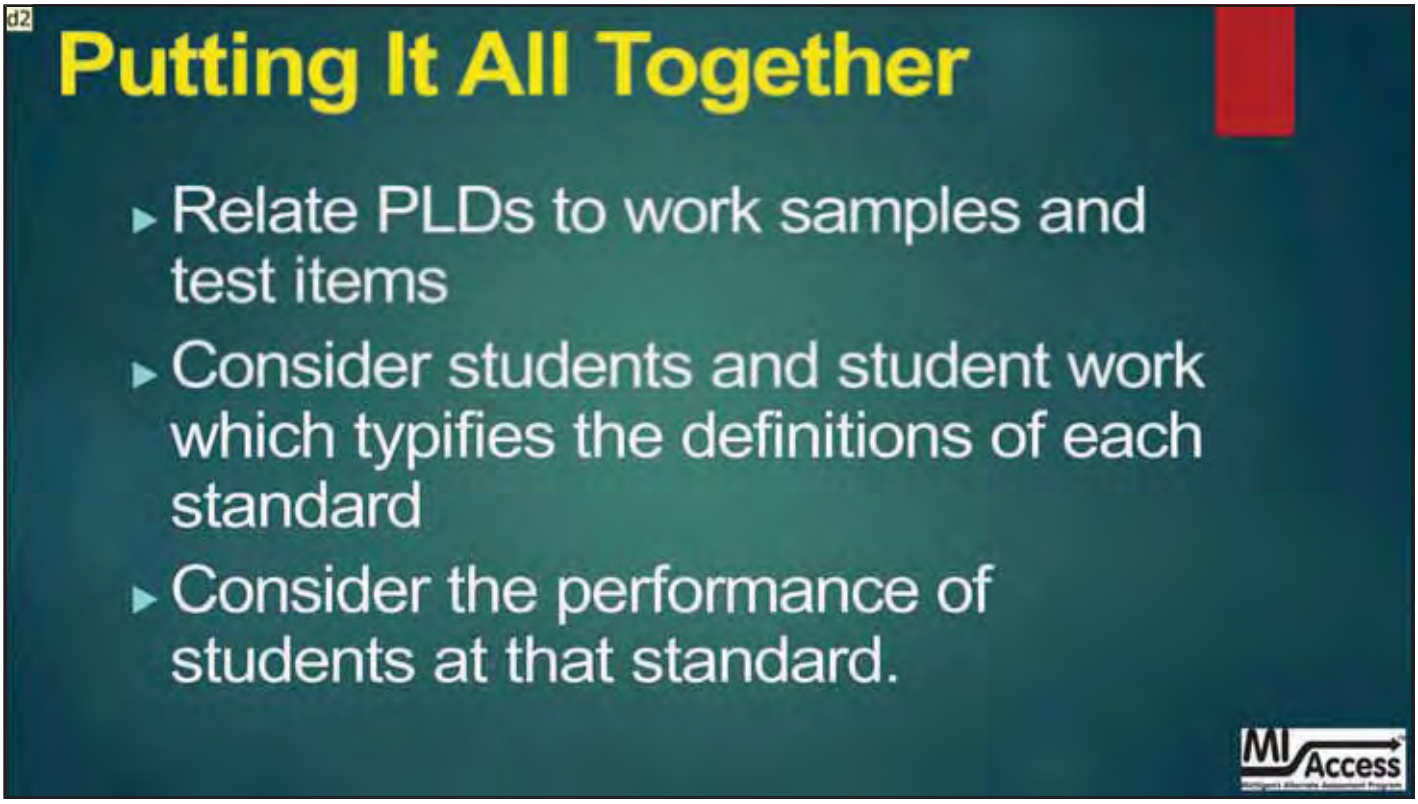
correctly or to receive high scores on them. Our chief aim is to find out what you thought was particularly easy or difficult and what you think would be particularly easy or difficult for Michigan students.

The purpose of this exercise is to have you become very aware of the content of the tests. These tests have gone through years of development. The items have been written, reviewed, field tested, and approved by several committees of Michigan educators. Your responsibility is to help set standards, not to criticize the tests. These exams may not be perfect but they are very good and contain content relevant to the Michigan standards and curriculum.

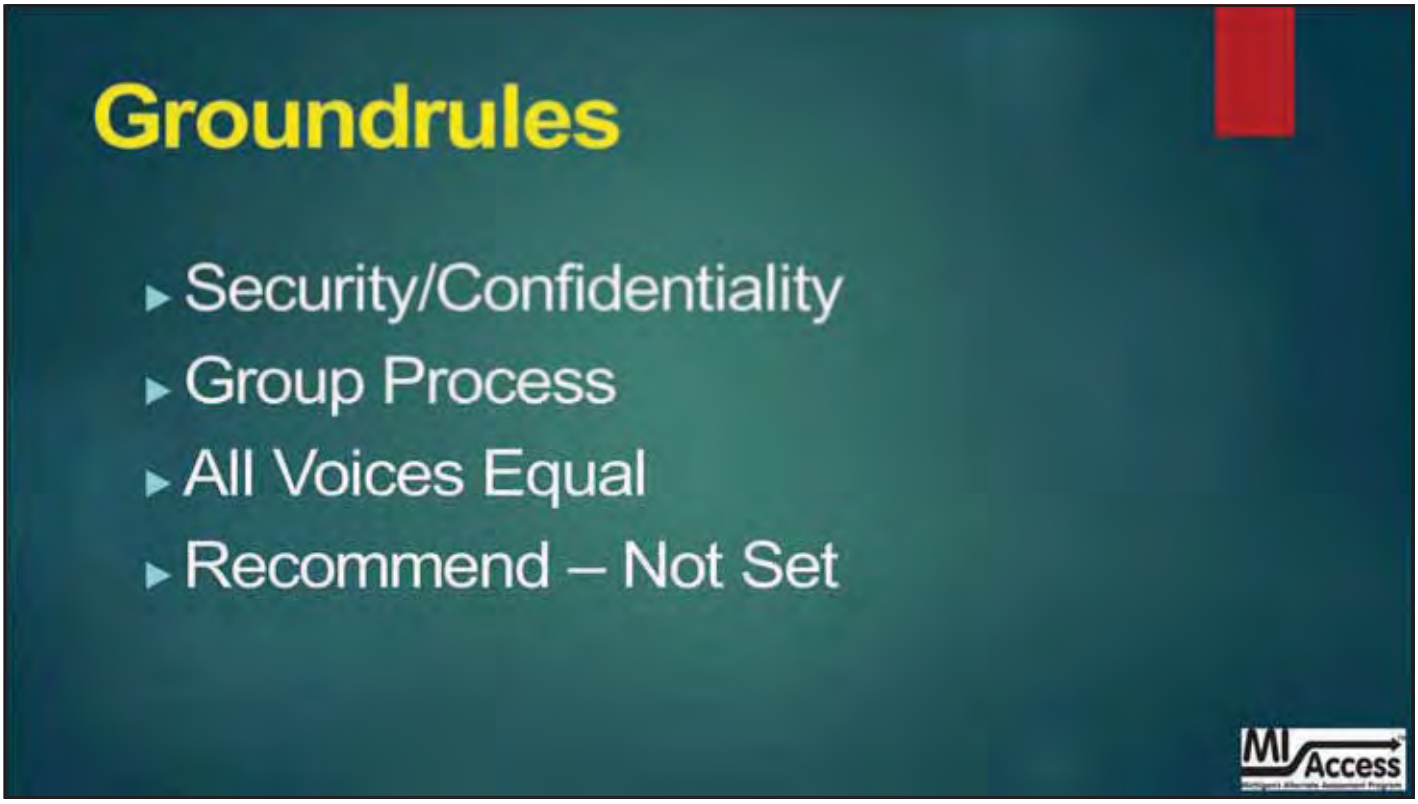


Right now, each of you probably has some idea what constitutes Emerging, Attained, and Surpassed. We also want to have a very detailed discussion of the PLDs. Again, a PLD is simply a description of the types of things that students at a particular level know and are able to do. The PLD for Surpassed (Level 1) describes things that these students can do that Attained (Level 2) students cannot do. Likewise, the PLD for Attained describes things that Level 2 students can do that Emerging (Level 3) students cannot do.

Later this morning, you will break into your separate groups, or panels, to study the PLDs for your grade band and subject. Read each description very carefully. Consider what it means to be at the Emerging, Attained, or Surpassed levels. Try to imagine students you have known who would fit the descriptions you are studying. Discuss those students with one another, and try to get a sense of the range of achievement within each level. Keep in mind that the PLDs are geared specifically to the state content standards that were used to create these tests. There are other aspects of performance that are not addressed here because they are not directly relevant to these content standards or to MI-Access.



So here's what you're going to do, once we get today's preliminaries out of the way:
First, you will examine samples of student work.
As you examine those work samples, you will consider the entire body of work relative to the PLDs.
After you have done that, you will make a decision about each work sample: Does it represent the work of a student performing at Level 3, Level 2, or Level 1?



Now, let me familiarize you with the groundrules for standard setting.
You will be seeing actual test items and actual student responses. We are going to ask you to sign a security/confidentiality agreement stating that you will not reveal any of these test items or any student response you might see while you are here. We will sign materials out to you in sequence and account for them by that same sequence when we take them back. We will identify those things that you can share with others. Anything else – you should consider secure and confidential.
You may have already gathered that this will be a group process. There will be activities that you will do completely alone, but we will have a lot of discussion. The purpose of the discussion is to allow everyone a chance to contribute and for each person to develop a greater understanding of the PLDs and how to interpret them, the standard setting process, and the rationales that were used in coming up with standard setting judgments.
We will encourage everyone to speak up during group discussions and will try to keep any one person from overshadowing others. In each round, we will ask each of you to cast, in effect, a secret ballot of your cut score recommendations, which we will tally. In the end, we will take the median of all your group cut score recommendations and report that as the group cut score recommendations. This will give each of you an absolutely equal voice in the final recommendations of cut scores.
Again – I am using the term "recommendation." Although the process is called standard setting, it is really standard recommending. We will work hard this week and employ a proven procedure that yields defensible cut scores. We will then ask the Department and ultimately the Board to consider not just our recommendations but the manner in which we arrived at them. In the end, we recommend, and the Superintendent sets the standards.

Questions?

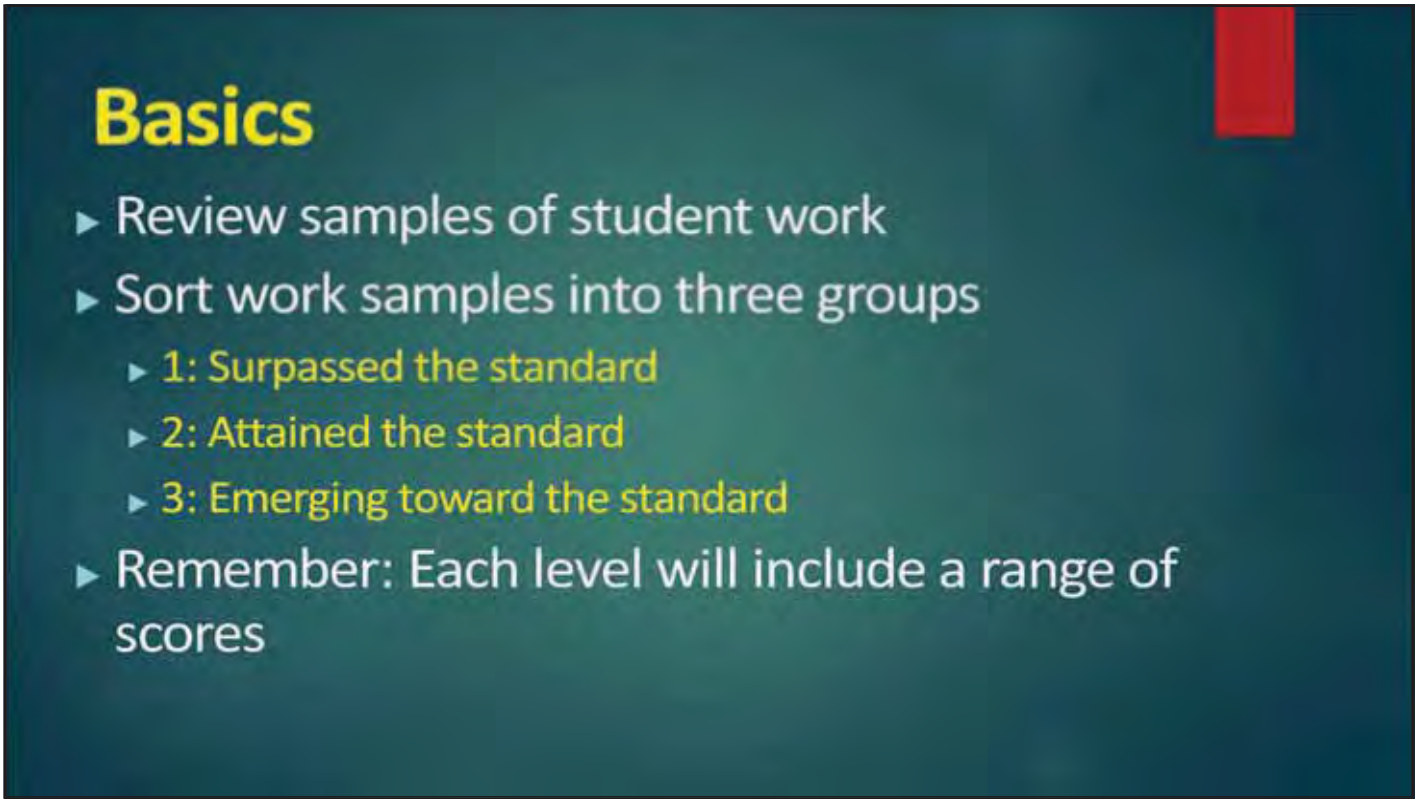
Panel Assignments

Panel	Room	Facilitator
Science 4/7	101	Corey Palermo
Science 11	102	Karen Kemp
ELA 3-5	Governor's Room	Tom Kelsh
ELA 6-8	Banquet Room 7	Craig Deville
ELA 11	Banquet Room 8	Tracy Robertson
Math 3-5	103	Winnie Reid
Math 6-8	104	Christina Luke
Math 11	203	Dan Bowen

Introduce facilitators and dismiss by room. Panel assignment will be on Panelist Packet.

MI-Access
Body of Work
Standard Setting Procedure

June 15-18, 2015

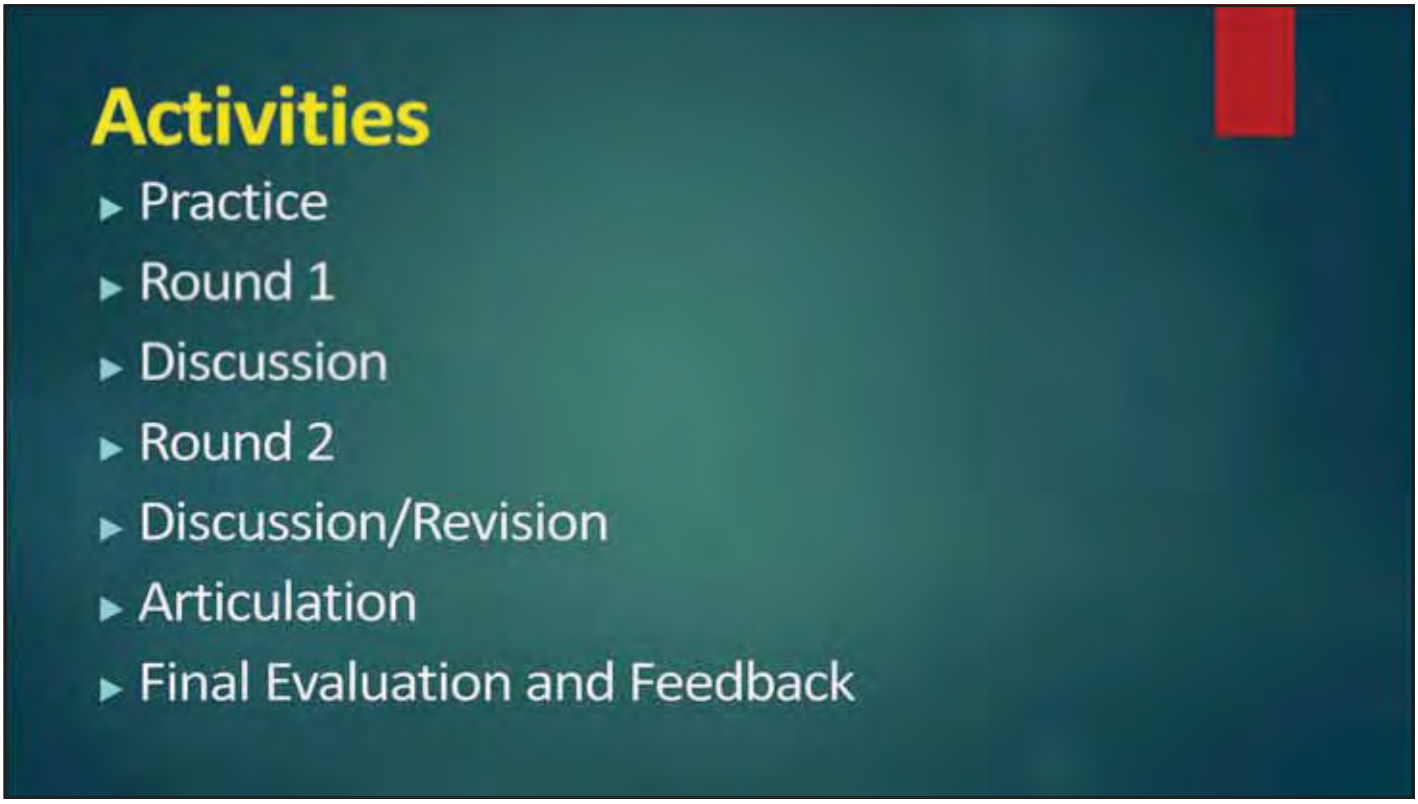


We have already scored all the tests. We know how well the students did individually and what the overall distributions of scores are. We have taken actual work samples from students that represent the range of total scores on the tests. Each work sample consists of the Scoring Document completed by the Primary and Shadow assessment administrator. Thus, for example, on a 10-item Participation Scoring Document with each item being scored on a 0-3 scale, it is possible to earn up to 30 points. When we sum the Primary and Shadow Assessment Administrators' scores, the total possible is 0-60. You will see some work samples with scores near 60 and some near 0. Most will be somewhere in between.
You will be asked to examine each work sample and place it into one of three groups:

1. Surpassed the standard
2. Attained the standard
3. Emerging toward the standard

To simplify the process, we have arranged the work samples from lowest to highest score. The first sample in the packet, say with a score of 0 or 2 would likely fall into Level 3: Emerging. How would you rate it? At some point, you should reach samples that you will have real difficulty assigning to Level 3 or Level 2. That's fine. You are probably nearing the score that divides Level 3 from Level 2. Even after you get to a sample that you think should be in Level 2, you may find that the next one seems to be a better fit in Level 3. That's fine too. After a while, you will find that you are assigning most samples to Level 2. Then you will reach a point at which you struggle deciding whether a sample belongs in Level 2 or Level 1. That's because you are probably approaching the cut score that separates Level 2 from Level 1.

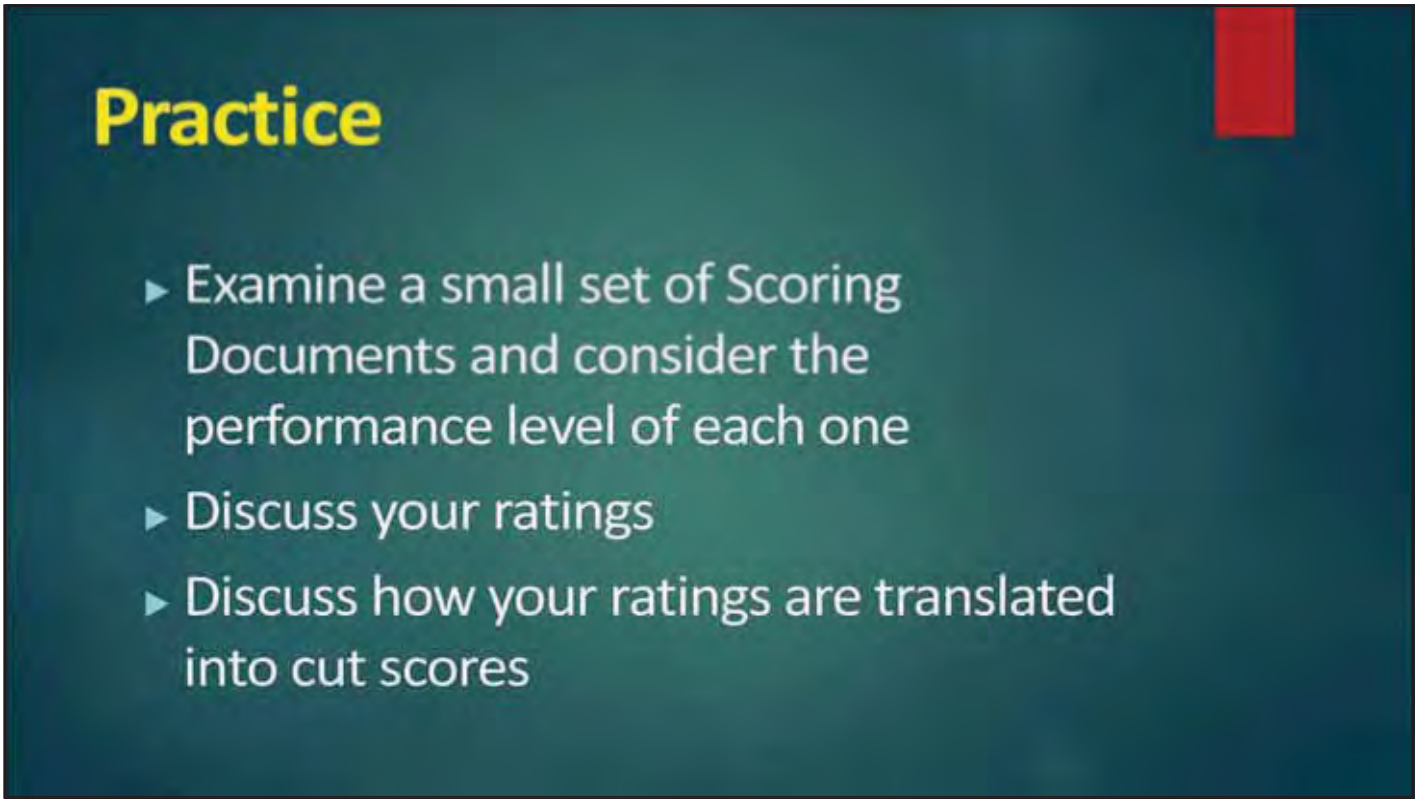
As you rate the samples, keep in mind that each level encompasses a fairly wide range of performance. For example, Level 1 will include perfect scores as well as scores that would just barely make it into Level 1. Level 2 will include scores that almost made it into Level 1 as well as those that just barely made it into Level 2. Level 3 will include scores that almost made it into Level 2 all the way down to 0.



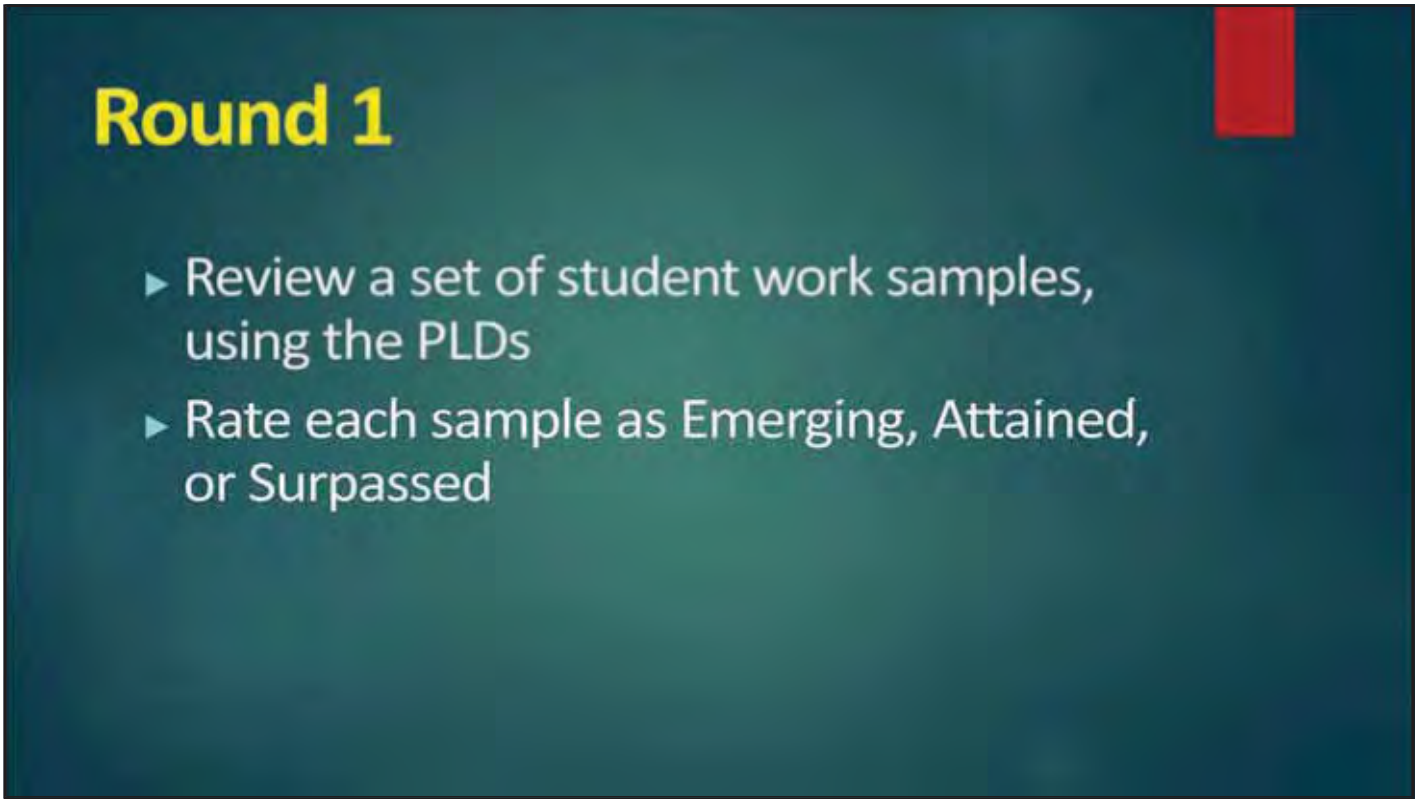
After we explain the Body of Work procedure, you will have a chance to practice sorting a small number of work samples. We will talk about that, answer your questions, and then start on Round 1. We will go through two rounds of reviewing work samples. After Round 1, we will discuss where the dividing lines appear to be, and then you will review some more work samples. Between Round 1 and Round 2, we will remove some of the work samples that didn't seem to contribute much to the placement of cut scores and insert some new ones with scores closer to the ranges where cut scores seemed to be. Before you start Round 2, we will share student score distribution data – also called impact data – with you so that you can see how many students would be classified at each level, based on your rough cut scores from Round 1.

As we mentioned earlier, after Round 2, some of you will go into a cross-grade review session; the rest of you will take part in a debriefing concerning the Body of Work procedure and how we implemented it.

We will also check to make sure that you understand what you are being asked to do, and we will answer any questions that you have throughout the course of the process. You will have the opportunity to provide us with feedback, how you understood different components of the process, and how you arrived at your judgments. These data will be collected in readiness and feedback forms and they are very important to us to make sure that process is working as intended. You also will have the opportunity to complete a final evaluation form of the whole standard setting process. Please fill out this information and provide us with your opinions.

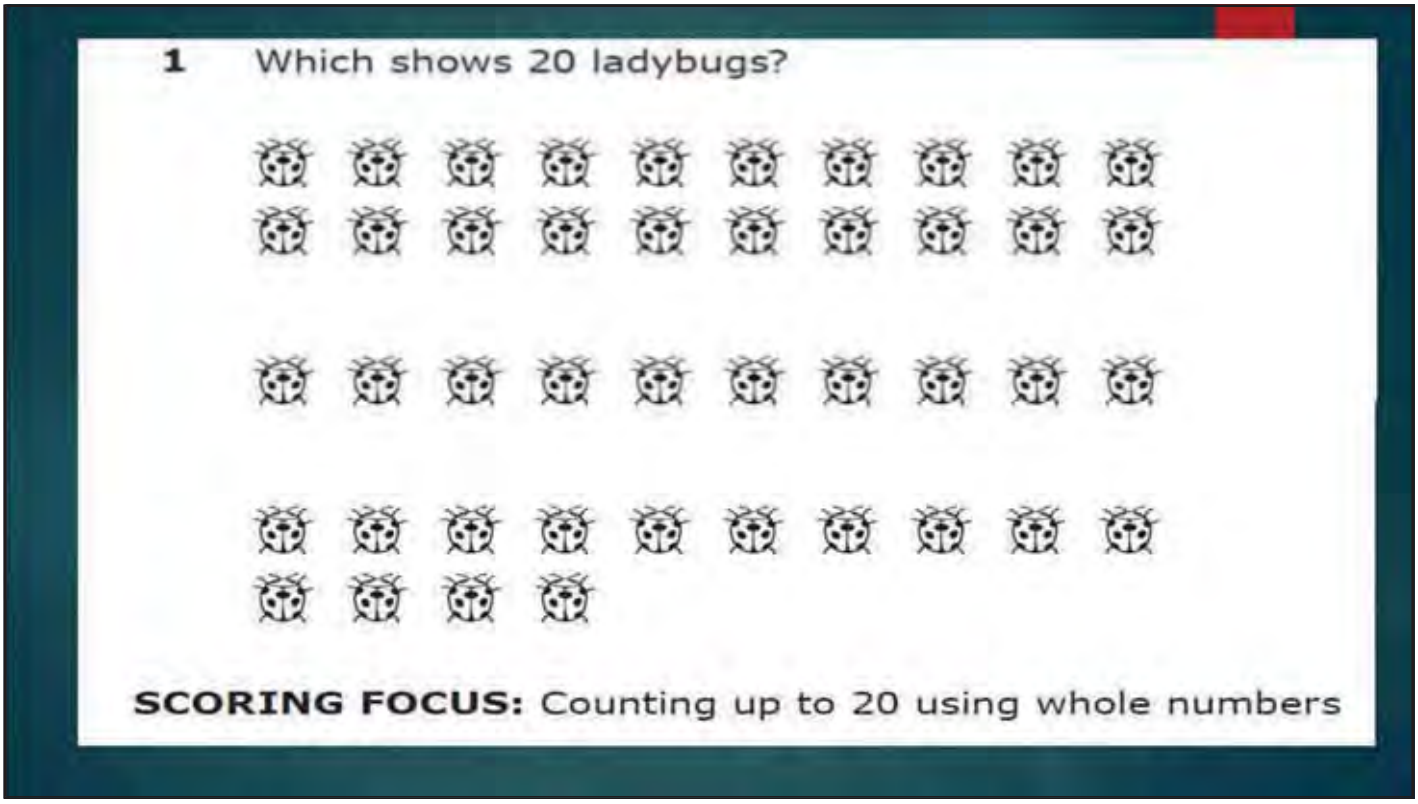


The practice set will consist of a small sample of Scoring Documents. We will ask you to examine these Scoring Documents and decide which ones represent Level 1, Level 2, and Level 3. After you do that, we will discuss your ratings. We will not have you set cut scores in this practice round, but we will discuss how you would go about providing actual cut score judgments from the materials that you received. The practice round is designed to get you acclimated to what you are being asked to do and to ask specific questions before we get started.

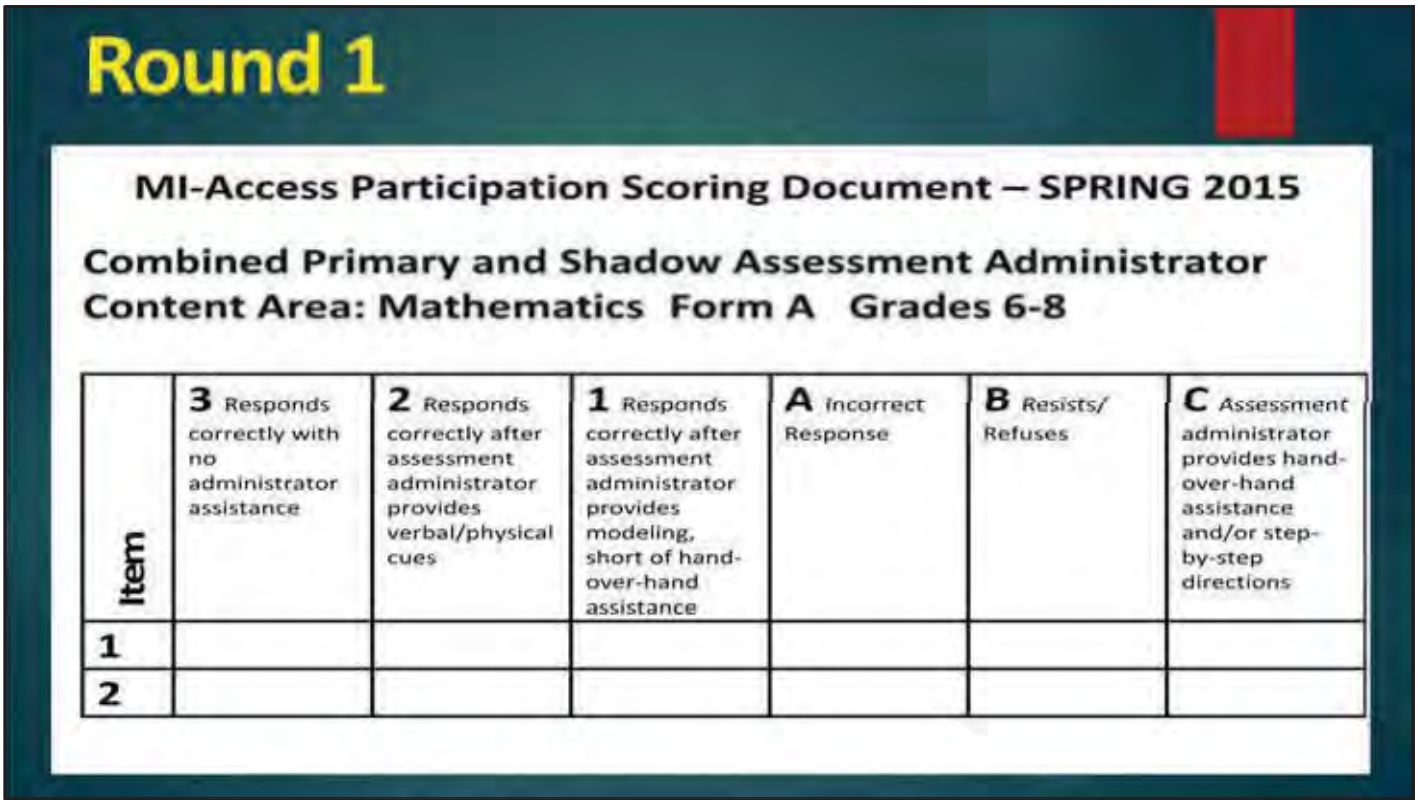


In the first round of standard setting, you will review a large sample of student work samples that spans the range of possible scores that a student can receive on the assessment. Keep in mind that the first sample will have the lowest score, and each succeeding sample will have a higher score than the one preceding it.

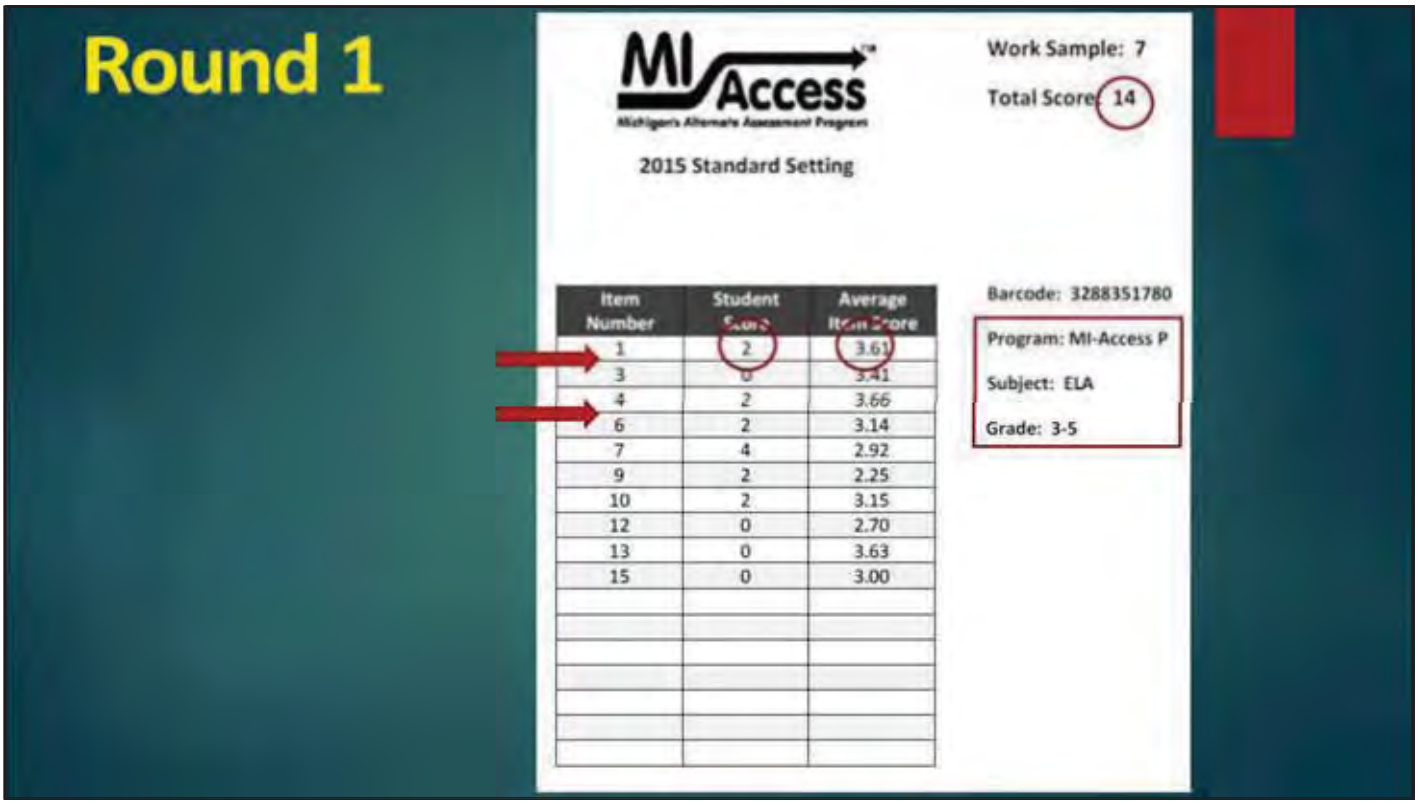
Let's take a look at how this will work.



As you will recall from earlier today, this is an example of the types of questions students will be asked to answer. They can answer these questions with no help, with physical or verbal cues, with modeling, or with hand-over-hand directions. Each type of response earns a different score.



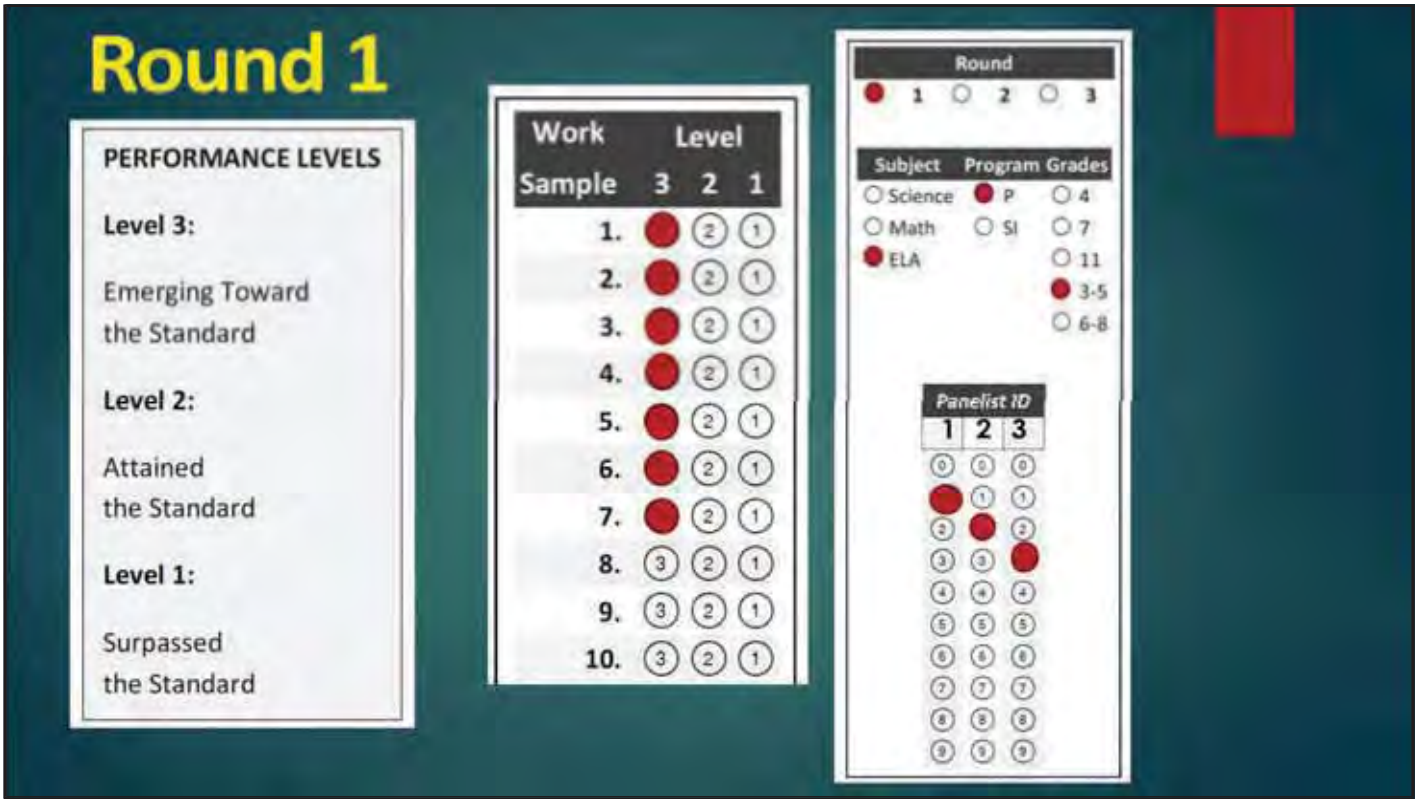
This is an example of a Scoring Document for **MI-Access Participation** that the Primary and Shadow Assessment Administrators used this spring to enter student scores. You have a sample in your packet. It's on page 67 of the Administrator Manual. Locate that, and let's take a moment to go over the scores and comments. Let's start with the top part showing what the scores mean. [**Review all score points** and A, B, and C. Ask for questions and answer, with assistance from MDE staff as necessary.] There's a similar one for Supported Independence on page 68. The video you watched this morning really covered all you need to know about this.



Here's an example of what you will actually be looking at. This particular set of work samples is for ELA grades 3-5/Participation. You will have a set of about 30 score sheets for Round 1, arranged in order from lowest to highest score. On each score sheet, you will see not only the total score the student got but the score that student got on each item as well as the average score for that item. Notice also, that there will be some gaps in the numbering of the items. Here, for example, you will see that there is no item number 2. Item 2 was a field test item and did not count toward the total score. Items 5, 8, 11, and 14 were also field test items and did not count toward the total score.

Keep in mind that for Participation, the highest score for any item is 3, and when we combine the scores of the primary and shadow raters, the highest total score for any item is 6. Thus, the highest total score for a 10-item test is 60. This student got a score of 14, so this work sample would be near the beginning of the set (actually #7 in this set).

Having examined the items for this test and seen the scores this student got on each item, your task will be to decide whether this particular student's test performance is most like that of a student who is Emerging toward the Standard, a student who has Attained the Standard, or a student who has Surpassed the Standard, based on the definitions of these performance levels given in the Performance Level Descriptors. This is the heart of the Body of Work Procedure, so I want to pause here to make sure this is clear. Are there any questions about this so far?

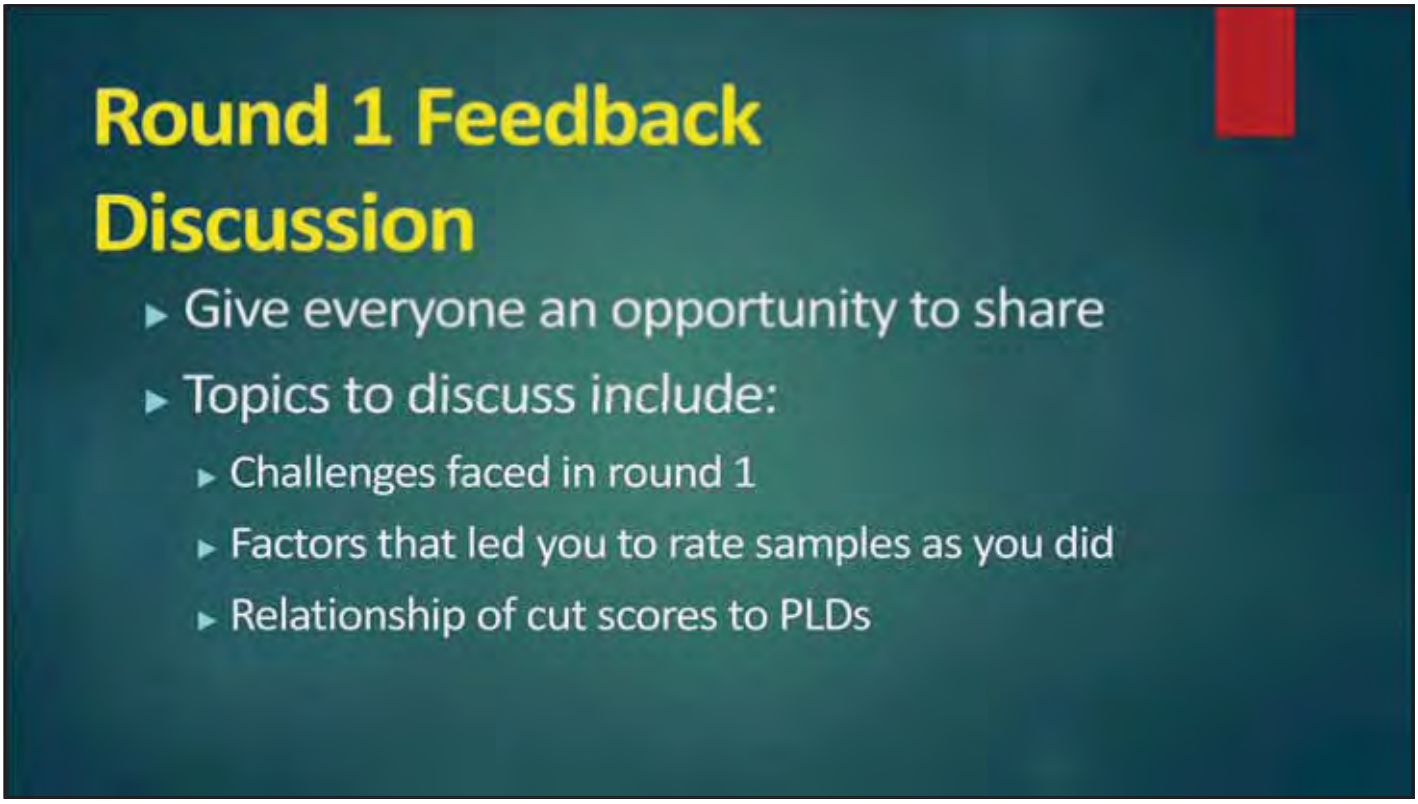


You will be completing forms that look like this. You have a panelist number on your packet that you will enter on the right side of the form and then bubble in the numbers below so that when we tally the results, each of you can get your rating sheet and comments back for the inter-round discussion. For example, if I am Panelist 123, working in the ELA grades 3-5 group reviewing work samples in Round 1, here's how I would fill out my information for Round 1…

Now, let's think about work sample 7 that we just looked at. It had a score of 14 out of 60 possible points. Knowing this student's scores on each item, as well as how well other students did on the items, and comparing the overall performance with the PLDs, let's say I decide that this performance clearly falls into the Emerging category, so I give it a rating of 3: Emerging. Once I make that decision, I bubble a 3 on line 7… I notice that I have entered a 3 for every work sample so far. Don't worry; remember that these are in score order. You will eventually see work samples you will want to give a rating of 2 (Attained) and later on work samples that you will want to give a rating of 1 (Surpassed). Others at my table may not agree with some of my ratings, and that's fine. You should discuss these work samples with others at your table, but when you enter your rating (1, 2, or 3), that should be your rating, not the group consensus. Of course, if everyone at the table agrees, that's fine too.

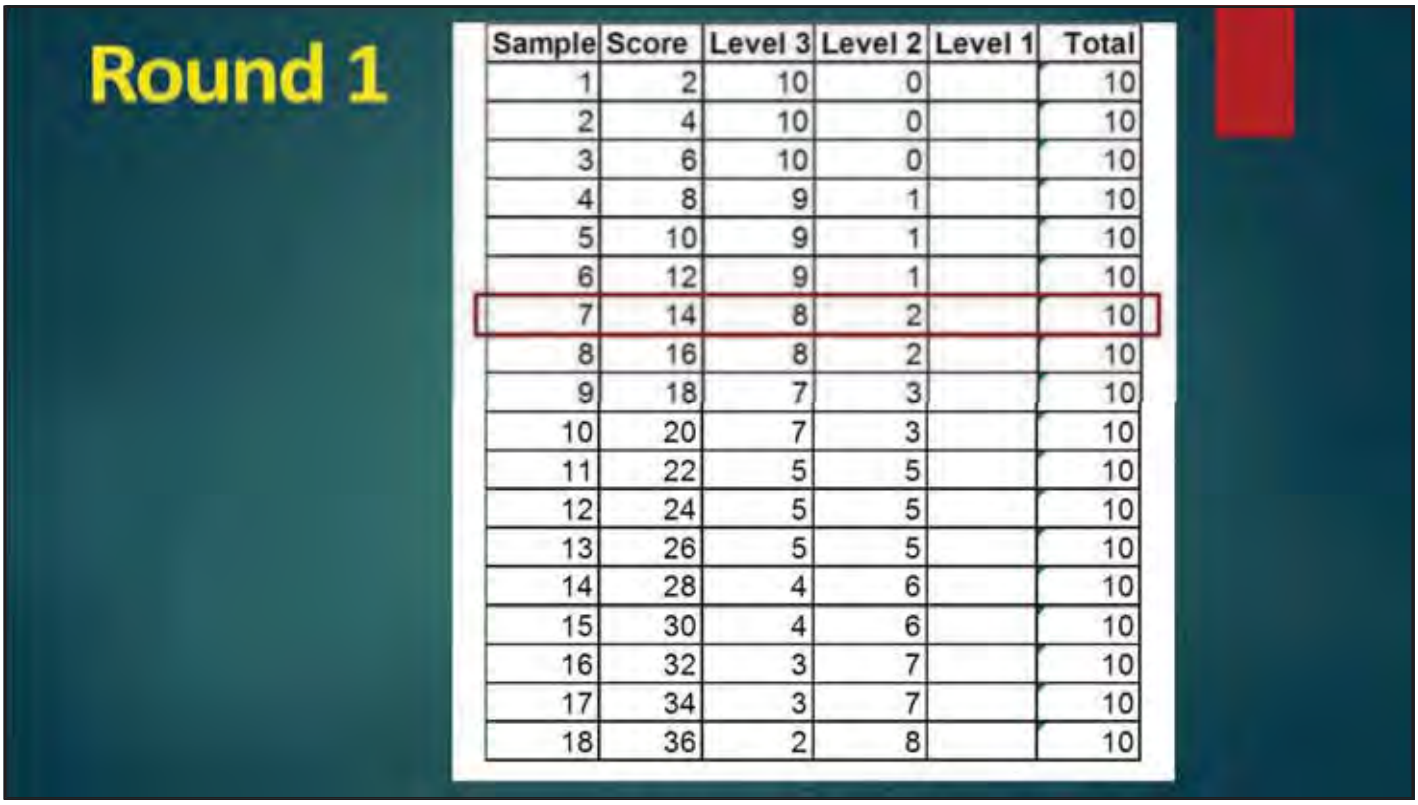
After you have completed your Rating Form for Round 1, you will turn it in to your facilitator. He or she will check all your materials, make sure we have everything we are supposed to have, and give you your next assignment or dismiss you for the day. Some of you will finish earlier than others. That's fine.

While you are having dinner and a good night's sleep, we will tally all the results, translate them into cut score ranges, do some more calculations, create tables and graphs, and have them ready for you to study and discuss tomorrow.



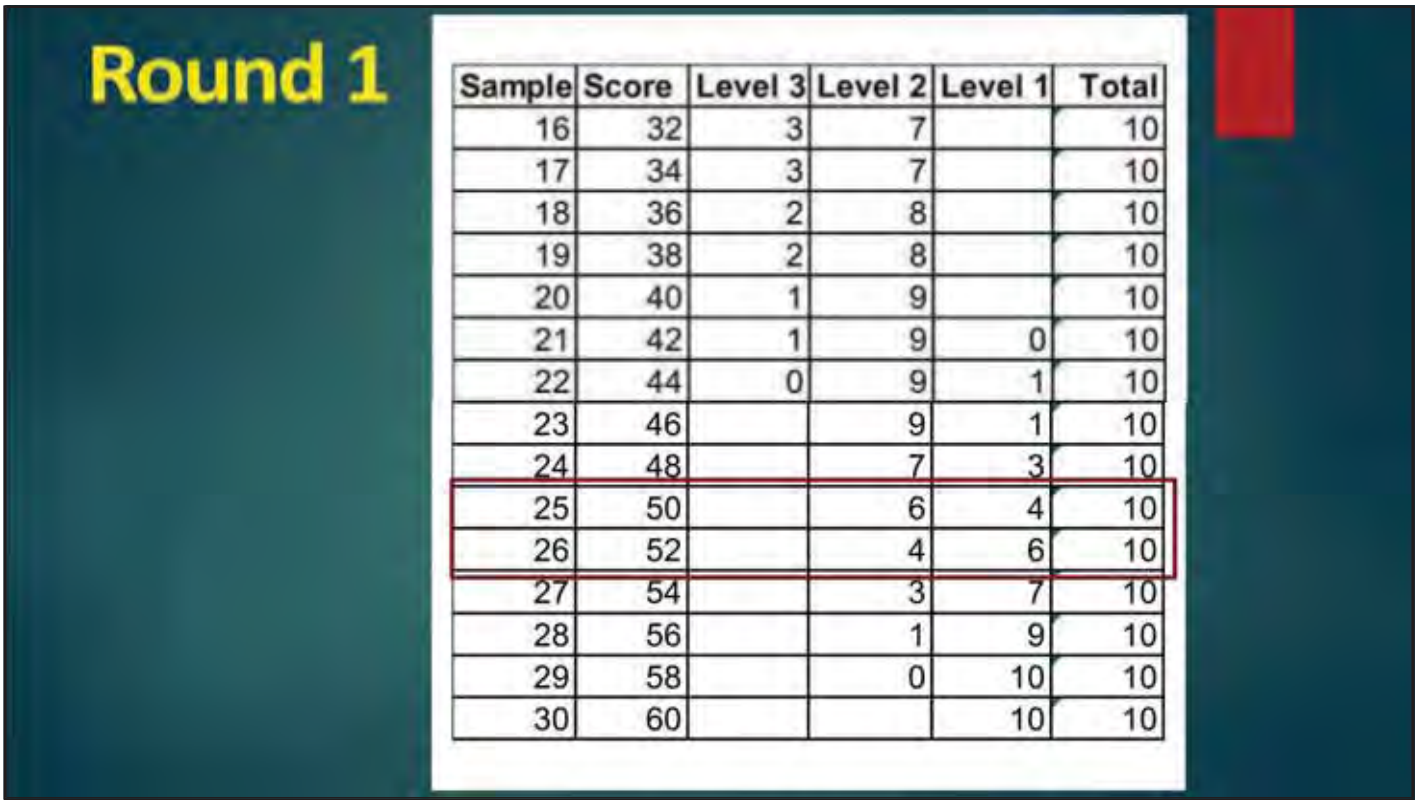
After we have completed all our calculations, we will share the results of Round 1 with you. We want to make sure everyone has a chance to be heard and that no one dominates the conversation. We would like you to discuss any challenges you faced in Round 1, any samples that were particularly difficult to categorize, what influenced your decisions, and how you used the PLDs to make your ratings.

If your judgment is different than the rest of the group's that's okay we will discuss this and give you the opportunity to provide new recommendations in Round 2. This discussion is important for everyone to check in and understand how others arrived at their judgments. Here's what you will see as you have these discussions…



Sample	Score	Level 3	Level 2	Level 1	Total
1	2	10	0		10
2	4	10	0		10
3	6	10	0		10
4	8	9	1		10
5	10	9	1		10
6	12	9	1		10
7	14	8	2		10
8	16	8	2		10
9	18	7	3		10
10	20	7	3		10
11	22	5	5		10
12	24	5	5		10
13	26	5	5		10
14	28	4	6		10
15	30	4	6		10
16	32	3	7		10
17	34	3	7		10
18	36	2	8		10

For each set of work samples, we will tally the number of Level 1, Level 2, and Level 3 ratings in a table that looks something like this.
Recall our previous work sample #7 that I rated Level 3 (Emerging). Most people agreed with me, but 2 people in my group saw this sample as Level 2: Attained. We'll have a chance to talk about why some of us saw this as a 3 and others as a 2 when we discuss our Round 1 ratings.

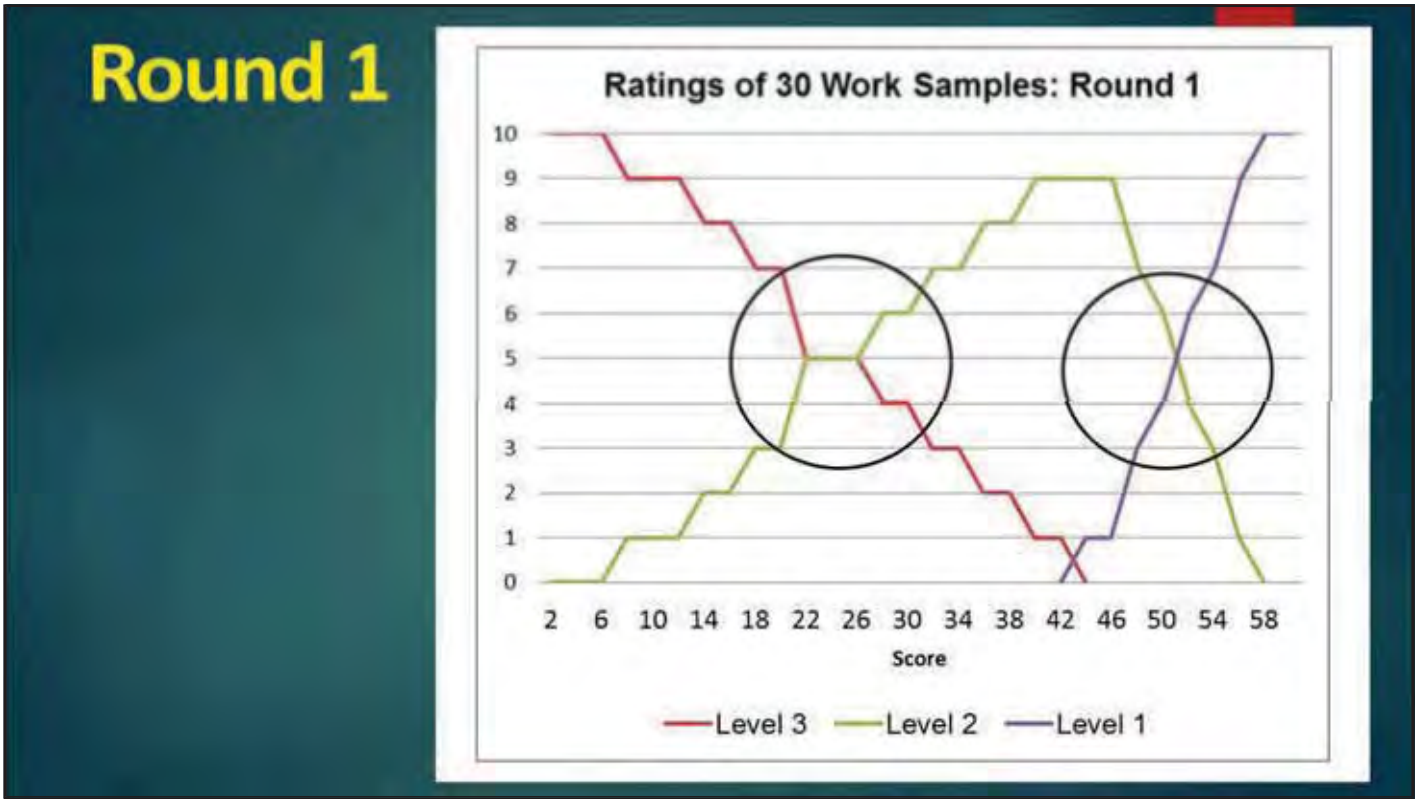


Sample	Score	Level 3	Level 2	Level 1	Total
16	32	3	7		10
17	34	3	7		10
18	36	2	8		10
19	38	2	8		10
20	40	1	9		10
21	42	1	9	0	10
22	44	0	9	1	10
23	46		9	1	10
24	48		7	3	10
25	50		6	4	10
26	52		4	6	10
27	54		3	7	10
28	56		1	9	10
29	58		0	10	10
30	60			10	10

Here's the rest of that table.

We'll want to talk about these tables, particularly about the work samples that led to large differences of opinion. For example, look at Samples 25 and 26: The room was fairly evenly divided on these two, and it looks like we may have found the point where Levels 2 and 1 meet. At this point, we would want to take out our PLDs and ask someone to explain why Sample 25 should be placed in Level 2, based on its contents, relative to the PLD for Level 2. Then we would ask someone to do the same for Level 1. We would repeat the process for Sample 26. We believe this discussion will help everyone fine tune their understanding of the PLDs and how they apply to the work samples. This is the primary reason we don't just do this once.

But wait; there's more!



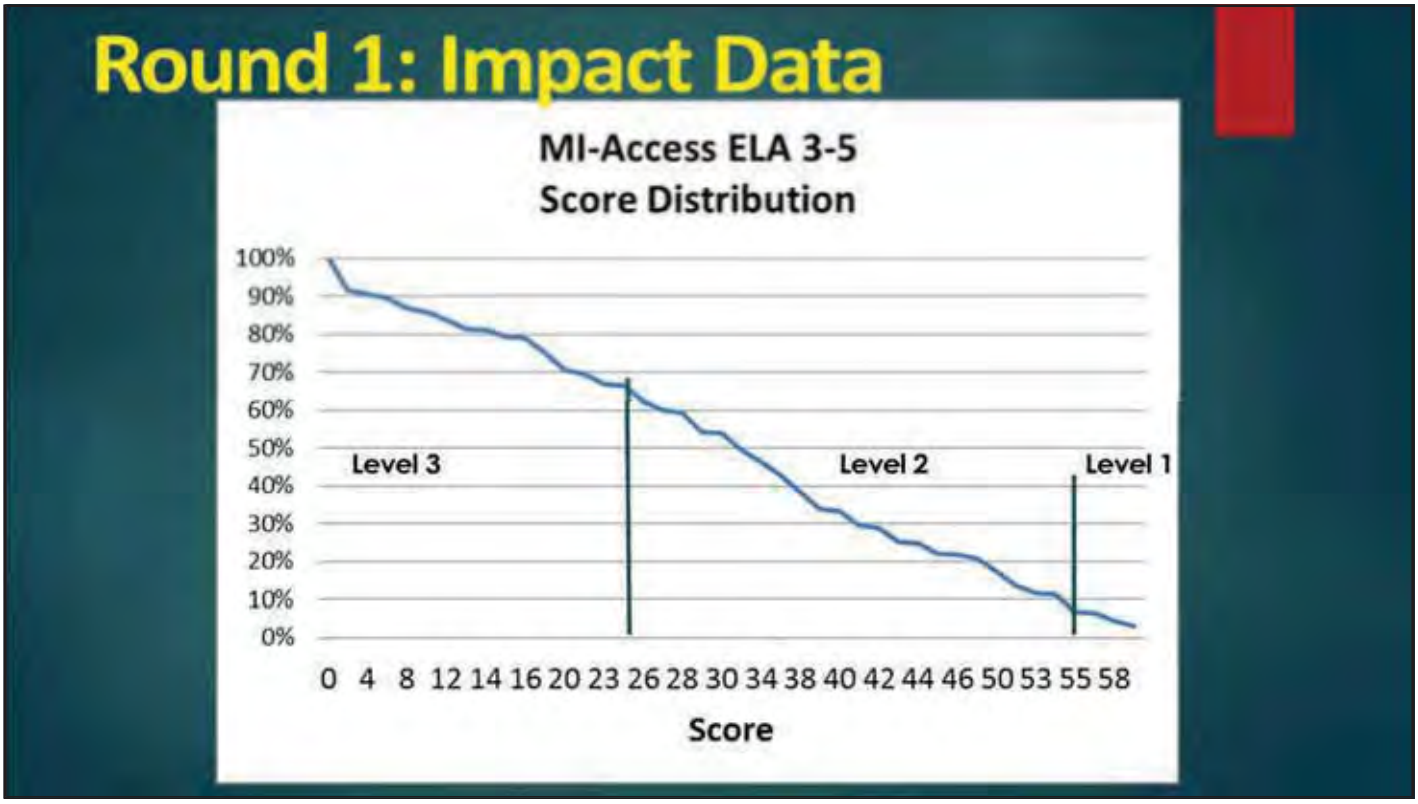
We will also plot those ratings against the scores in a chart that looks something like this. By the way, this is a made-up example. [Explain structure of chart.]

Notice first that we have arranged scores from low to high. Look where the Level 2 and Level 3 lines cross – at around 22-26 points. Below a score of about 18, hardly anyone assigned any work samples to Level 2, so in Round 2, we probably wouldn't look at many samples with scores below 16 or 18.

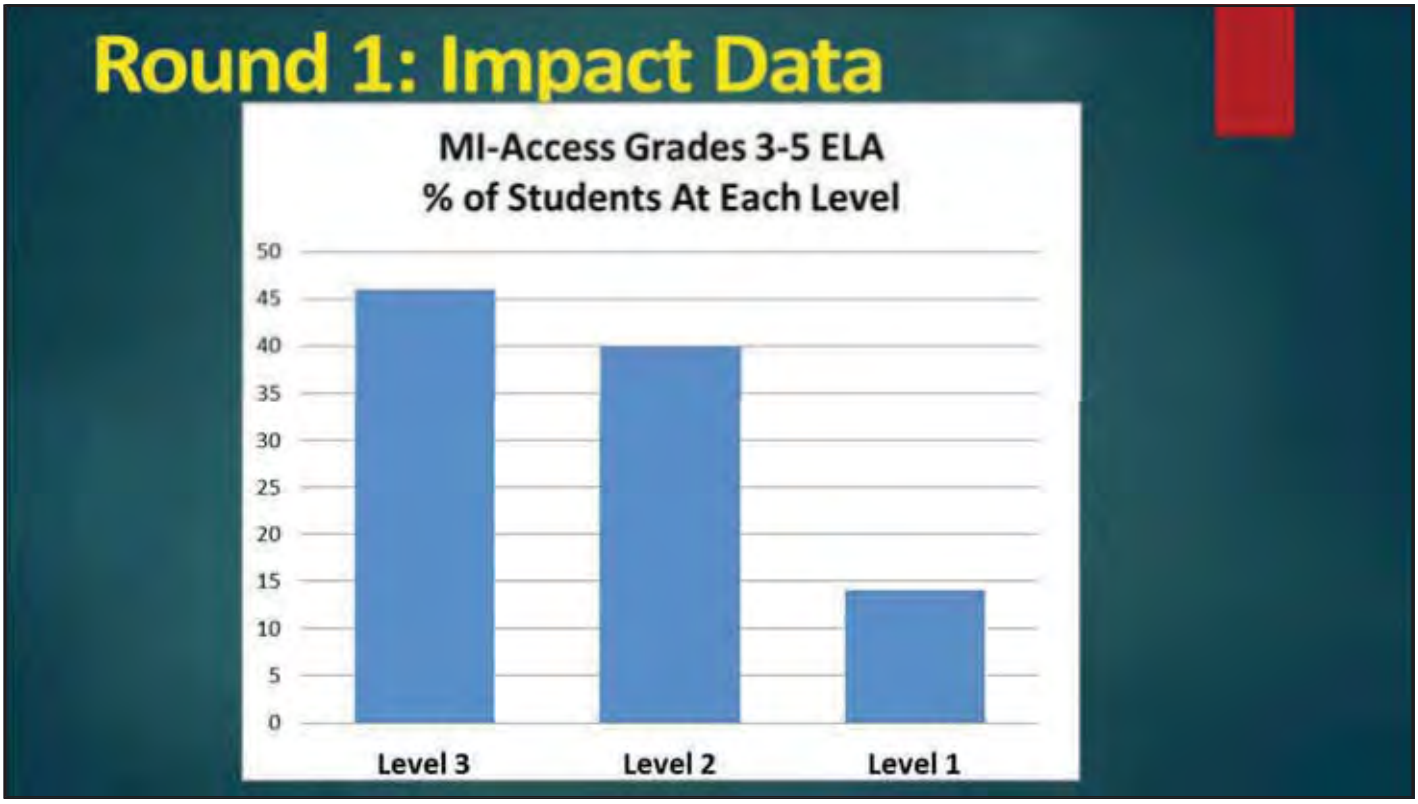
Now notice that the Level 1 and Level 2 lines cross at a score of about 52 in this example. Above a score of 56, hardly anyone rated any samples below Level 1. Therefore, in Round 2, we probably wouldn't look at many samples with scores above 54 or 56.

We haven't actually set any cut scores yet, but we now have a general idea where they might be. Therefore, while we were preparing these charts, we were also culling from the Round 1 work samples all the samples with scores above 56 and below 16 or 18 and reinforcing the packet with samples having scores between 16 and 56. In Round 2, you will have more samples that will challenge your decision-making process.

Before we move on to Round 2, however, we have one more bit of information we will want to share with you…



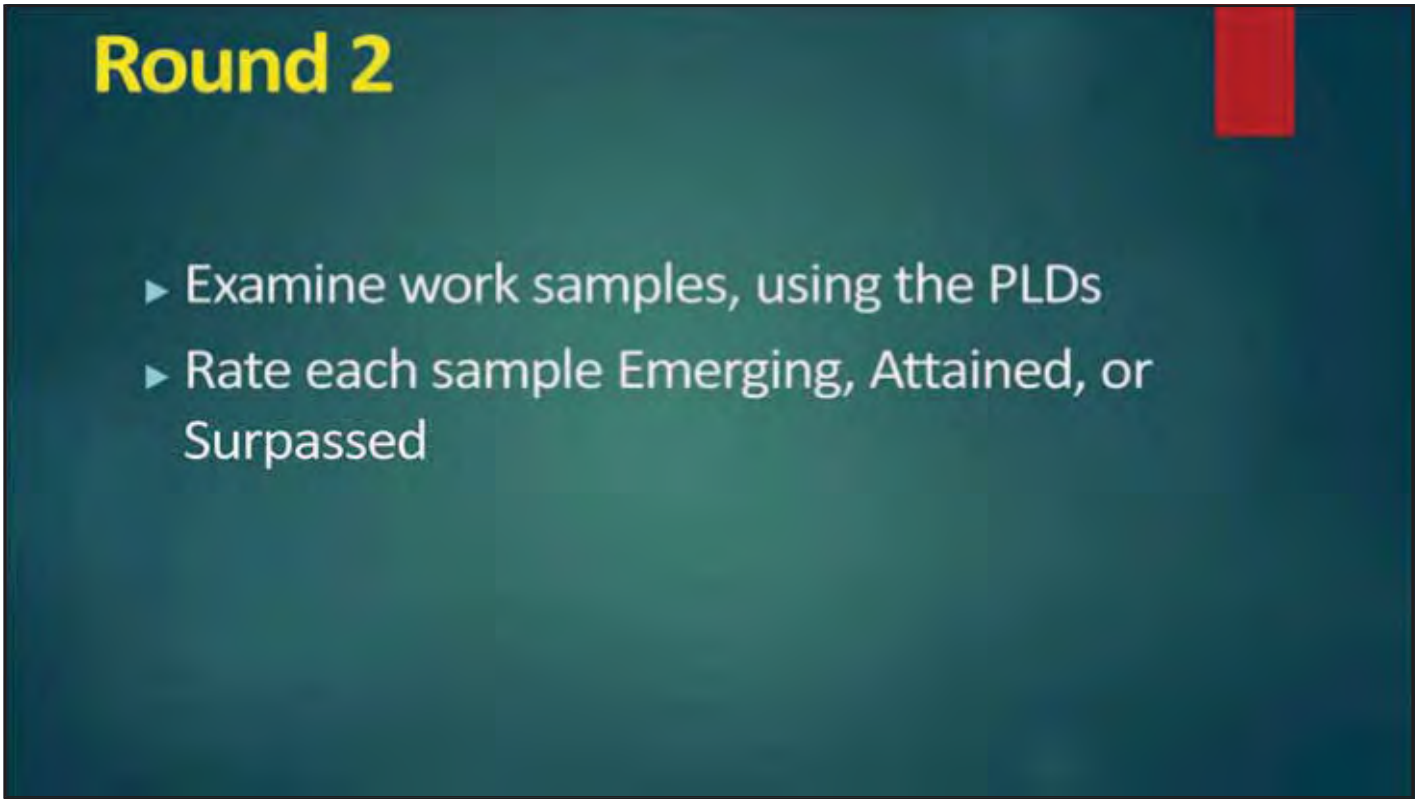
After Round 1 we will also be giving you impact data: percentages of students at or above each score, based on this spring's scores. We will superimpose the rough cut scores from Round 1 on the impact table or chart so you can see the relationship between cut score and impact. This is one way of looking at it – percentages of students scoring at or above each score point, with lines separating Level 3 from Level 2 (at about score point 24-25) and Level 2 from Level 1 (at about score point 55). But that's just one way of looking at the results. Here's another…



For each test, we will use the cut scores to determine how percentages of students at Level 3, Level 2, and Level 1, based on the spring 2015 test performances.

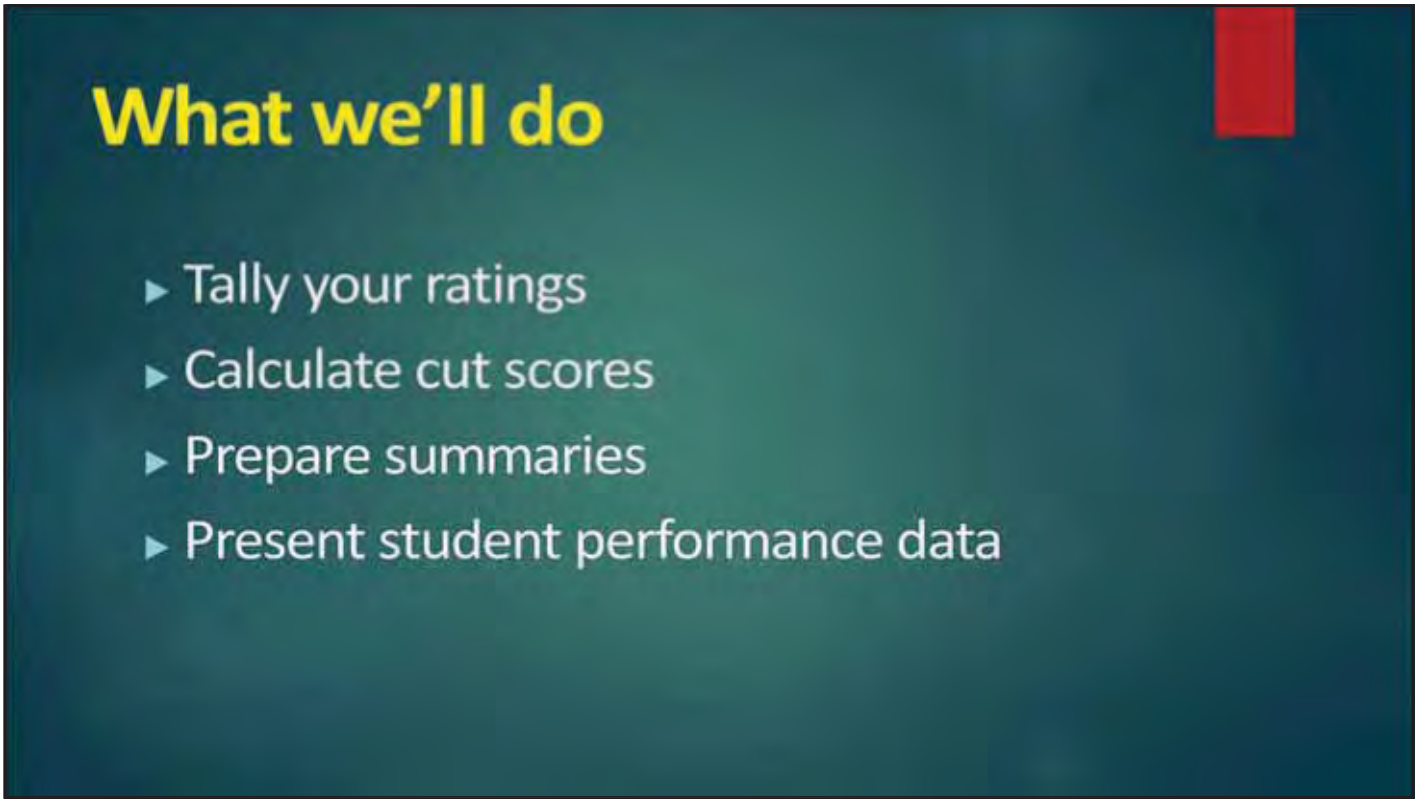
As you consider the feedback from Round 1, pay close attention to

1. How other people in your room rated the work samples you rated – did you pick up any new ideas that might affect your Round 2 ratings?
2. How other people in your room described the way they interpreted the PLDs – did you pick up any new ideas that might affect your Round 2 ratings?
3. The percentages of students who would be at each level based on the Round 1 cut score ranges – did those seem reasonable or not?



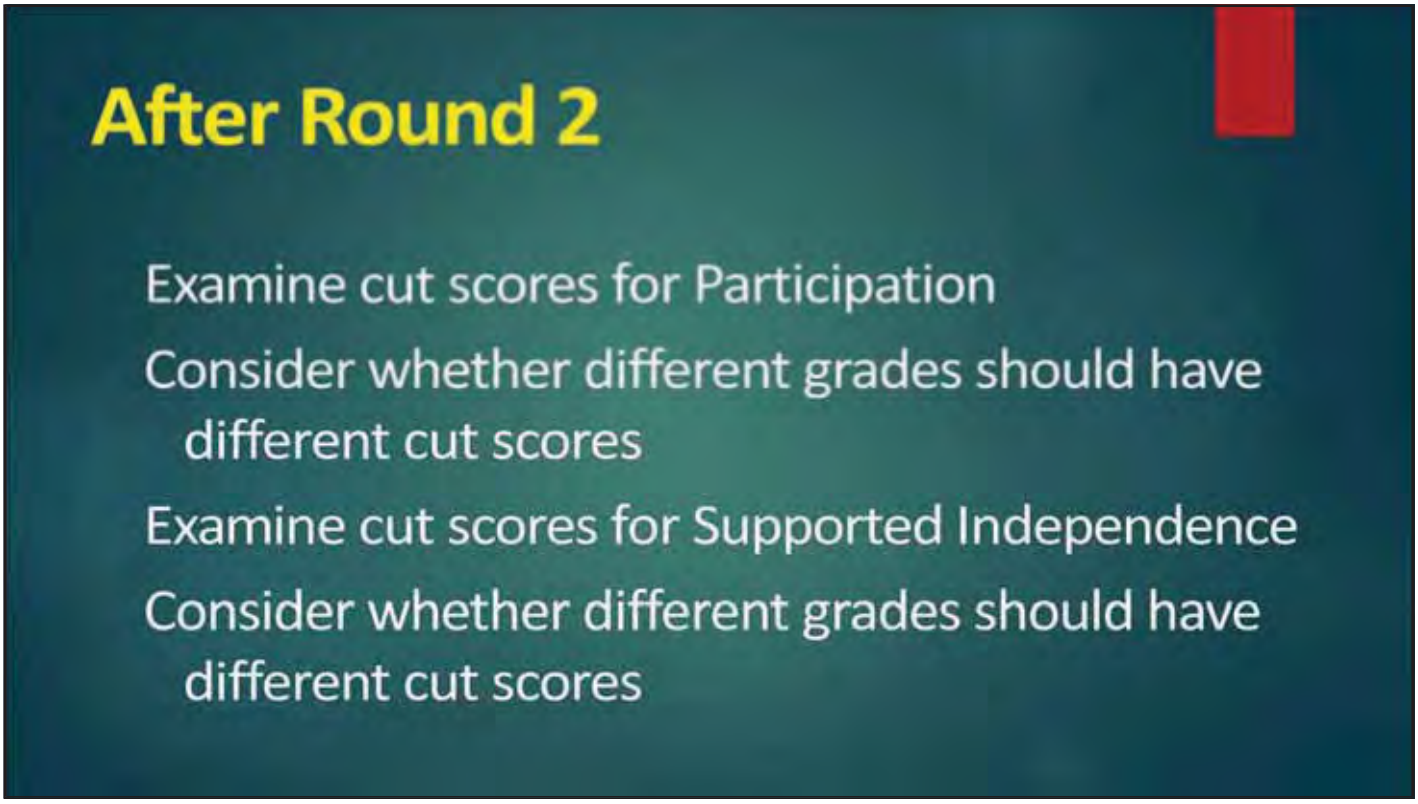
The process you will use in Round 2 is very similar to the process that you will have used in Round 1. The major differences for the second round are that the work samples will now be targeted around each of the Round 1 rough cut scores, and you will receive additional information that shows the impact data. You will still use the PLDs to sort the work samples.

As in Round 1, feel free to consult with others at your table, but please do not talk across tables, as that tends to create confusion.



When you finish Round 2, make sure you have completed Round 2 of your rating form, and turn it in to your facilitator. Again, he or she will make sure you have turned in all the materials you should turn in and give you your next assignment or dismiss you.

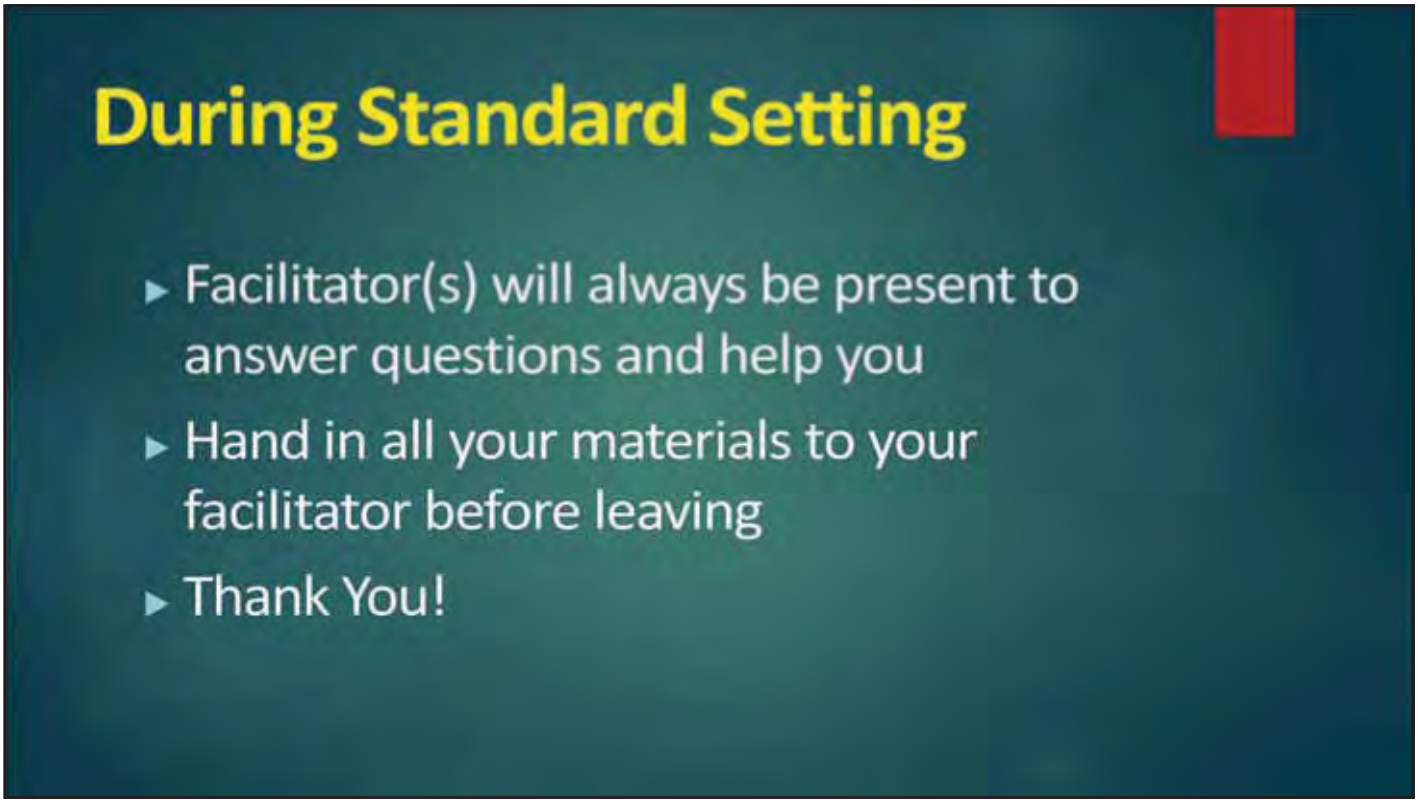
After Round 2, we will once again tally your ratings, calculate cut scores, and create tables and charts for you. We will also prepare tables and charts showing the impact of your Round 2 cut scores. At that point, you have one last opportunity to review your Round 2 decisions before going either to the Cross-Grade Review Panel or to a debriefing panel. Tomorrow, you will learn which of those two groups you will be in.



Most of you will be considering cut scores for grade bands: grades 3-5 or 6-8. You will be setting a single cut score for Attained and a single cut score for Surpassed for the entire grade band. After two rounds of considering these band-wide cut scores, you will have an opportunity to decide if there should be any degree of differentiation from one grade to the next. You do not have to differentiate by grade, but you will have an opportunity to do so if you should so desire. Your facilitator will lead a general discussion of the cut scores and impact data – both globally and by grade level. You can decide as a group how you wish to proceed from there.

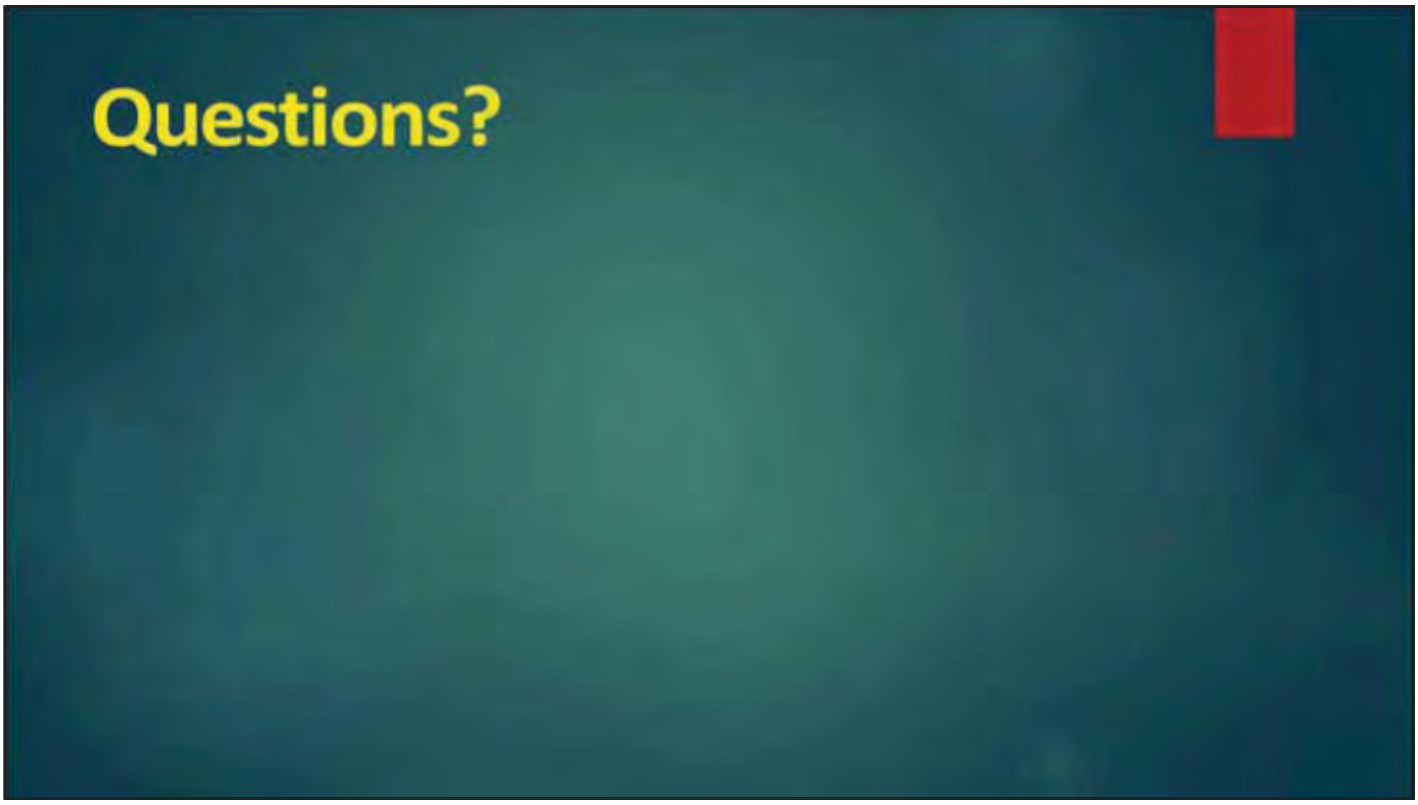
Should you decide to differentiate by grade within band, we will ask you to defend your decision in terms of a specific work sample or small set of work samples. For example, if we set the ELA grades 3-5 cut score for Attained at 24 and you would like to drop that cut score to 23 for third graders or raise it to 25 for fifth graders, you should be able to point to a work sample in your packet and say, "A third grader performing at Level 2 should be able to perform at this level (showing the work sample with a score of 23) but not necessarily at this level (pointing to a work sample with a score of 24)."

We will do this democratically. Someone will make a motion to make that change, and someone else will second it. Given that we will have spent two days or more coming up with the Round 2 cut scores, we will require a 2/3 majority vote to approve the change. Think of it as overriding a veto or amending the constitution.



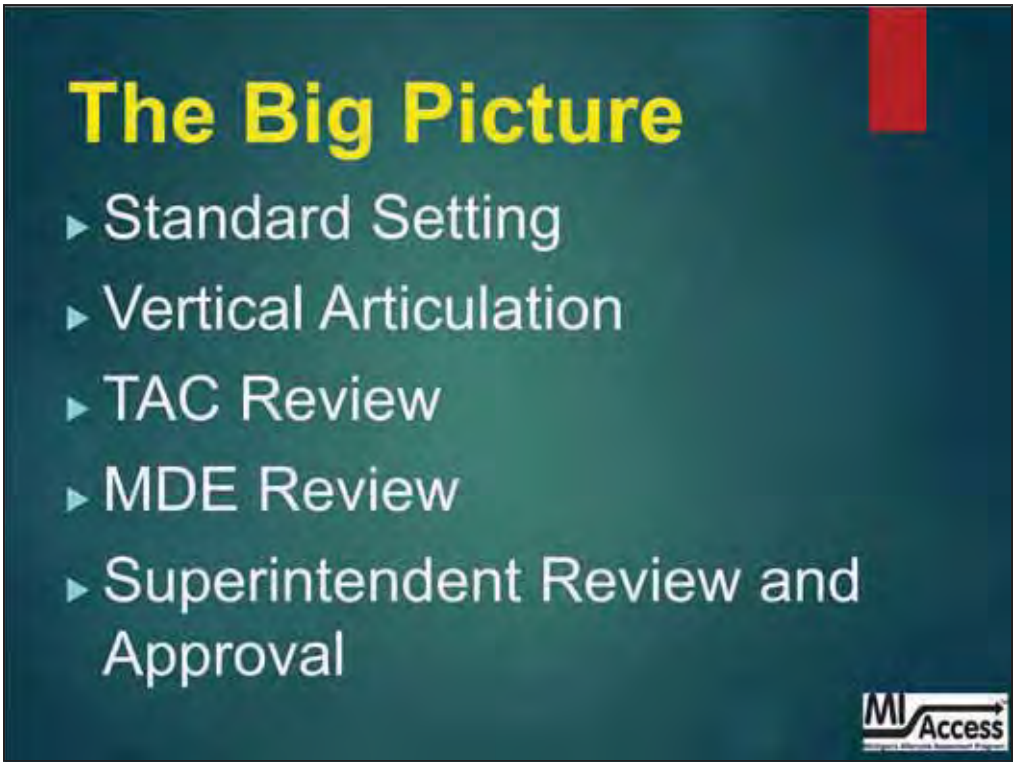
Some final notes….Facilitators and MDE staff will also be on hand to help you and answer your questions. MI staff will help guide you through each of the steps of standard setting. Once you finish any activity, you should make sure to hand in your materials to MI staff that will be in your room.

Most importantly, we hope you enjoy the standard setting process and we greatly appreciate your time and your willingness to participate.



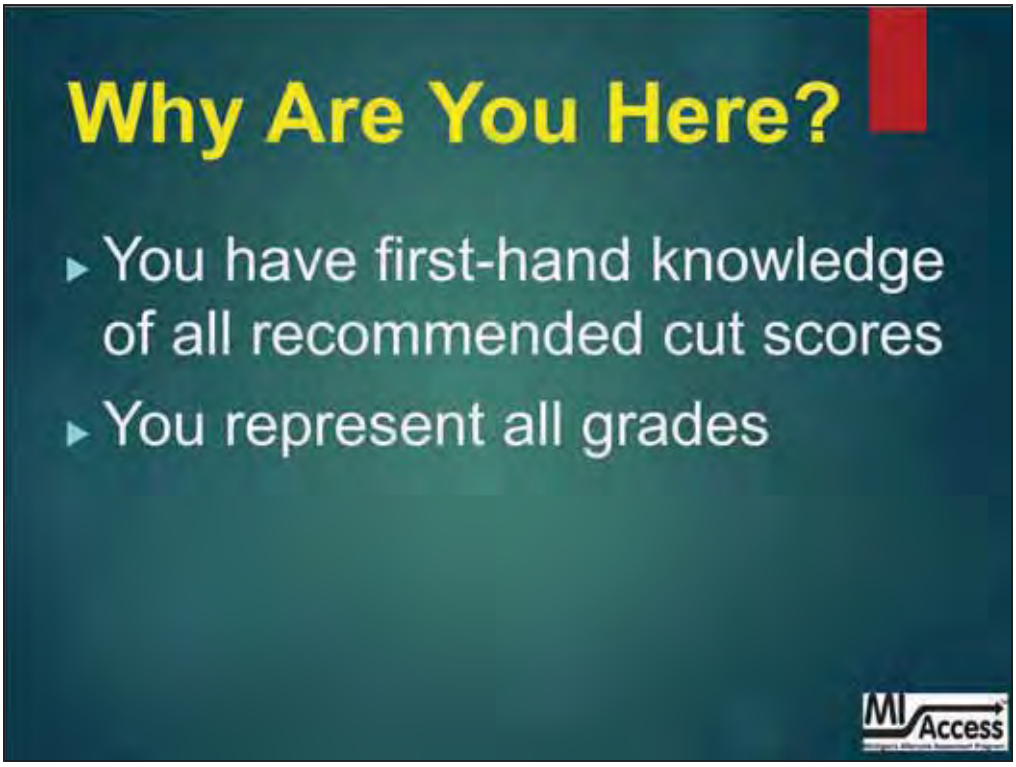
Do you have any questions about any part of the presentation before we tell you about your room assignments?
[Answer any procedural questions. For policy questions, defer to MDE. For subject or grade-band-specific questions, defer to facilitators, who will answer them once they have been dismissed from the large-group session.]

MI-Access
Vertical Articulation
MICHIGAN Department of Education
MI Access
Michigan's Alternate Assessment Program

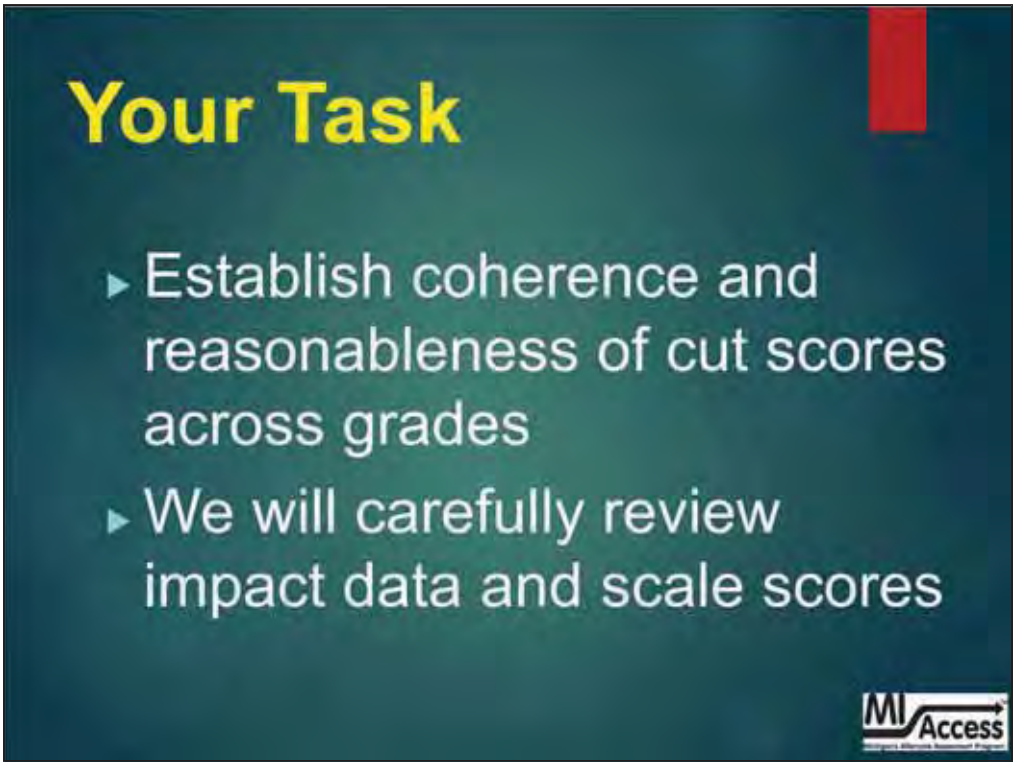


For the past three days, you have been involved in standard setting for a single grade or grade band. That's part of a larger enterprise of setting cut scores for all grades in such a way that when a superintendent or a school board or the general public looks at the full impact across grades, it seems reasonable. Therefore, now that we have recommended cut scores grade by grade, we want to take a more panoramic view of the results and see if we want to make any adjustments. Just as you discussed your individual cut scores with others at your table and then with others in your room and made certain adjustments, we can now expand that conversation to include panelists in other rooms considering other grades.

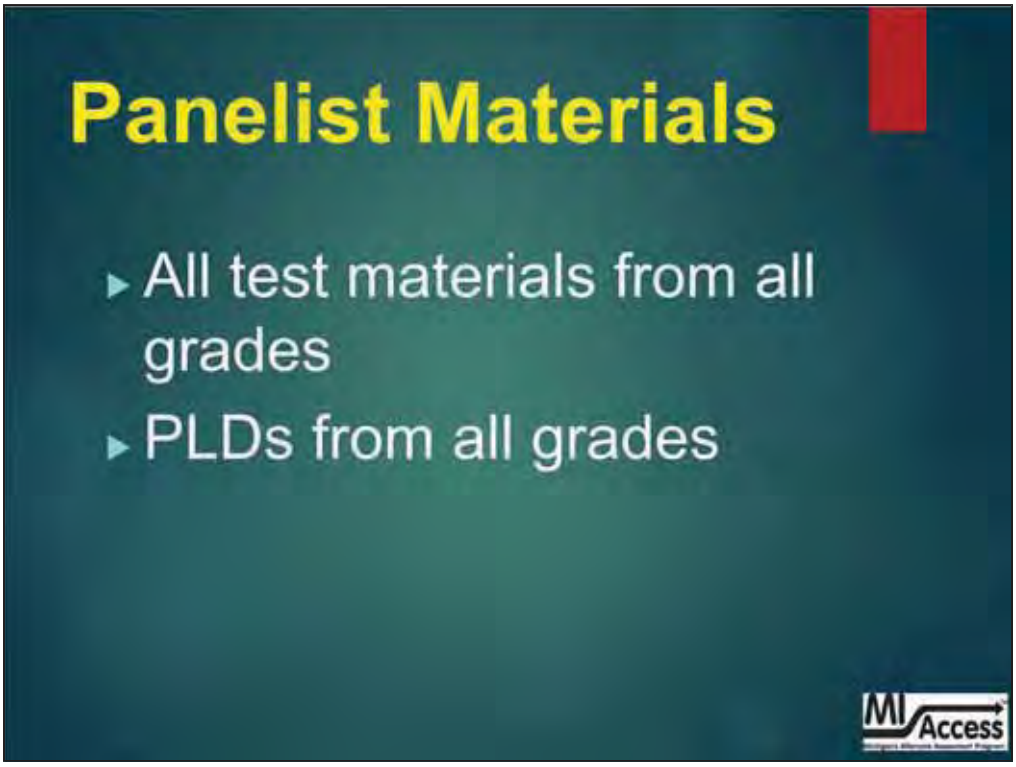
After we finish our work here today, the process still won't be finished. The technical advisory committee will examine what we did here this week to make sure we followed generally accepted practice and will make their recommendation to the Department as to whether or not any further adjustments are in order. The Department will then review all input and may make additional adjustments before making final recommendations to the State Board of Education. The Board, as I have mentioned previously this week, has the final authority to accept, modify, or reject the cut scores recommended to them.



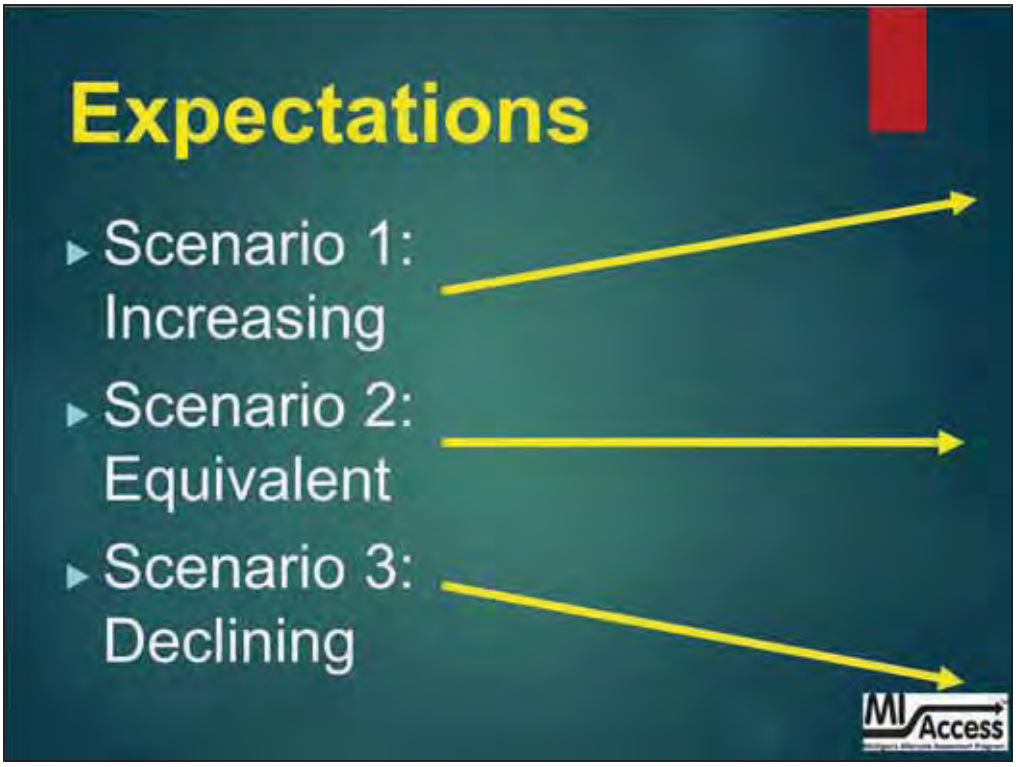
You were selected because you know the students and the tests. You also know what went on in the various breakout rooms over the past three days. We need your advice as we look over all the recommendations to see if there are any conflicts we need to resolve. I will explain momentarily what I mean by "conflict." the bottom line is that you are here because you are the most knowledgeable people in the state with regard to these students, these tests, and these recommendations.



In a few minutes, we are going to look at a chart showing all cut scores and their impacts on all grades. Your task is to tell us whether the results we have obtained so far are generally reasonable when we consider all grades or if we should make some changes.



During this activity, you will have access to all the materials you used over the past three days plus all the materials the other panels used.

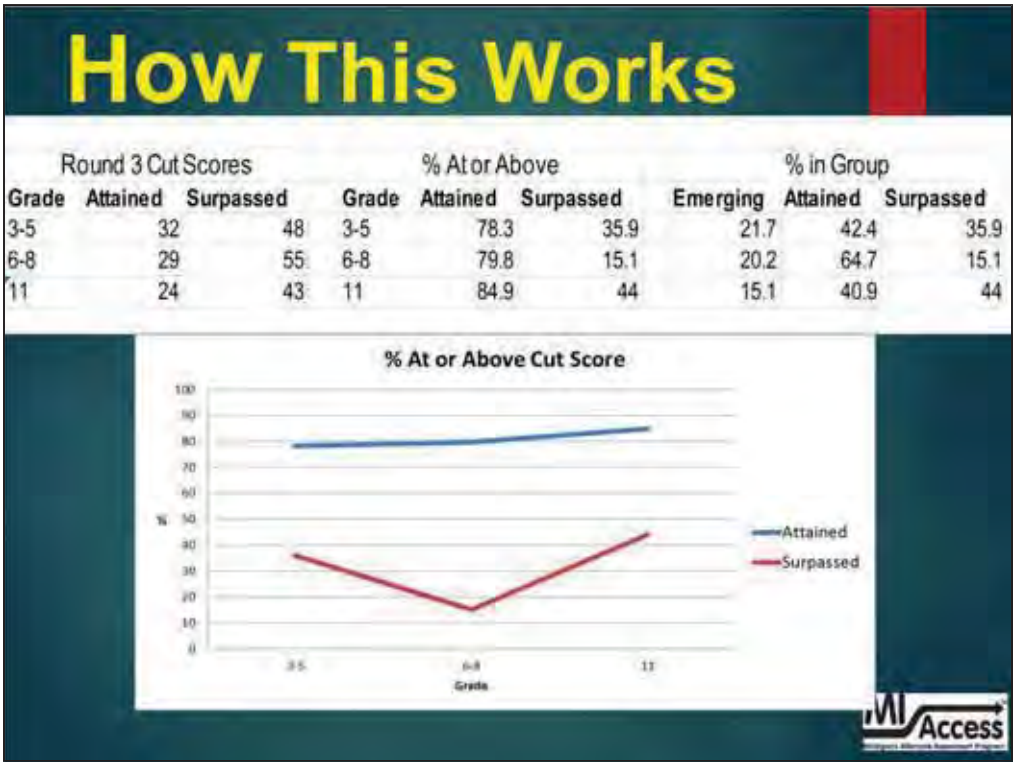


When I speak of the reasonableness of cut scores and their associated impacts, I am talking about expectations. Think about how schools and systems perform over time. There are typically three scenarios in terms of students at or above some cut score:

1. Student performance increases over time – each year, students do a little better than they did the year before; in any given year, a larger percentage of 8th graders attain the standards than they did as 7h graders the year before.
2. Student performance is fairly stable over time – each year looks pretty much like the year before, and within any given year, 8th graders, 7th graders, 6th graders and so on all achieve at about the same level.
3. Then there's generally declining performance – each year, students perform a little less well than they did the year before; 8th graders don't do quite as well as 7th graders, and 7th graders don't do quite as well as 6th graders.

There are variations on these three scenarios; for example, generally increasing except for 6th grade, which in our district is the first year of middle school, and everyone knows what happens then. Then there's generally declining except for 8th grade where we have our reading enrichment program. These tend to be school- or district-specific phenomena, and we are going to be looking at trends for over half the country.

What we don't expect is something like this – where the percentage of 4th graders at Level 2, for example, is much higher than that of 3rd graders. But then at 5th grade, they drop off again, only to rise at 6th and 7th grades, drop off again at 8th grade and high school. How would you explain this to parents? "Your 4th grader did quite well this year, but the likelihood that he or she will do well again next year is not so good."



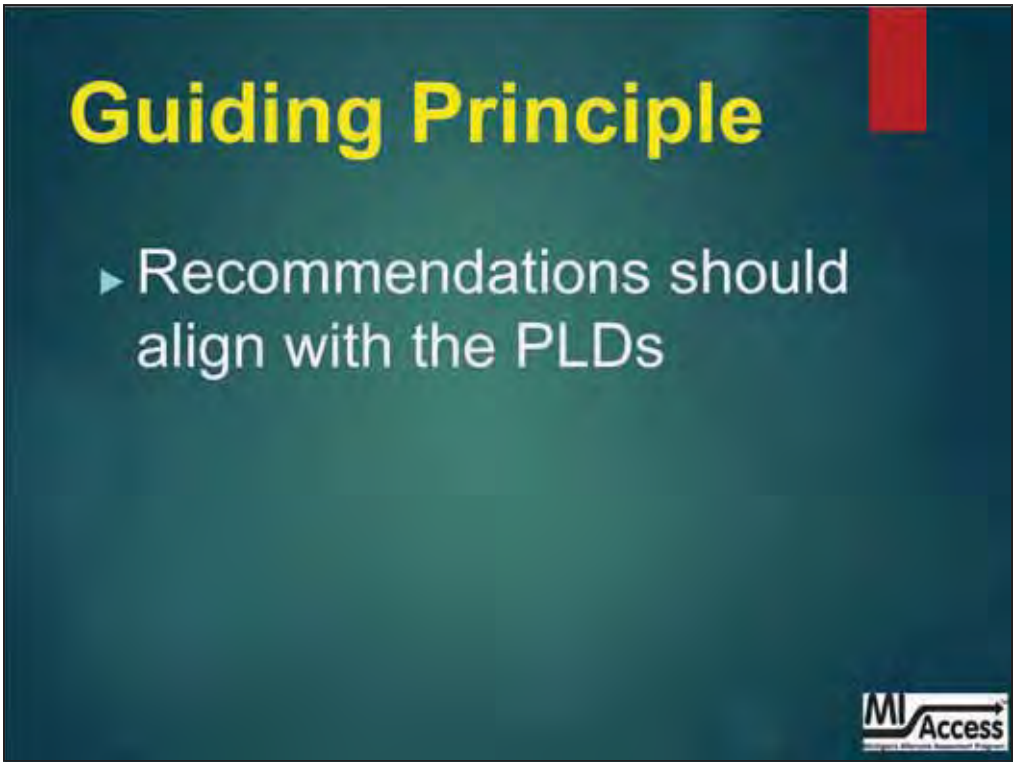
When we break into two groups – one for ELA and one for Math – you will see a set of tables and a graph like this. The first table will show all the final cut scores expressed in scale scores. The second table will show the percentages of students at or above Level 2 and Level 1, by grade, given the cut scores in the first table. This same information is also depicted in the down below. The third table will show the percentages of students in each of the three levels, based on the cut scores in the first table.

As we consider the information in these tables and this graph, we will be able to recommend a change in any cut score. When we change any cut score in the first table, the other tables and the graph automatically update. Let's work through an example.

The graph you see here generally conforms to the second scenario I mentioned earlier – generally flat. All grades show about 50% of students scoring at or above Level 2, except for 7th grade. Why are the results for grade 7 so different from all the others? Notice that in the first table, the Level 2 cut score is 3050. If we look at grades 5, 6, and 8, we see a much larger jump from grade 6 to grade 7 than for grade 5 to grade 6 and a very small jump from grade 7 to grade 8. Is there something about 7th graders, the test, the items that would make these differences seem reasonable?

In this situation, we would go back to the [tests/work samples – depending on procedure] and let everyone take a look. At some point, someone may suggest lowering the cut score for grade 7 to make the progression from grade 5 to grade 6 to grade 7 to grade 8 more even. Eventually, someone suggests moving the Grade 7 Level 2 cut score to 2965. Here's what happens…

Here, we have changed the Grade 7 Level 2 cut score from 3050 to 2965. You will notice that the percentages in the second and third tables have changed, and that the dip in the Level 2 graph has been reduced. It is not entirely gone, however. That was not really our objective. Our objective was to reset the cut score where it would seem more reasonable.

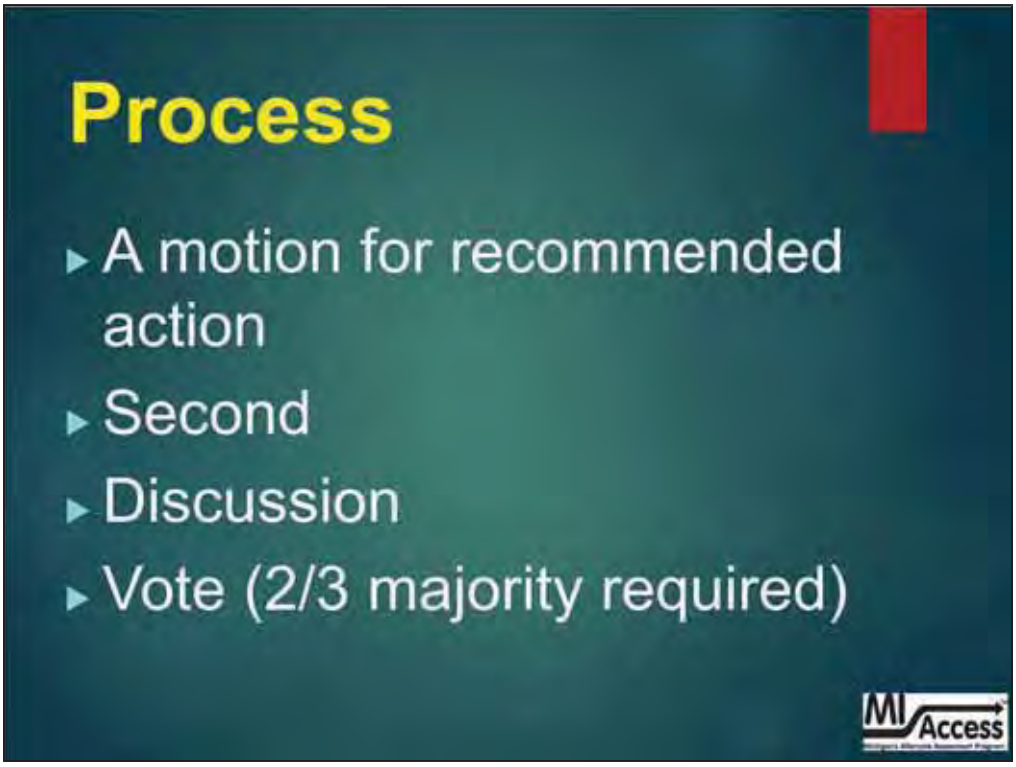


The scale scores for these tests are set up so that there is a general increase from grade 3 to grade 4, and so on up through grade 11. If the scaled cut score for Level 2 for grade 6 is higher than the cut score for Level 2 for grade 7, we may want to look into that. If the percentages of students at Level 2 and above varies significantly from grade to grade, we may also want to look into that. I'm not saying that neither of these two things should ever happen; I'm just saying they would be unexpected. We would first want to find out if the departures from expectation are justified or can be explained. If we can't explain them, then we consider changing something.

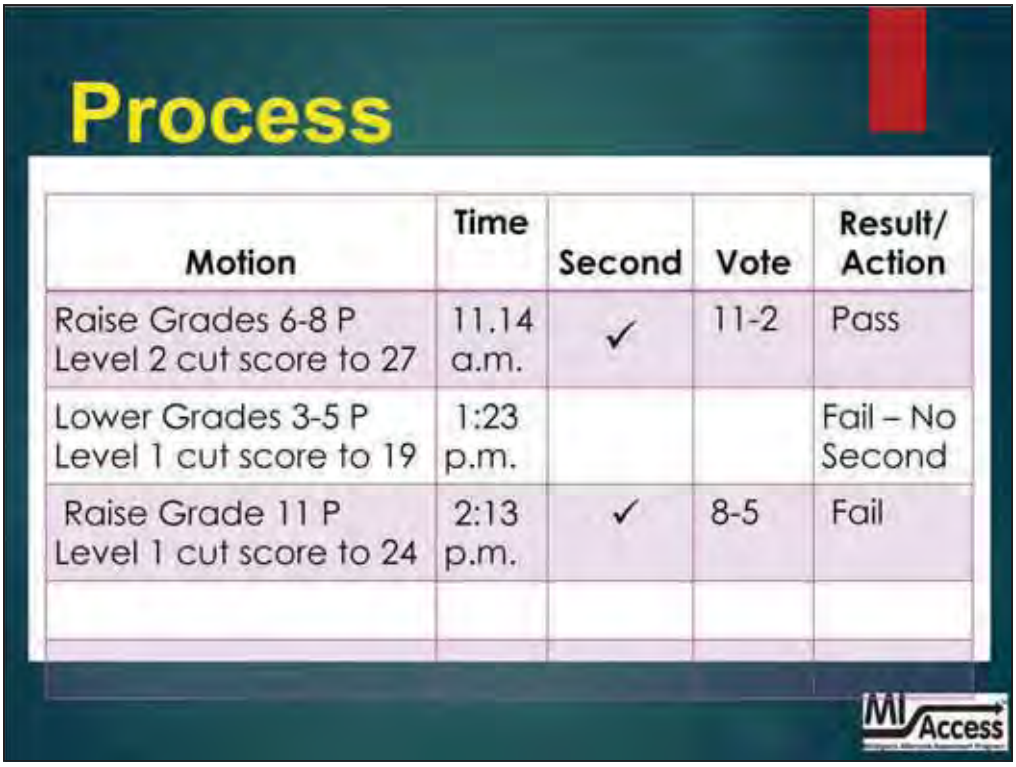
However, we don't want to move cut scores just to make the lines smooth. Every recommendation for a cut score change should be grounded in the PLDs

BoW Version: In changing a cut score, we are saying that a work sample that we previously thought was in one level is now in another level. Can you look at this work sample and the PLD and honestly say that this work sample belongs in this level? If the answer is Yes, then the change is justified. If the answer is No, the change is not justified. We must either accept the original cut score or find another substitute.

Bookmark Version: To change a cut score, we have to move our collective bookmark. Before accepting the change, we need to look at the page that most closely aligns with the new recommended cut score and ask ourselves if the item on the preceding page (remember what the bookmark signifies) aligns with the PLD for that level. If the answer is Yes, then the change is justified. If the answer is No, the change is not justified. We must either accept the original cut score or find another substitute.



We're going to do this democratically. To change any cut score, we will need a motion and a second. After the second, there will be discussion, not before. At some point, Craig or I will call for the vote. Since we are essentially overriding someone else's decision, we will need a 2/3 majority to pass the motion. We will have someone in each room recording the vote count and noting whether or not the motion passed. At the end, we will ask for a motion to accept all cut scores, changed plus remaining unchanged, and follow the same rules.



Here's an example.

The first motion passed by a vote of 11 to 2.

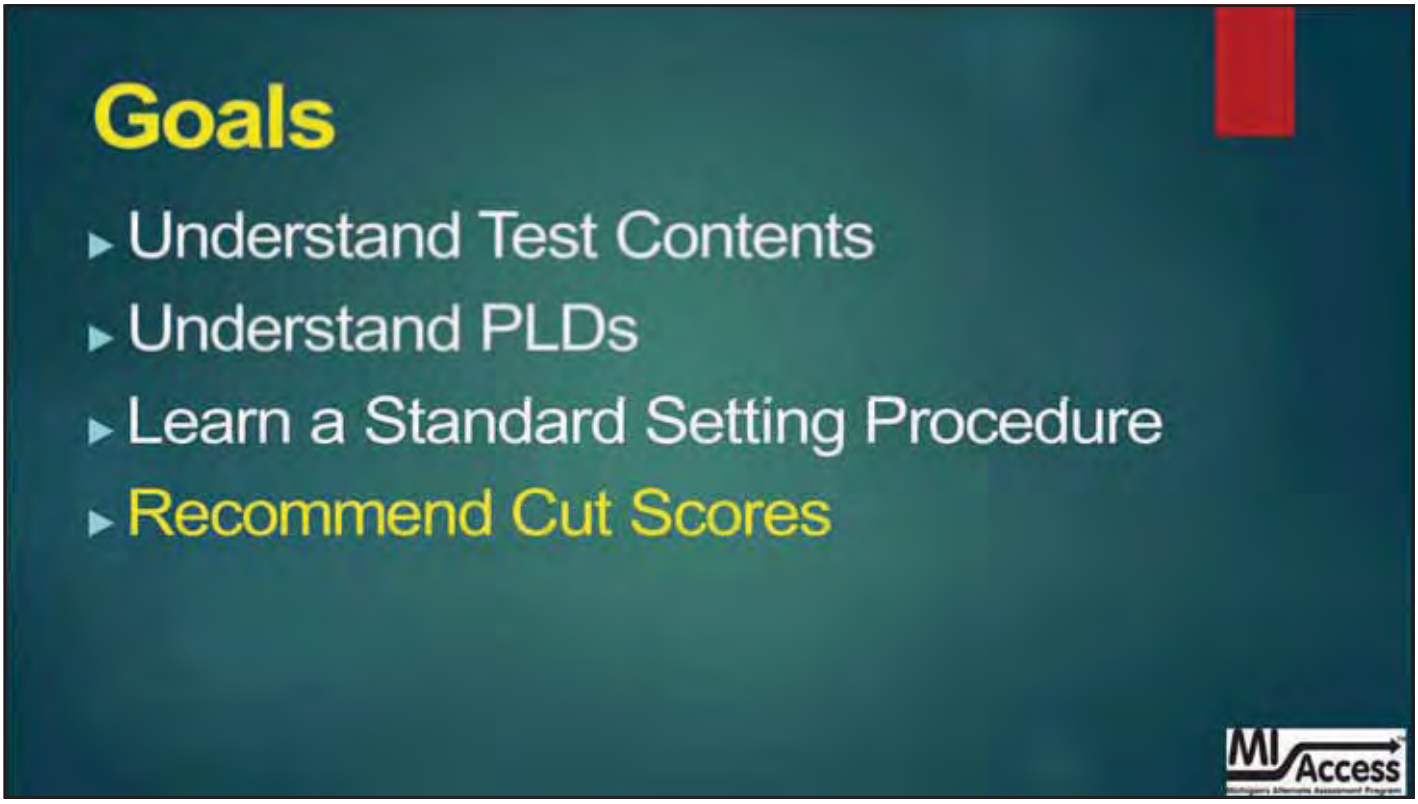
The second motion failed because a majority voted against it.

The third motion also failed. Even though a majority voted for it, it did not get a 2/3 majority.

We're going to do this democratically. To change any cut score, we will need a motion and a second. After the second, there will be discussion, not before. At some point, Craig or I will call for the vote. Since we are essentially overriding someone else's decision, we will need a 2/3 majority to pass the motion. We will have someone in each room recording the vote count and noting whether or not the motion passed. At the end, we will ask for a motion to accept all cut scores, changed plus remaining unchanged, and follow the same rules.

MI-Access
Standard Setting

June 29 – July 2, 2015



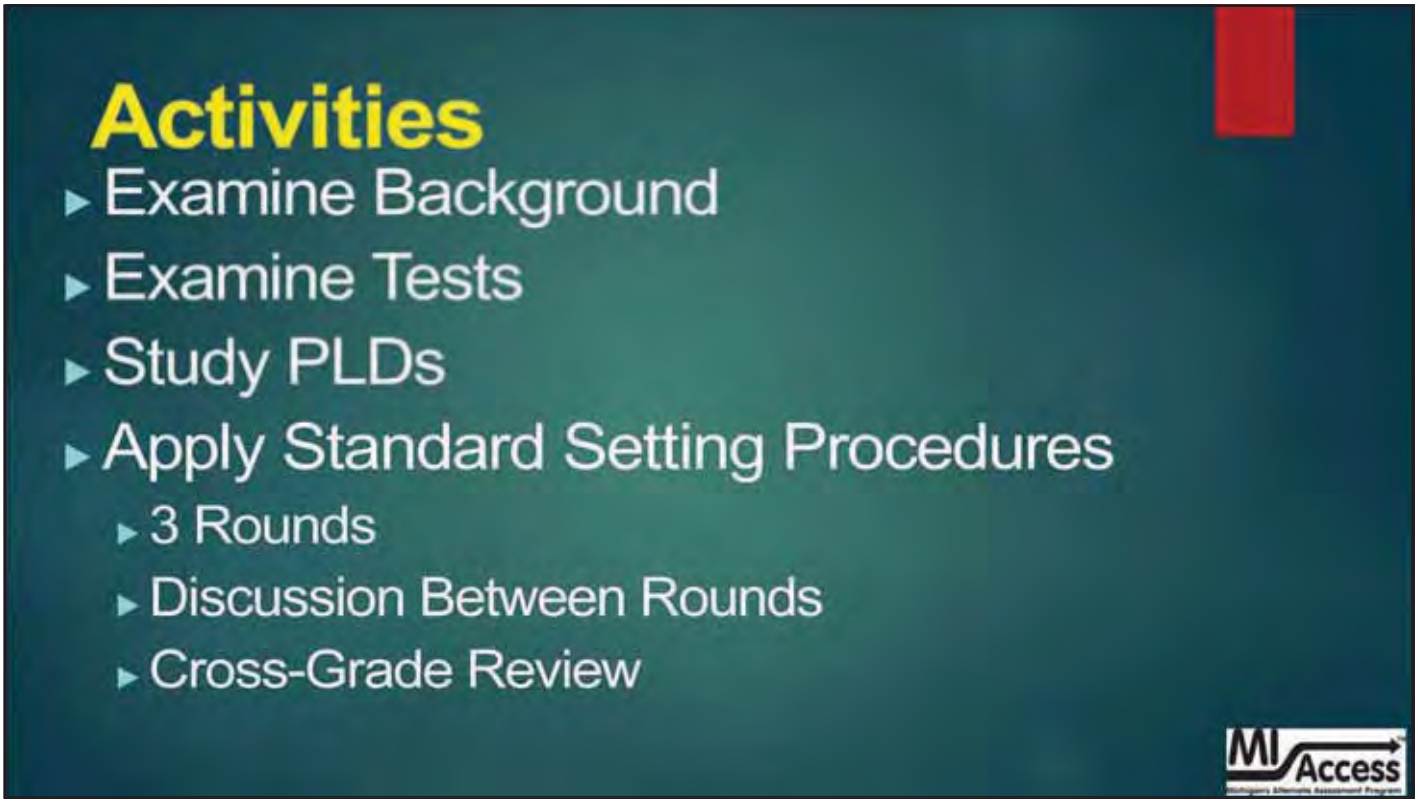
We are here to consider the tests students took this spring under MI-Access: Functional Independence; and to recommend performance standards – cut scores – to the State Superintendent. With that in mind, here are our goals for the next four days:

Understand Test Contents – In order to recommend meaningfully what scores a student should earn on a given test in order to be considered Partially Proficient, Proficient, or Advanced, you should be very familiar with the contents of those tests, starting with the state content standards and ending with the individual items and their scoring rubrics.

Understand PLDs – We want you to be very familiar with the Performance Level Descriptors (PLDs) that describe what students at each performance level know and can do. Your recommended cut scores will translate those descriptions into numerical goals for students.

Learn a Standard Setting Procedure – You will be using a process known as the Bookmark Procedure.

Recommend Cut scores – When all is said and done, the main thing you do this week will be recommend two cut scores for each test, one to separate Emerging from Attained performance and one to separate Attained from Surpassed performance. Everything else you will do will be primarily to prepare you to meet this specific goal.



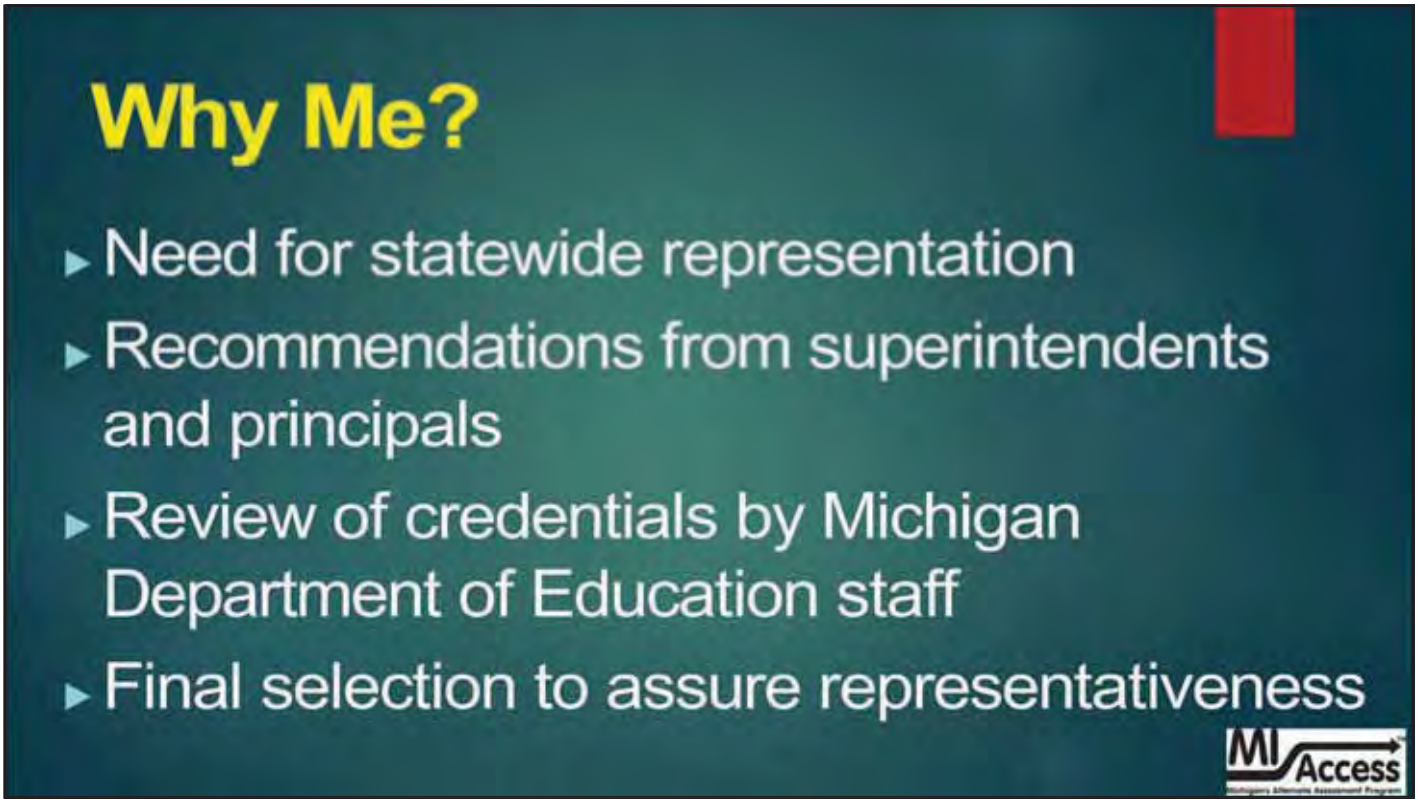
To accomplish these goals, we have developed a series of activities that will lead to the development of defensible cut scores. If you would take out your agenda, we will look at the activities we have planned for the next four days.

Later this morning, you will examine the tests in some detail. Afterwards, you are going to study the Performance Level Descriptors or PLDs in detail. Every recommendation we make this week must be firmly grounded in the Performance Level Descriptors. When we submit your recommendations to the Superintendent and when he acts on them, every action needs to be based on the PLDs because at some point, these performance standards will be reviewed by outside agencies, and the first question they will ask is whether or not we set our cut scores on the basis of clearly worded PLDs.

This afternoon, you are going to learn a specific way to use the information you have to make cut score recommendations. Let me state now, and we will no doubt remind you frequently later, that your job is to recommend cut scores. The Superintendent will make the final decisions about cut scores, based on your recommendations and other considerations.

After you have learned the standard-setting procedure and had a chance to practice, each of you will then examine the test for your subject/grade band and recommend cut scores. We will tally results for your panel and share the results with you, after which time, you will discuss those results in your panel and do the same thing again. Between rounds of examining the tests, we will give you additional information to consider.

At the end of the week, some of you will participate in a cross-grade review. You will examine recommended cut scores for all grades or grade bands and consider whether some of them ought to be adjusted so that the overall impact of the cut scores would seem reasonable to parents, teachers, school administrators, and the general public. We will provide specific instructions on how that will work. The rest of you will participate in a debriefing about the process we are using this week. We will use the feedback you give us in our presentation to the Superintendent and to help us improve the process for future standard-setting activities.



Having heard all this, you may be wondering how you happened to be chosen for this singular honor.

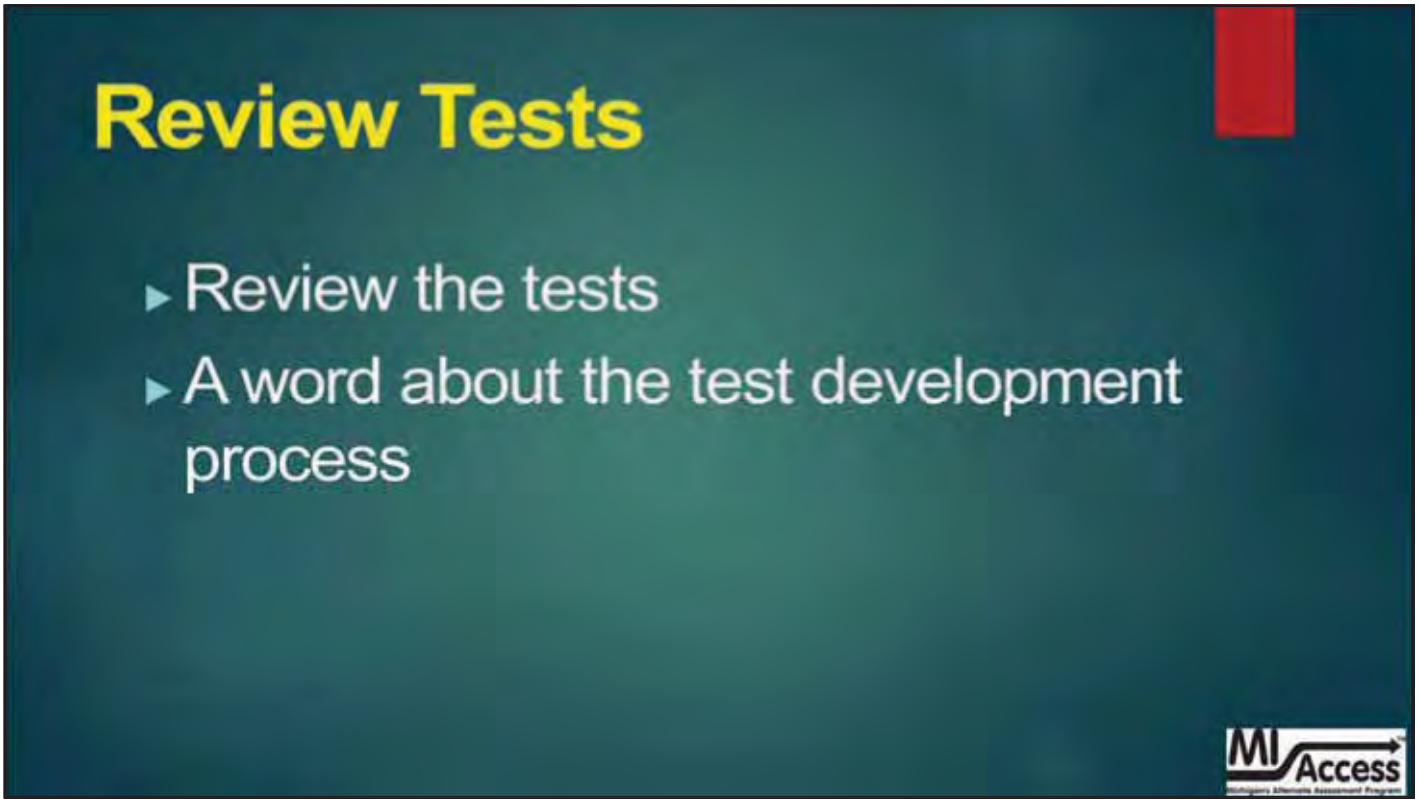
Standard setting is a high-profile activity, and we want as many people as possible involved in it. More importantly, we want those people to be representative of the State of Michigan as a whole. We have sent invitations to all parts of the state in order to find panelists who could fairly represent the state in terms of gender, ethnicity, length of service, and type of student population served. Staff of the Michigan Department of Education reviewed credentials of many people and chose you as the most representative and best qualified to carry out this important task. They put a lot of thought into selecting you, and we trust that you will put a lot of thought into what you do here this week. The performance standards we recommend this week, once approved or modified by the Superintendent, will be applied to all MI-Access Participation and Supported Independence students in Michigan not only this year but for years to come.

This may be a good time to provide a little background about standard setting and clarify some terms.
First, there are many different kinds of standards. Many people, when we refer to standards, automatically think of content standards. Actually, we couldn't set cut scores unless we had content standards that tell us what we are trying to teach and therefore what we are trying to test. But performance standards are numerical standards that specify how much we expect students to learn.

Part of the process of establishing performance standards is establishing performance levels. In Michigan, for the MI-Access test we have three:
3 – Emerging toward the standard
2 – Attained the standard
1 – Surpassed the standard

Each level has its own detailed description of what students at that level know and are able to do. No Child Left Behind requires states to spell out these levels with Performance Level Descriptors – PLDs. The law requires at least three levels; MEAP has four and MI-Access has three. Some states have even more. Groups of Michigan educators have drafted these descriptors, Michigan Department of Education staff have worked with those groups to refine and polish them, and the Superintendent ultimately has the responsibility for implementing them.

Again, the three performance levels are
1 – Surpassed the standard
2 – Attained the standard
3 – Emerging toward the standard



As I mentioned earlier, you will review the tests before you start recommending cut scores. A note about the tests: These tests were developed over a period of 2-3 years and have had considerable input already from groups of Michigan educators, for both content and fairness/sensitivity issues. You may not be thrilled with each and every item or how it is scored. That's OK. This is not the time to critique the tests because these are the tests we gave to students this year, and these are the tests from which they will receive their scores. If you see something you think is not as it should be, we would certainly like to know about it.

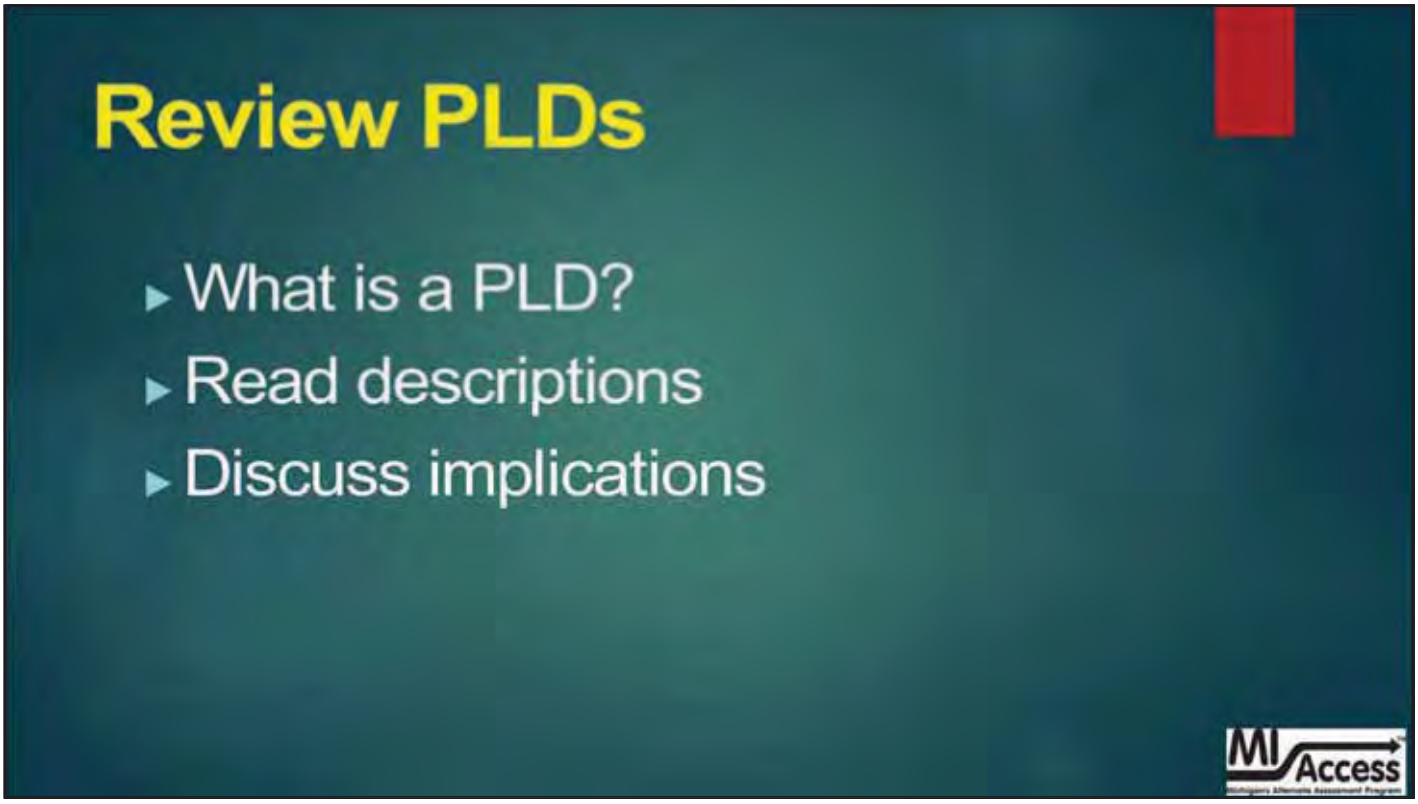
Please share your comments with us at the breaks or at other times, but we will not be conducting a test or item review during this workshop. It is appropriate, however, to keep those concerns in mind as you recommend cut scores. For example, if you see something that you believe would hinder an otherwise proficient student from answering an item correctly, you may take that into consideration when you make your cut score recommendation.

After you review the tests and receive instruction in the standard-setting procedure, you will be ready to start recommending cut scores. As you look at each test item , you may find something that you disagree with—again please feel free to mention it to one of us, and take that into consideration when you make your cut score recommendation.

After you review the tests, we will conduct a brief discussion about your reactions to them. Primarily, we will be interested in what you think it takes to answer particular items

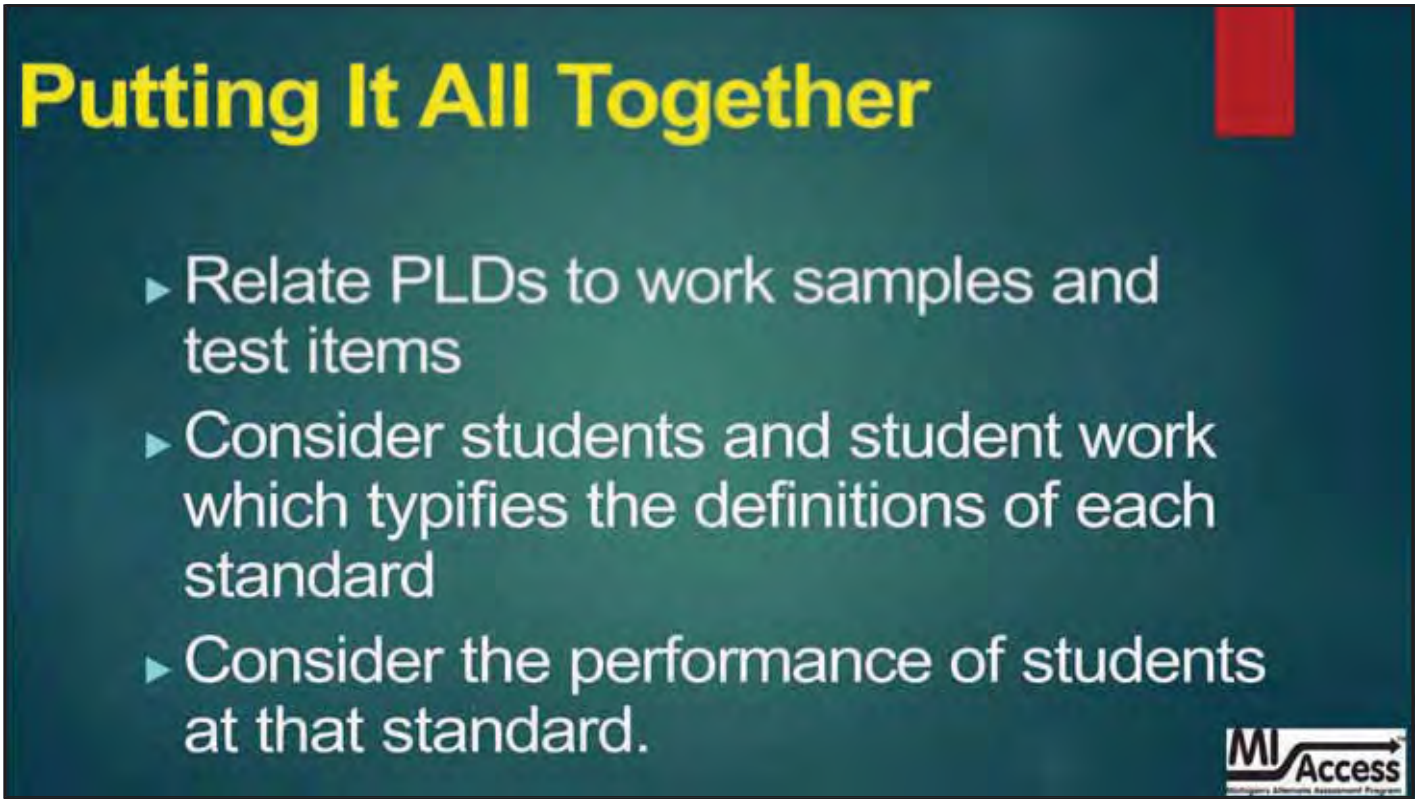
correctly or to receive high scores on them. Our chief aim is to find out what you thought was particularly easy or difficult and what you think would be particularly easy or difficult for Michigan students.

The purpose of this exercise is to have you become very aware of the content of the tests. These tests have gone through years of development. The items have been written, reviewed, field tested, and approved by several committees of Michigan educators. Your responsibility is to help set standards, not to criticize the tests. These exams may not be perfect but they are very good and contain content relevant to the Michigan standards and curriculum.

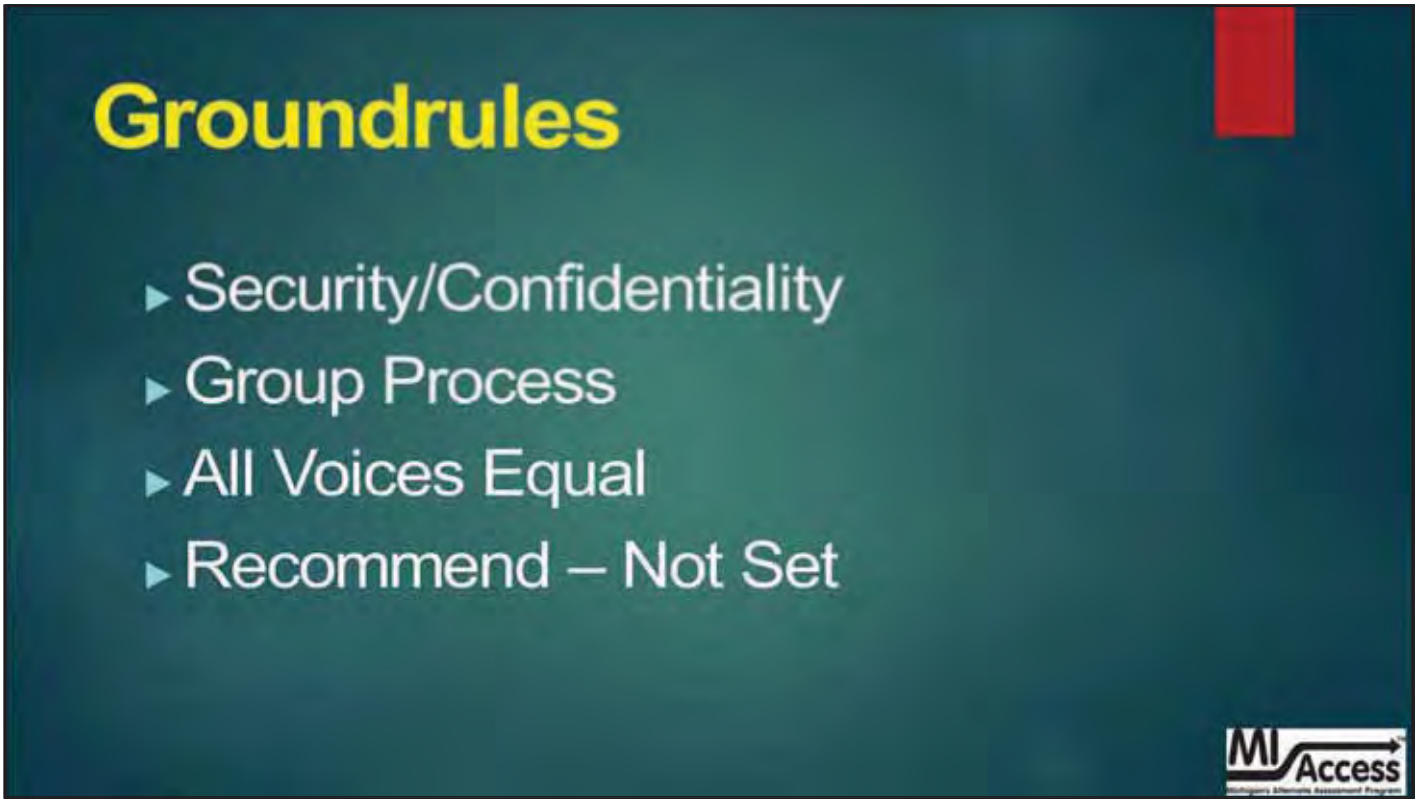


Right now, each of you probably has some idea what constitutes Emerging, Attained, and Surpassed. We also want to have a very detailed discussion of the PLDs. Again, a PLD is simply a description of the types of things that students at a particular level know and are able to do. The PLD for Surpassed (Level 1) describes things that these students can do that Attained (Level 2) students cannot do. Likewise, the PLD for Attained describes things that Level 2 students can do that Emerging (Level 3) students cannot do.

Later this morning, you will break into your separate groups, or panels, to study the PLDs for your grade band and subject. Read each description very carefully. Consider what it means to be at the Emerging, Attained, or Surpassed levels. Try to imagine students you have known who would fit the descriptions you are studying. For this standard-setting procedure, we will be paying close attention to a thin slice of students just barely at the Attained level and another thin slice of students just barely at the Surpassed level. Discuss those students with one another, and try to get a sense of the range of achievement within each level. Keep in mind that the PLDs are geared specifically to the state content standards that were used to create these tests. There are other aspects of performance that are not addressed here because they are not directly relevant to these content standards or to MI-Access.

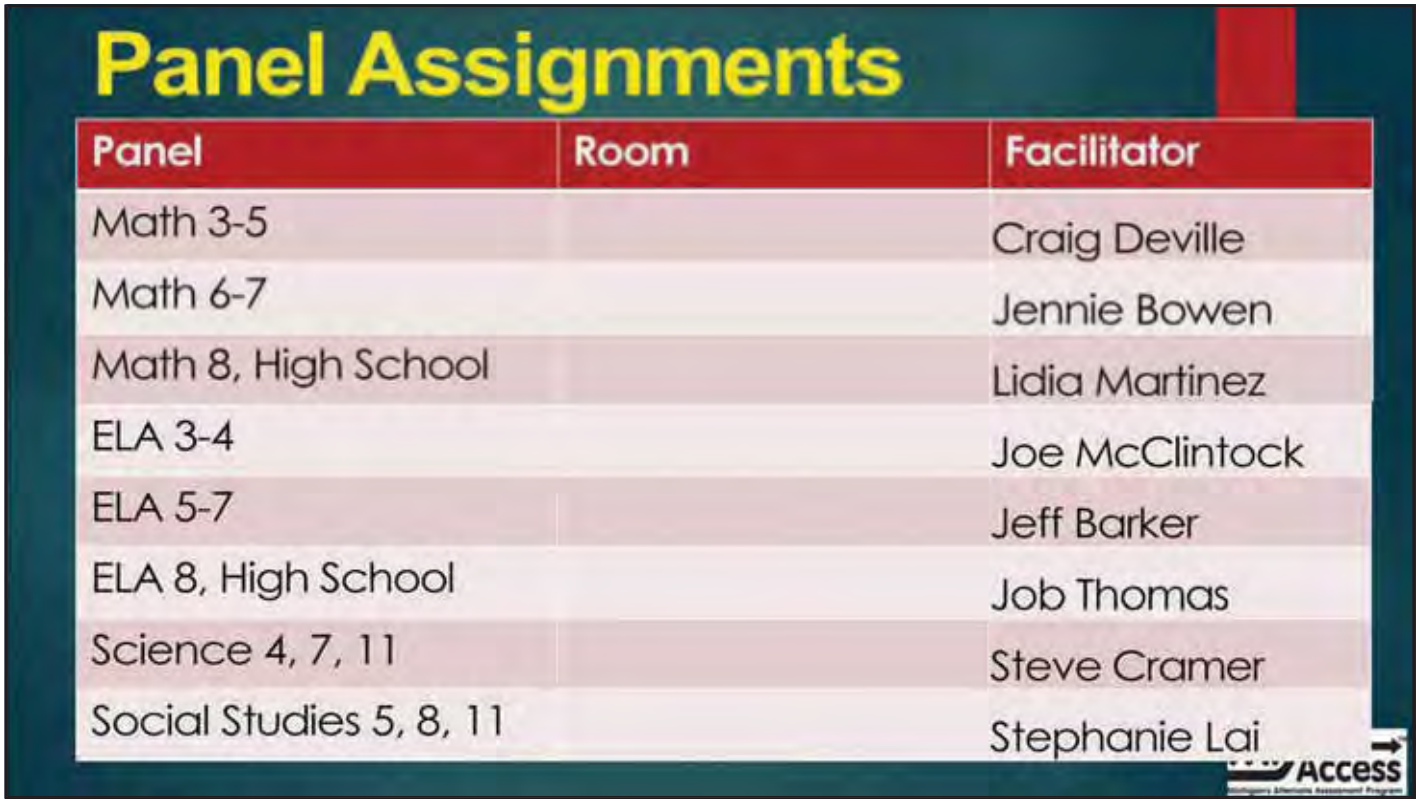


So here's what you're going to do, once we get today's preliminaries out of the way:
After you have learned about the Bookmark procedure, you will examine items in a booklet that has its items ordered from easiest to hardest. Your job will be to find the place in that booklet where the items become too hard for a student just barely at the Attained level to answer correctly and then find the place where the items become too hard for a student just barely at the Surpassed level to answer correctly.



Now, let me familiarize you with the groundrules for standard setting.
You will be seeing actual test items and actual student responses. We are going to ask you to sign a security/confidentiality agreement stating that you will not reveal any of these test items or any student response you might see while you are here. We will sign materials out to you in sequence and account for them by that same sequence when we take them back. We will identify those things that you can share with others. Anything else – you should consider secure and confidential.
You may have already gathered that this will be a group process. There will be activities that you will do completely alone, but we will have a lot of discussion. The purpose of the discussion is to allow everyone a chance to contribute and for each person to develop a greater understanding of the PLDs and how to interpret them, the standard setting process, and the rationales that were used in coming up with standard setting judgments.
We will encourage everyone to speak up during group discussions and will try to keep any one person from overshadowing others. In each round, we will ask each of you to cast, in effect, a secret ballot of your cut score recommendations, which we will tally. In the end, we will take the median of all your group cut score recommendations and report that as the group cut score recommendations. This will give each of you an absolutely equal voice in the final recommendations of cut scores.
Again – I am using the term "recommendation." Although the process is called standard setting, it is really standard recommending. We will work hard this week and employ a proven procedure that yields defensible cut scores. We will then ask the Department and ultimately the Superintendent to consider not just our recommendations but the manner in which we arrived at them. In the end, we recommend, and the Superintendent sets the standards.

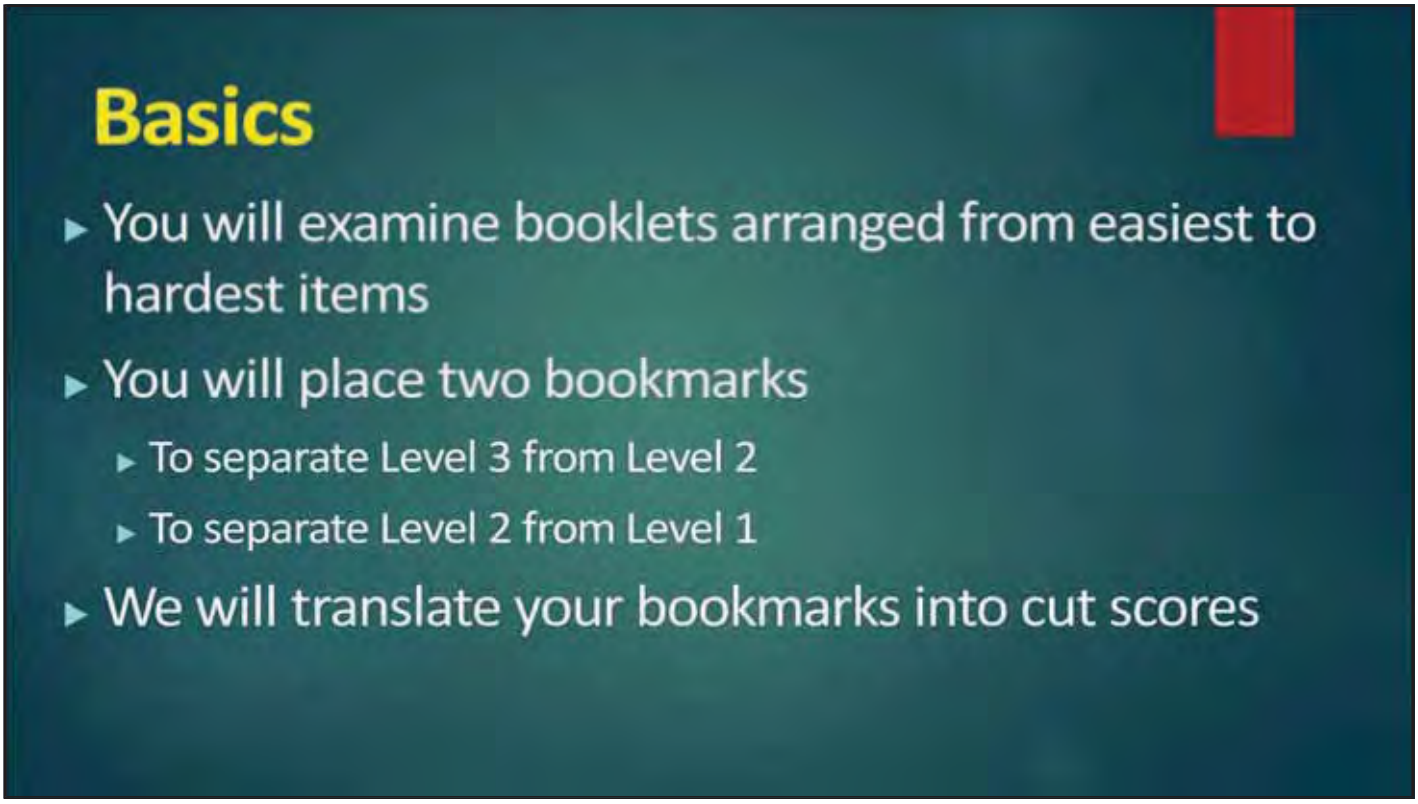
Questions?



Introduce facilitators and dismiss by room. Room assignment will be on name badge.

MI-Access Bookmark Standard Setting Procedure

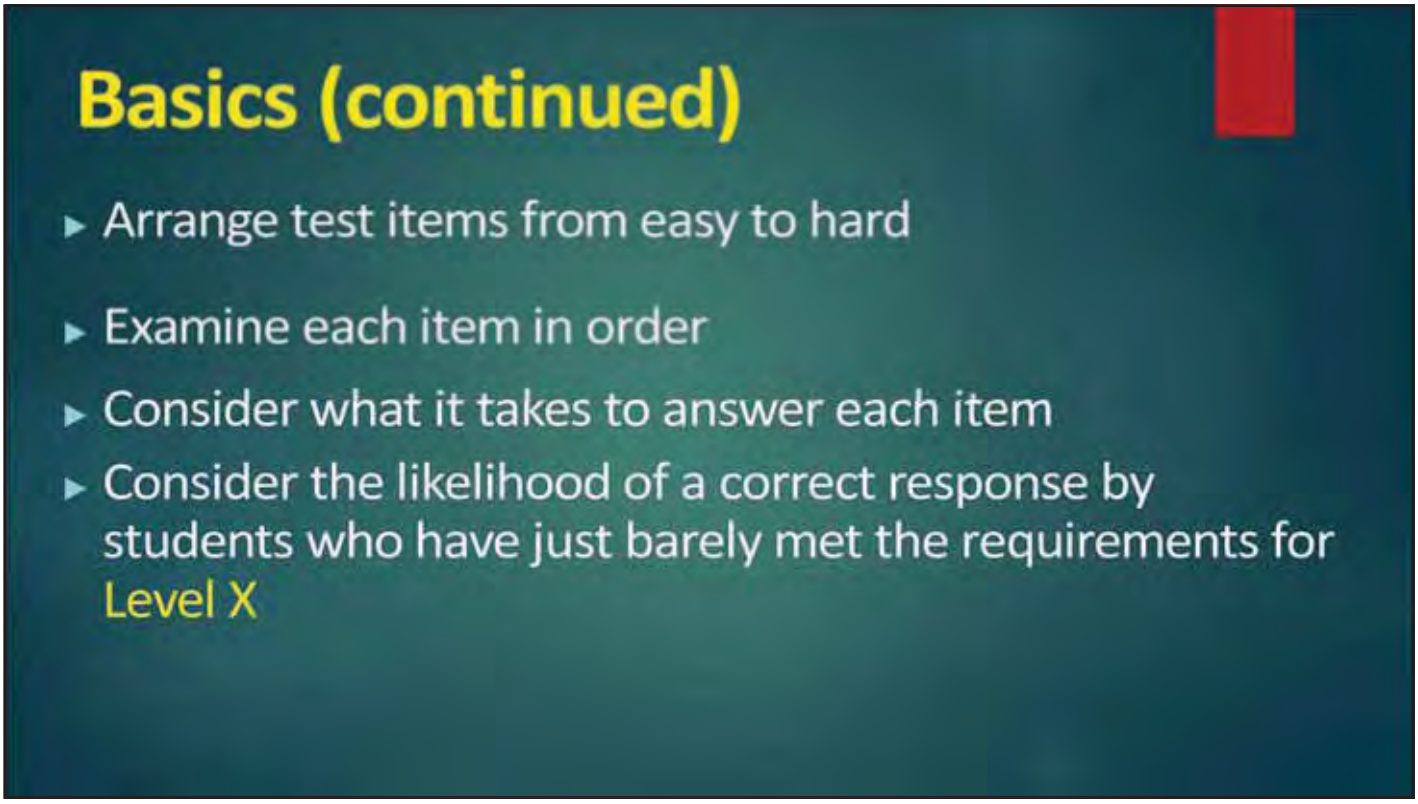
June 29 – July 2, 2015



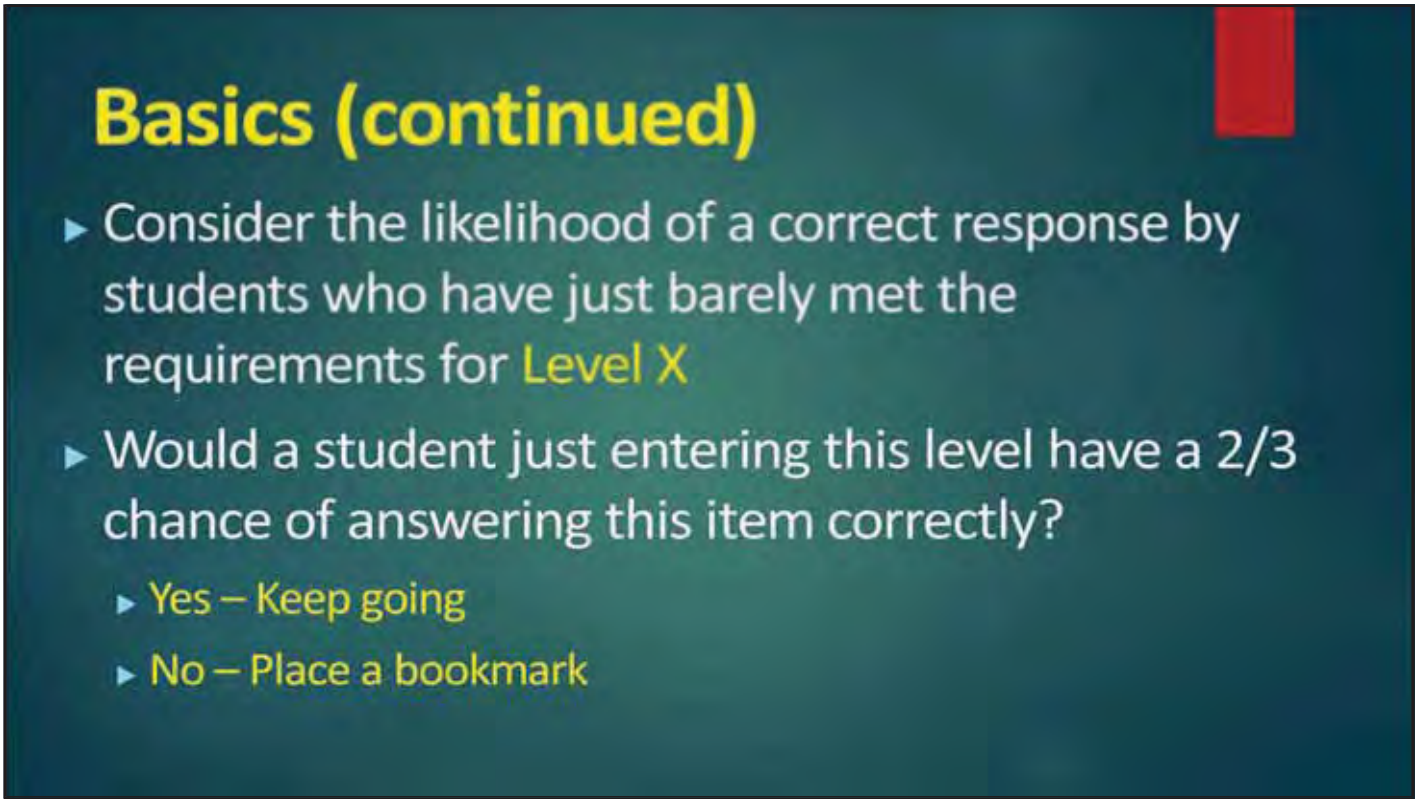
We have already scored nearly all the tests. We know how well the students did, and we know how difficult each item is. We have taken the test booklet students used this spring and arranged the items from easiest to hardest. The booklet you will work with will therefore look quite different from the one students saw, but it will contain exactly the same items.

We will ask you to read every item and help us find the dividing line between Level 3 and Level 2 and then the dividing line between Level 2 and Level 1. You will do that by placing bookmarks at two points in your ordered item booklets (OIBs) where you believe the items become too difficult for students at the threshold of Level 2 and Level 1 to answer correctly. We will elaborate on difficulty and threshold momentarily.

We will take the bookmarks you place in your booklets and translate them into cut scores. I will also explain the process by which we do that.



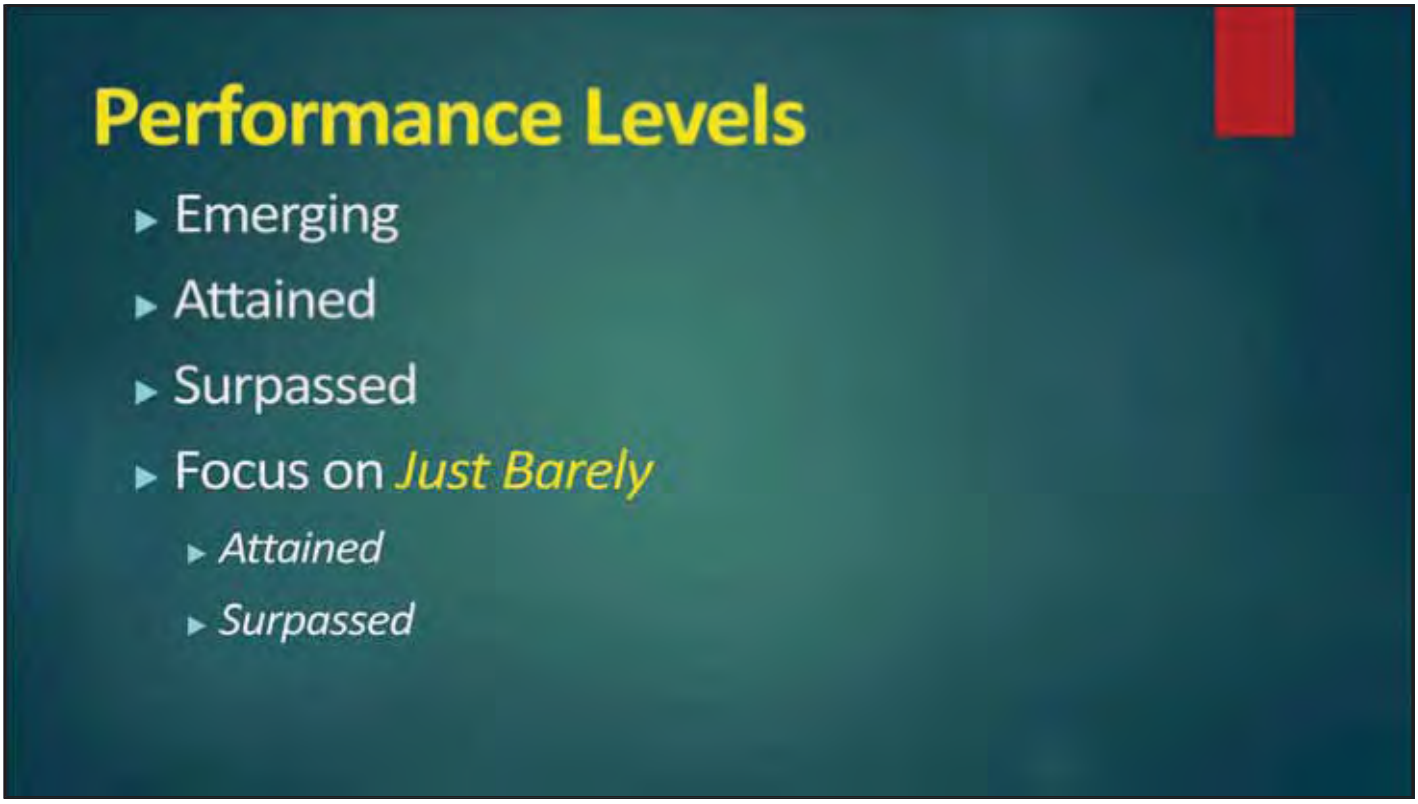
When I say that the items are arranged from easiest to hardest, I am referring to how students performed on them this spring, not our subjective judgment about the difficulty of the items. The item that the most students got right is on page 1; the item that the fewest students got right is on the last page. We want you to examine each item in order and consider what a student would have to know or be able to do in order to answer it correctly or in the case of the writing prompt, what a student would have to know or be able to do to receive a score of 1, 2, 3, or 4. As you consider each item, think about the student who is just at the threshold of Level 2. Is that student likely to answer this item correctly or earn this score point (or better) on the writing prompt? Once we have identified a point in the booklet where we have to say goodbye to the threshold Level 2 student, we start asking the same question about the threshold Level 3 student.



Specifically, as you begin to go through the ordered item booklet, I want you to think about the student who just barely makes it into Level 2, as defined by the PLDs you studied this morning. You will have a copy of those with you at all times as you complete this task, so please refer to them often. When we say "students who have just barely met the requirements for Level X," we mean this: As you start through the booklet, you are looking for the point at which a student who jus barely qualifies for Level 2 will no longer have a reasonable chance to answer correctly. For the purposes of this activity, we are defining reasonable chance as 2 out of 3. We can look at this in two ways:

1. If we put together several items just like this one, would the student just barely in Level 2; i.e., the threshold Level student, be able to answer 2/3 of them correctly?
2. If we found several threshold Level 2 students, would about 2/3 of them be able to answer this item correctly?

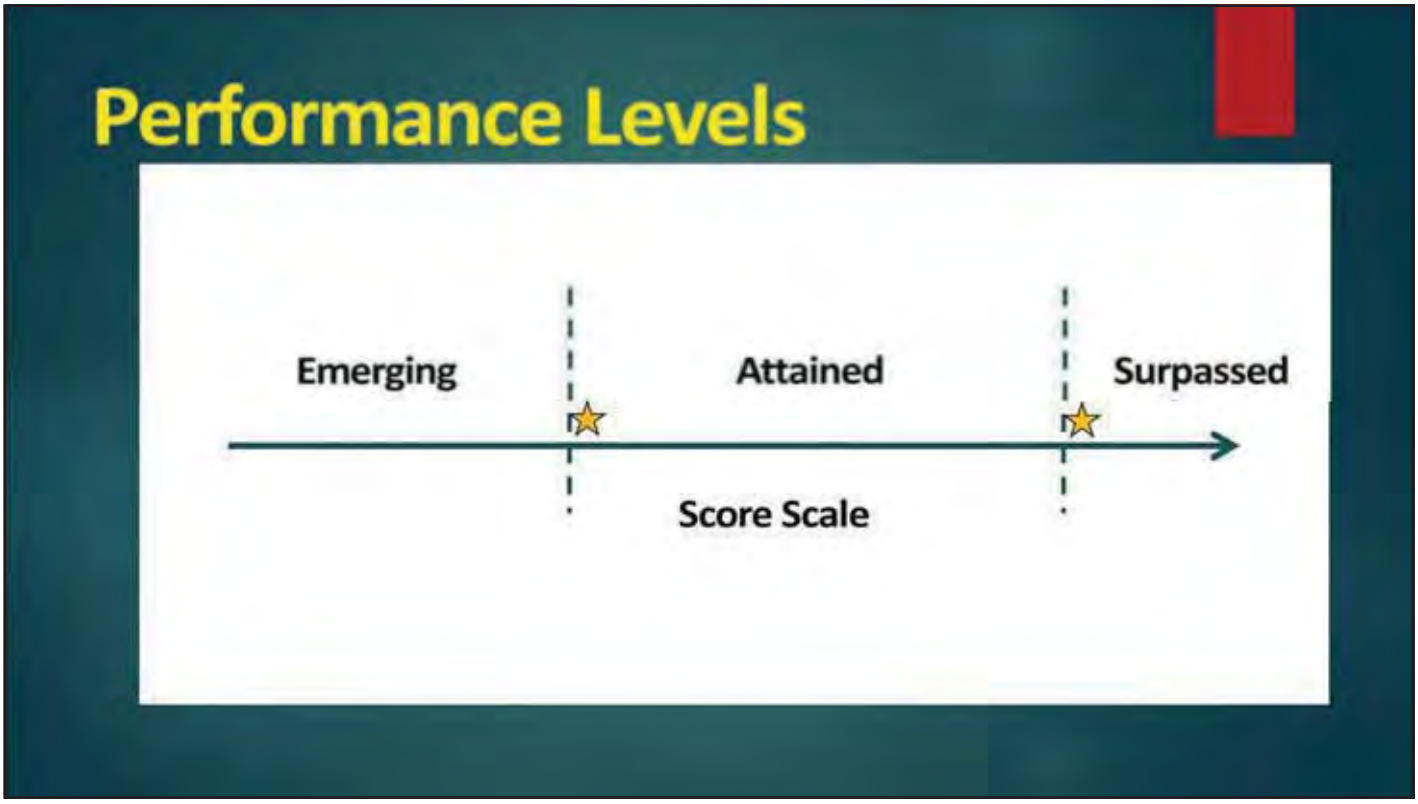
However we look at it, we want to answer that question about each item in the ordered item booklet. If we can answer the question Yes, we keep going. If we have to say No, we stop and place a bookmark because that is the place in the booklet where the threshold Level 2 student no longer has a reasonable chance of answering correctly. If all the rest of the items in the booklet are harder than this one, it is very unlikely that that student would have a reasonable chance of answering any of them correctly. However, once you have placed a bookmark, check the next two or three pages, just to make sure.



Let's remember what we saw earlier today when we examined the achievement levels. Recall that you spent time discussing what it means to achieve at the Attained and Surpassed levels.

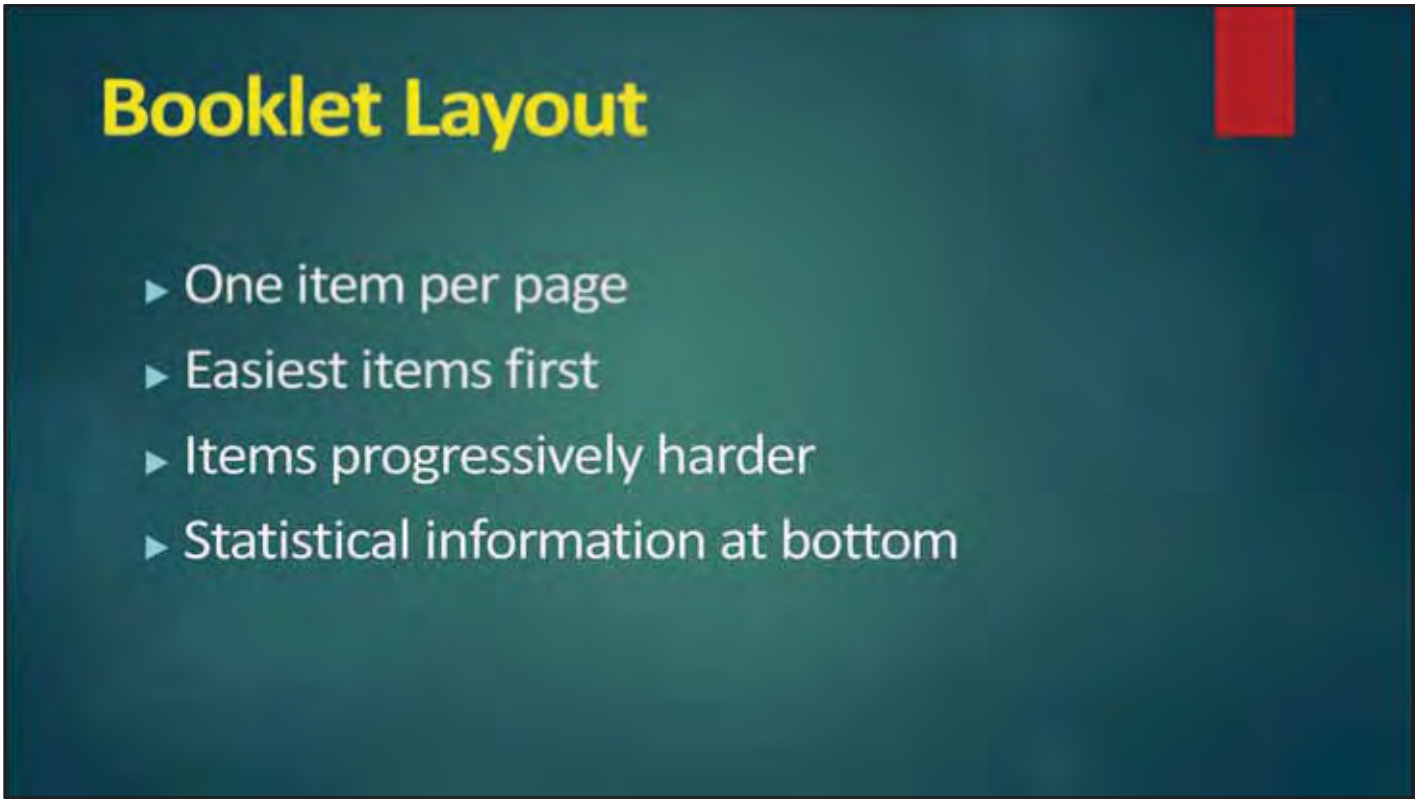
When you do standard setting you will be considering a special segment of each of those populations, students who just barely qualify as Met Expectations or just barely qualify as Exceeded Expectations. These are the students that we just tried to describe and think of.

You may be wondering about Emerging. We will not focus specifically on that group. You will recall from earlier that we are trying think of lines to separate different levels of performance. For example, we tried to think of a line to separate Emerging and Attained. Once we establish that line, everyone below it will be in the Emerging level.
The PLDs and your understanding of what students *just* barely at the Attained and Surpassed levels can do are vital. They should guide the standard setting process and the cut scores you recommend.



Let me explain why I keep referring to *Just Barely*. If we consider the score scale and the three levels of student performance (Emerging, Attained, and Surpassed), we have divided this large group of students into three still fairly large groups. Within any one group or level, there is still a range of performance or achievement. We are interested in finding the points at which Level 2 (Attained) begins and at which Level 1 (Surpassed) begins. Once we find the score where Level 2 begins, everyone with a score at or above that score will be in Level 2 until we reach the score where Level 1 begins. After we get to that score, everyone who receives that score or higher is in Level 1.

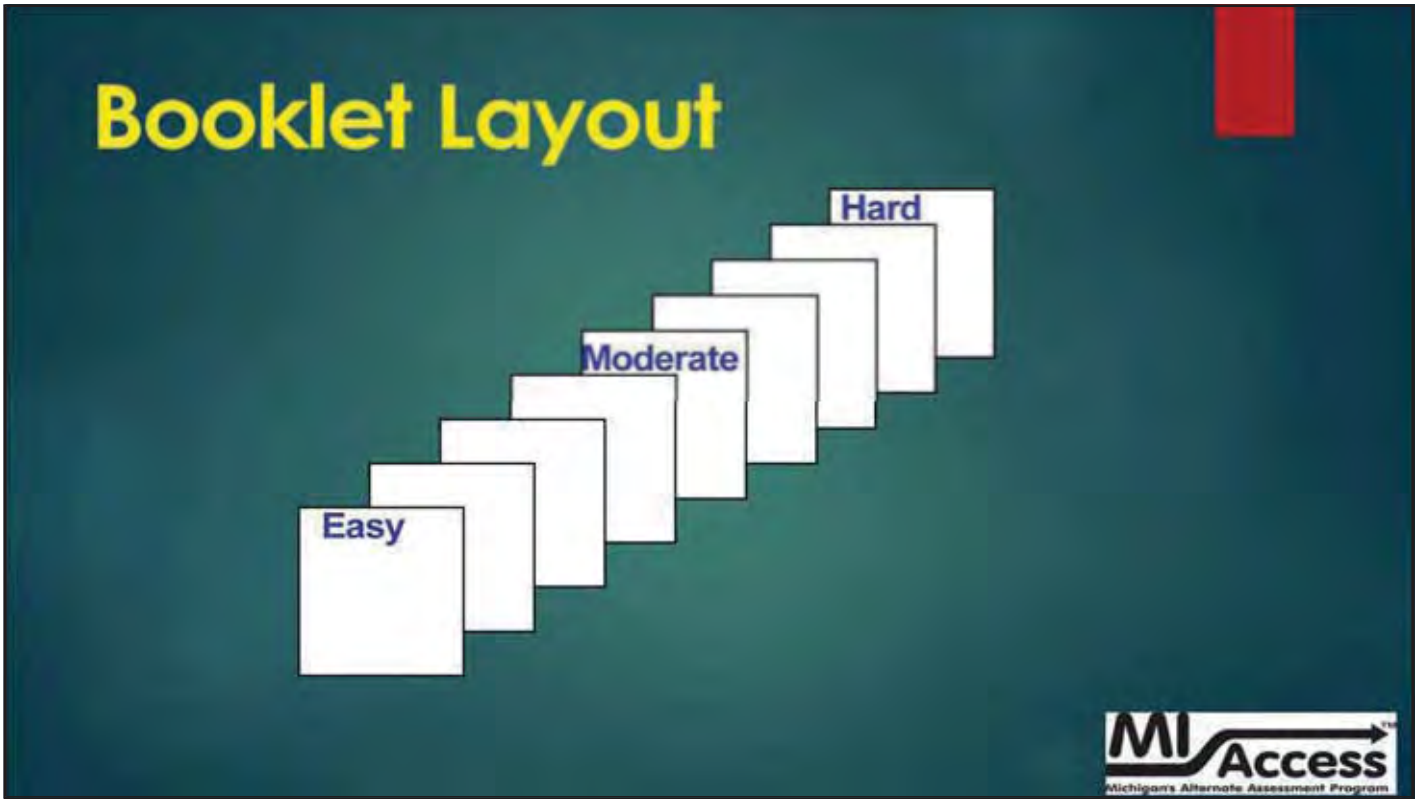
This is fundamental to what we are doing, so I will stop and see if there are any questions.



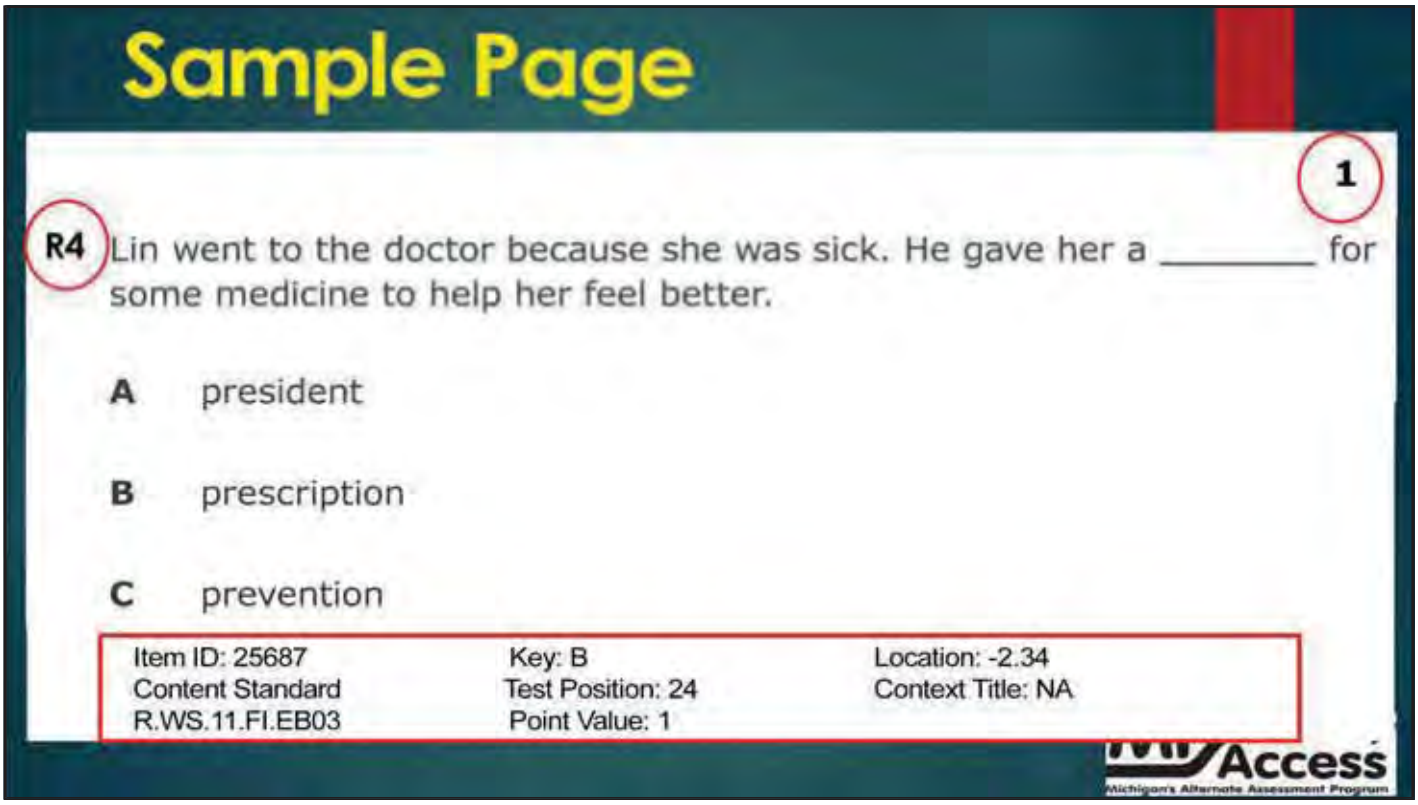
Now, let's consider how an ordered item booklet is actually laid out.

There will be only one item on each page, as opposed to several items on a page in the students' booklets. Each multiple choice item will be on a page by itself. The easiest item in the entire test will be on page one; the hardest will be on the final page. In between, each item will be harder than the one just before it. Occasionally, there will be two items in a row that are equally difficult. I have to warn you that you will probably see items later in the booklet that you believe are easier than some items earlier in the booklet. Remember that these items are ordered in terms of how students actually performed, not on the basis of anyone's judgment about the intrinsic difficulty of any items.

At the top of each page, you will see the order of the item in the OIB. At the bottom of each page, we have placed information you may want to use as you make your judgments. We have included some statistical information that will later prove useful. We have used a statistical model that lets us assign a scale value to each item and achievement levels to each student. These difficulty and achievement levels are on the same scale so that if we know the Rasch difficulty index for an item, we can calculate the achievement level a student would need in order to have a 2/3 chance of answering the item correctly. This model allows us to calibrate all the items and all the students and place them on a common scale that describes item difficulty and student ability in a way that allows us to compare one to the other.



Here's what a booklet will look like. This is an abbreviated form of a booklet, but it should give you a good idea.
Here we see an easy item on the first page
In the middle, items are of moderate difficulty, getting more and more difficult as you get further into it.
Finally, the most difficult item on the test is on the last page.

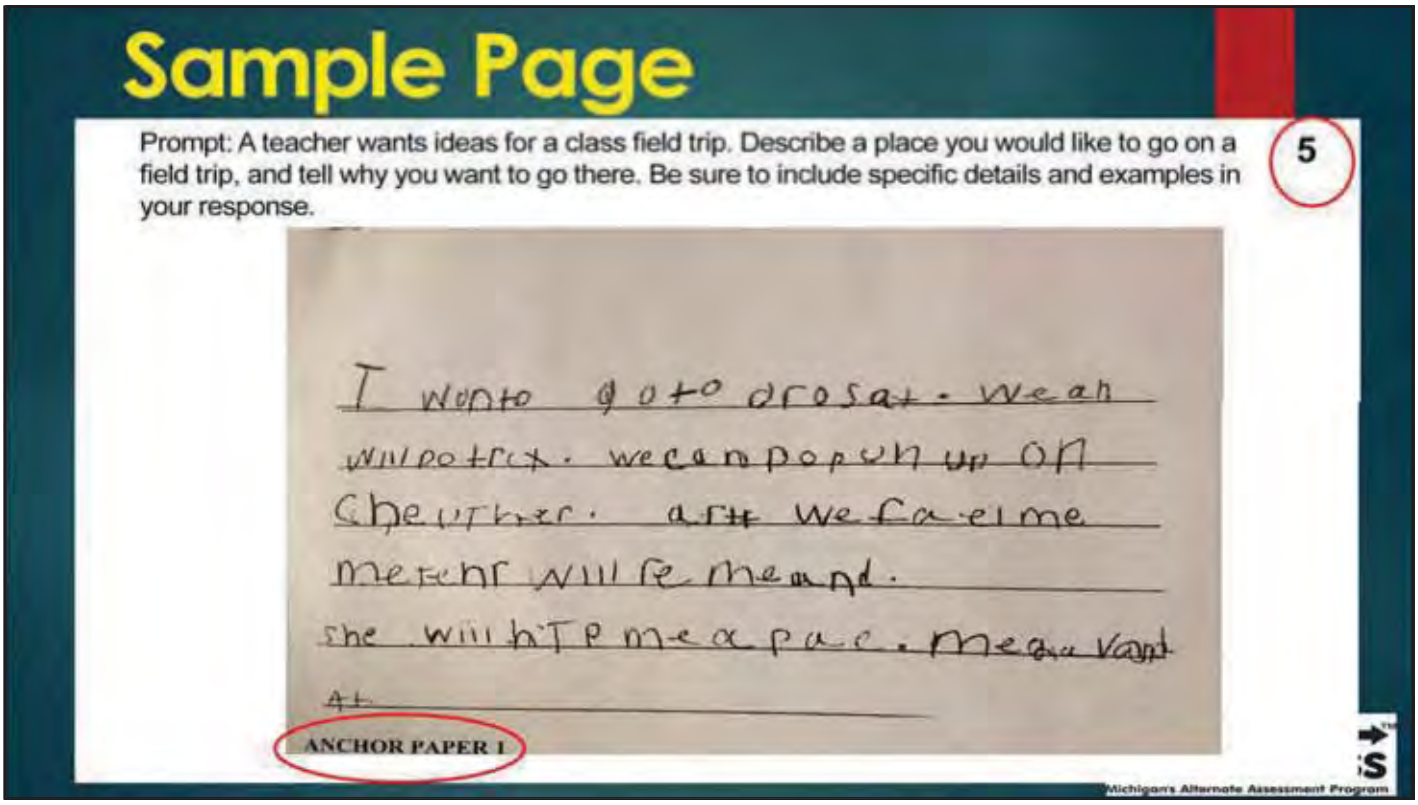


Here's a sample multiple choice item page.

You will notice that the whole item is here: the question or stimulus, the three answer choices, and where the item appeared in the original test booklet. This item was the fourth item in the original test booklet but the first item in the OIB because it turned out to be the easiest.

A student would not need to be at a very high level of achievement to answer it correctly. Easier items are associated with lower location values, and harder items are associated with higher location values..

To speed things along, we have indicated the correct answer for each item, which you can see here is B. All OIBs are laid out in this fashion.



For ELA, there will be one writing item worth up to four points. The OIB will have a sample for each of the four score points. Since it is easier (we assume) to get a score of 1, the sample of a 1 response will appear early in the booklet. Samples for score points 2, 3, and 4 will appear later. Here's a sample of a page with a writing response on it.

You will notice that the whole item is here: prompt, response, and score. Those of you working with ELA booklets saw the rubrics this morning and will have those available as you review your OIBs.

At the bottom right, you see the item order and the Scale score associated with this item. This item is on page 1 of the booklet, thus it is the easiest item on the test. A student would not need to be at a very high level of achievement to answer it correctly. The hypothetical scale ranges from 100-300. Easier items are associated with lower scale score and harder items are associated with higher scale scores. This item's scale score is 110, at the easy end of the scale.

To the right of the keyed response you will see an X identifying, in this example, that answer choice B is the correct response.

This example item is a Math item. The reading items will be in the exact same format, except that they will also have reading passages listed above the stem of the item.

Immediately after this introduction, you will break out into your individual work groups – which we call panels – to practice using the Bookmark procedure. You will have a chance to apply the procedure, ask questions about it, and then let your group facilitator know that you are ready to begin Round 1. There will be three rounds in all, with review and discussion in between.

After Round 1, you will have the opportunity to engage in discussion within your panel. This discussion gives everyone the opportunity to explain how they arrived at their standard setting judgments, to sort through and discuss potential differences, and try to and come to a common understanding of how to interpret the PLDs. Your facilitator will help lead these discussions.

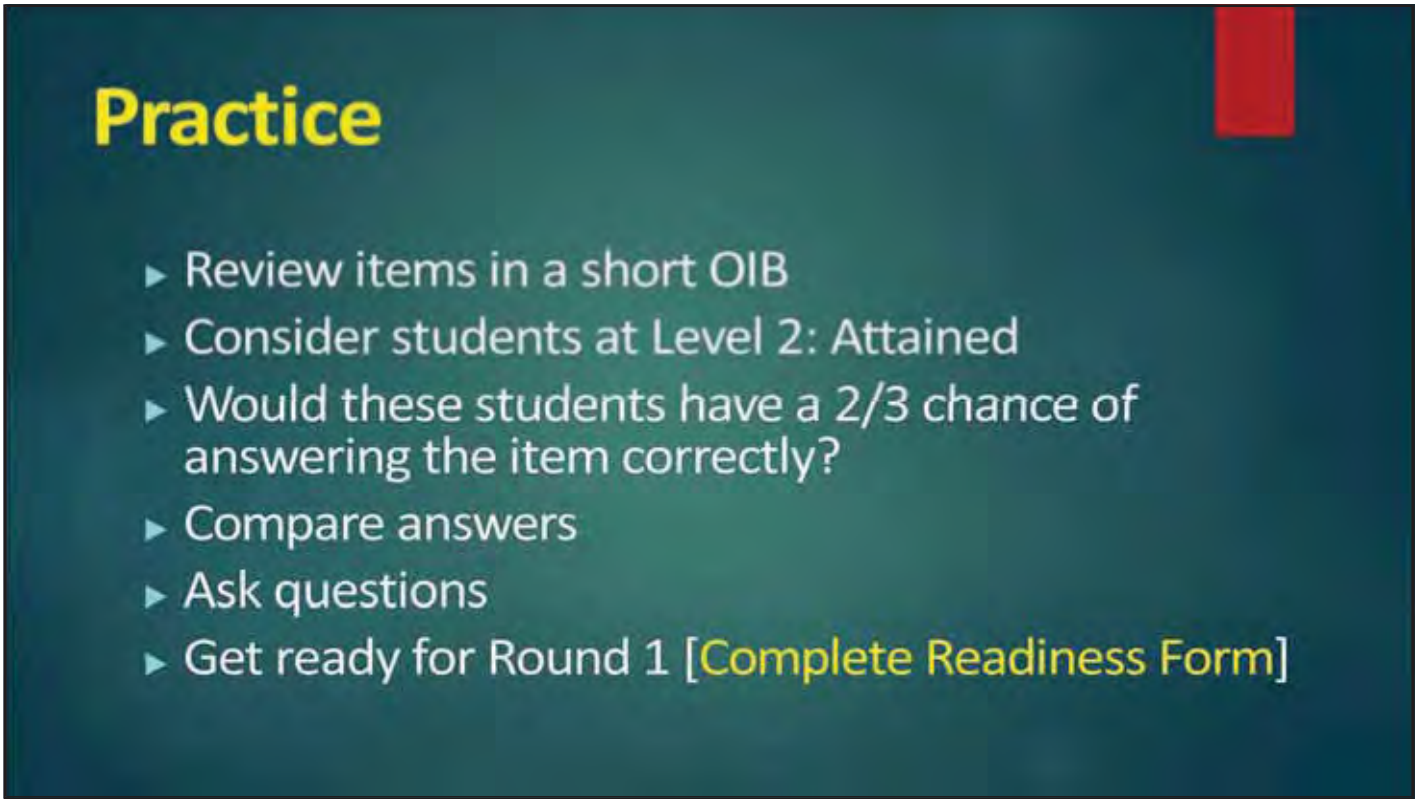
We will also show you impact data. The impact data will show you what the practical implications are of your cut scores on MI-Access students for your particular grade and subject. Following the presentation and consideration of these data, we will ask you to provide your second cut score recommendations. These recommendations should be your best judgments for cut scores based on all the information you have received during standard setting.

After Round 2, you will have the opportunity to engage in discussion within your panel. This discussion gives everyone the opportunity to explain how they arrived at their standard setting judgments, to sort through and discuss potential differences, and try to and come to a common understanding of how to interpret the PLDs. Your facilitator will help lead these discussions.

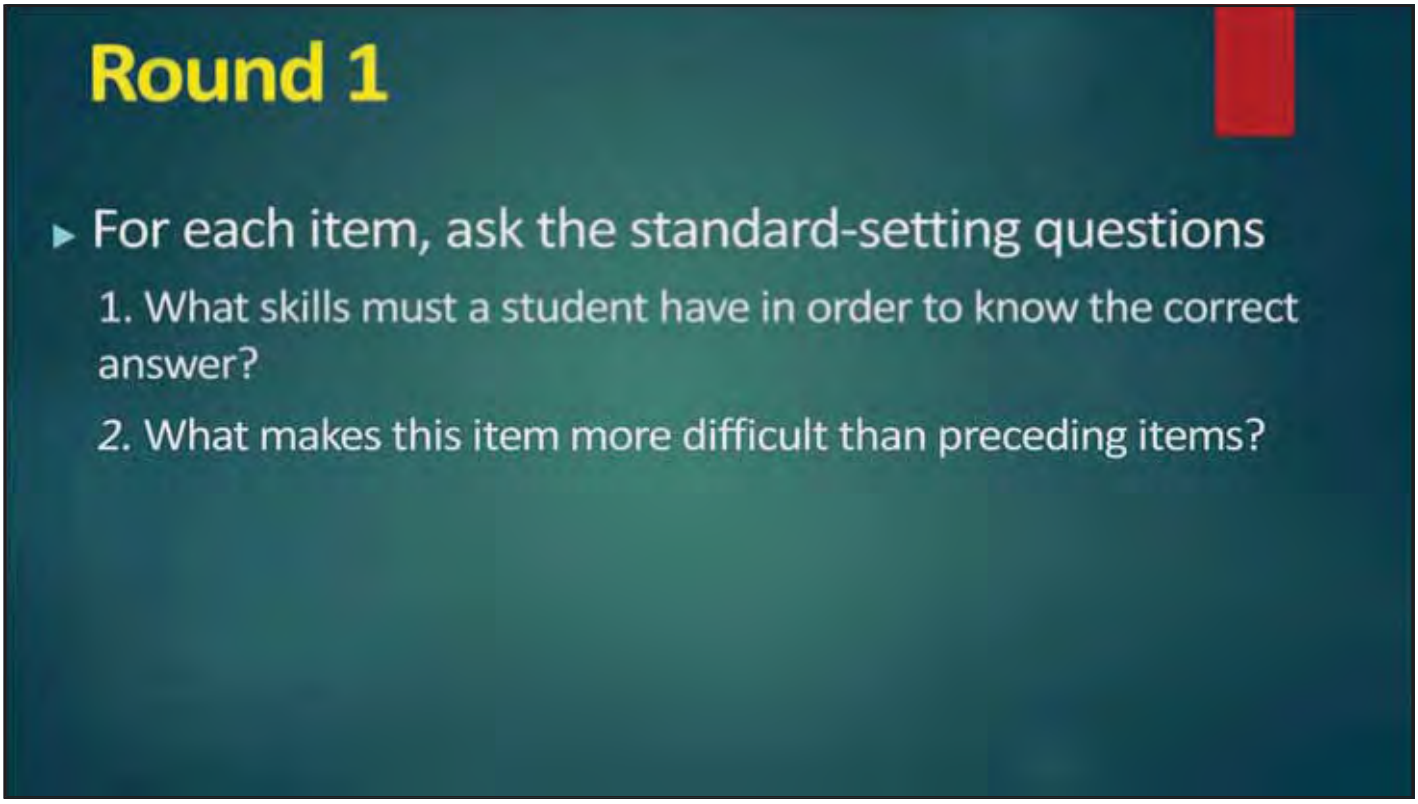
We will also show you impact data from this year as well as impact date from last year for comparison. Following the presentation and consideration of these data, we will ask you to provide your third and final cut score recommendations. These recommendations should be your best judgments for cut scores based on all the information you have received during standard setting.

After Round 3, some of you will be involved in a review of all cut scores for your subject across grades. The rest of you will take part in a review and critique of this process. Your facilitator will let you know on Wednesday which group you will be in.

We will also check to make sure that you understand what you are being asked to do and we will answer any questions that you have throughout the course of the process. You will have the opportunity to provide us with feedback, how you understood different components of the process, and how you arrived at your judgments. These data will be collected in readiness and feedback forms and they are very important to us to make sure that process is working as intended. You also will have the opportunity to complete a final evaluation form of the whole standard setting process. Please fill out this information and provide us with your opinions.



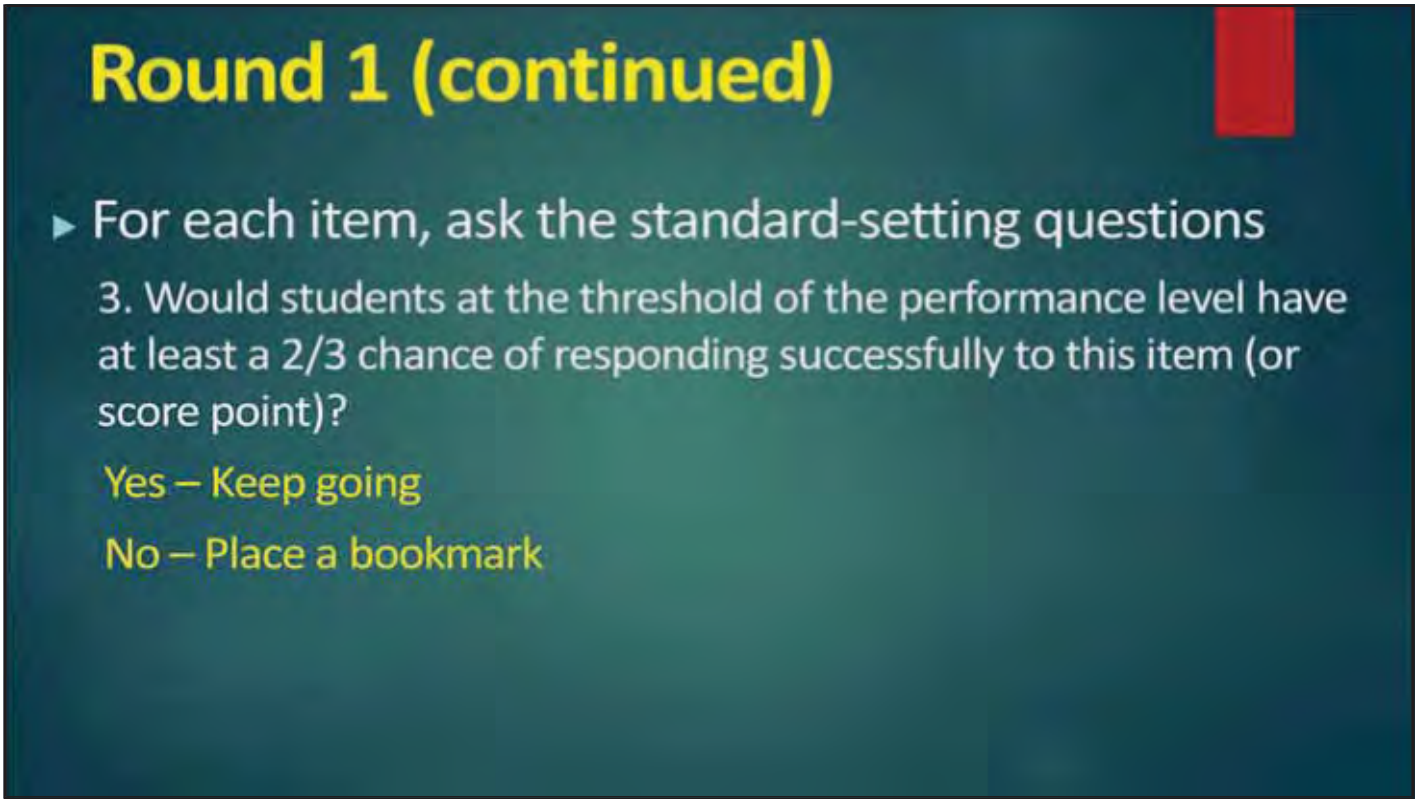
This practice set will consist of a smaller ordered item booklet. This will occur before you do any actual standard setting. We will ask you to examine items and decide if you think just barely Met Expectations students would have a 2/3 chance of answering the item correctly. After you participate in examining the practice ordered item booklet and decide on a practice cut score we will discuss your ratings and try to reach some form of consensus. We will not give you set actual cut scores in this practice round, but we will discuss how you would go about providing actual cut score judgments from the materials that you received. This is designed to get you acclimated to what you are being asked to do and to ask specific questions before we get started.



In the first round of standard setting, you will review an ordered item booklet that consists of all the operational items on the MI-Access test for your subject and grade band. You will get to see how the probability of success on each item relates to the scale score. You will ask yourself two basic questions about each item:

1. What skills must a student have in order to know the correct answer?
2. What makes this item more difficult than preceding items?

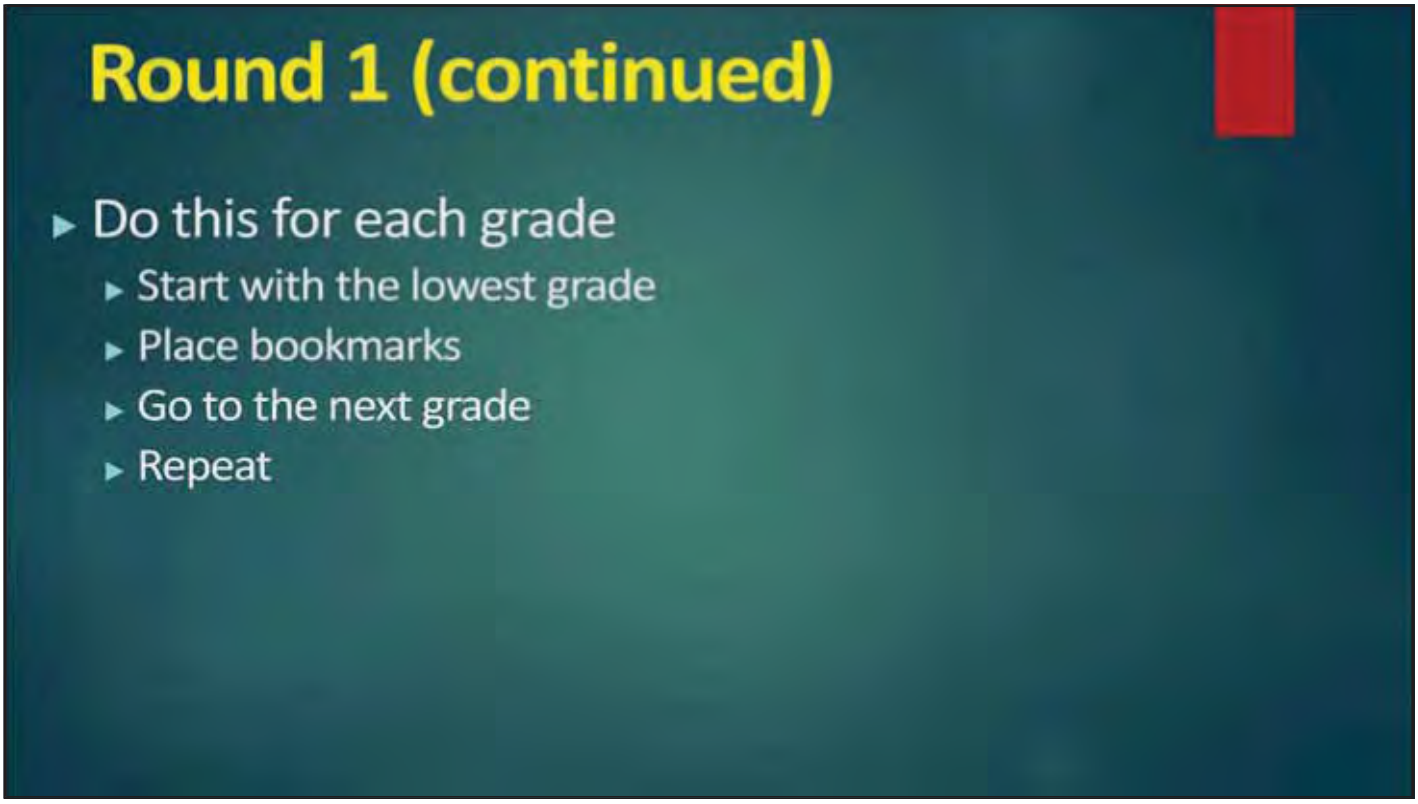
On this second question, you may not find much or even any difference in difficulty between one item and the next, but over the course of several items, you should notice that the sixth or seventh is more difficult than the first or second.



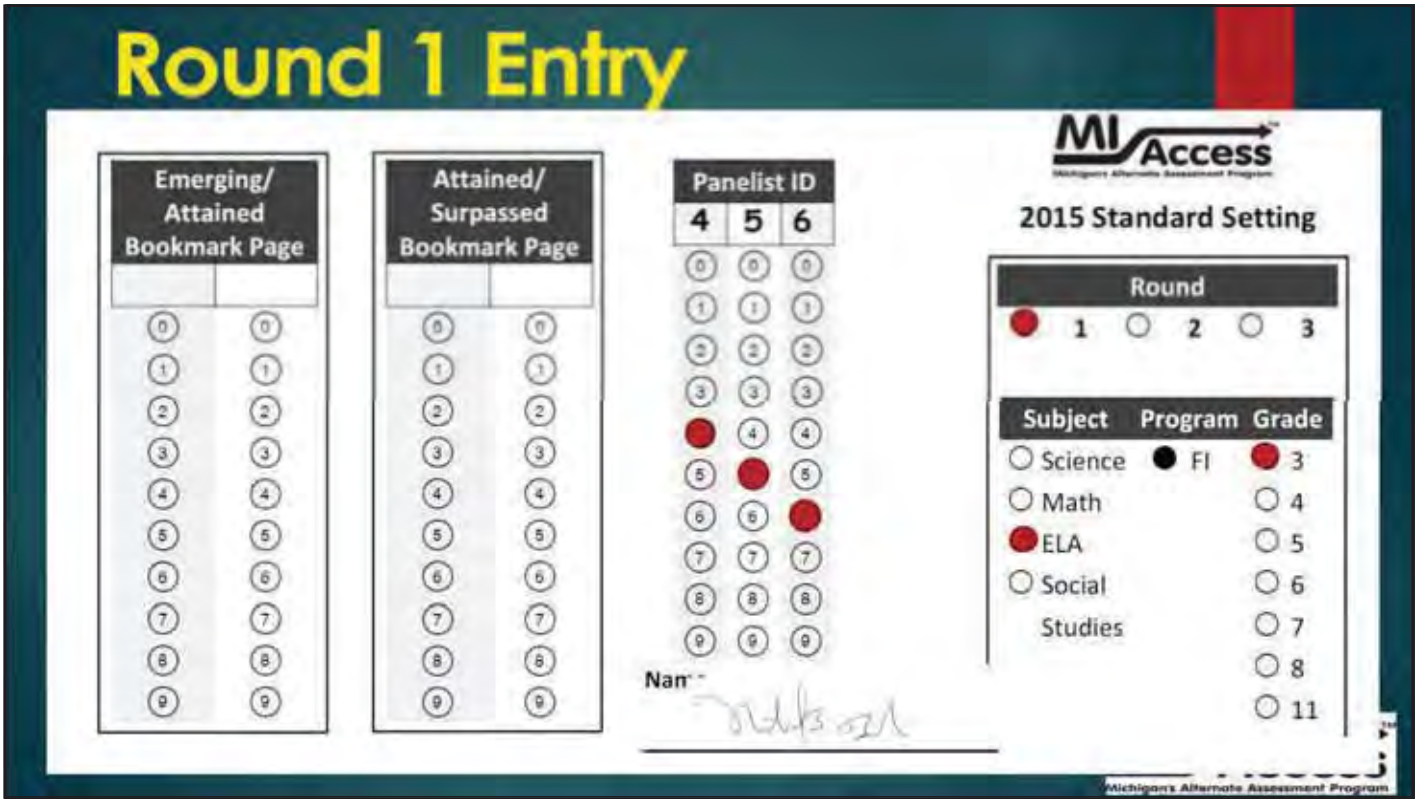
There's also a third question:
3. Would students at the threshold of the performance level have at least a 2/3 chance of responding successfully to this item (or score point)?
If you answer Yes, keep on going. If No, stop; place a bookmark here. Assuming that you started with the threshold Level 2 student in mind, you have found the place where that student no longer has a reasonable chance of answering correctly. The item on the page before your bookmark contains the last item the threshold Level 2 student has a 2/3 chance of answering correctly. Now you can start looking for items that would be too difficult for the threshold Level 3 student to answer correctly.

At the conclusion of the first round of standard setting, you will have provided cut score judgments, and you will have transferred them to a rating form that we will use to collect each of your individual judgments. We will then compile all of your recommendations using the median and compute each of the group cut scores.

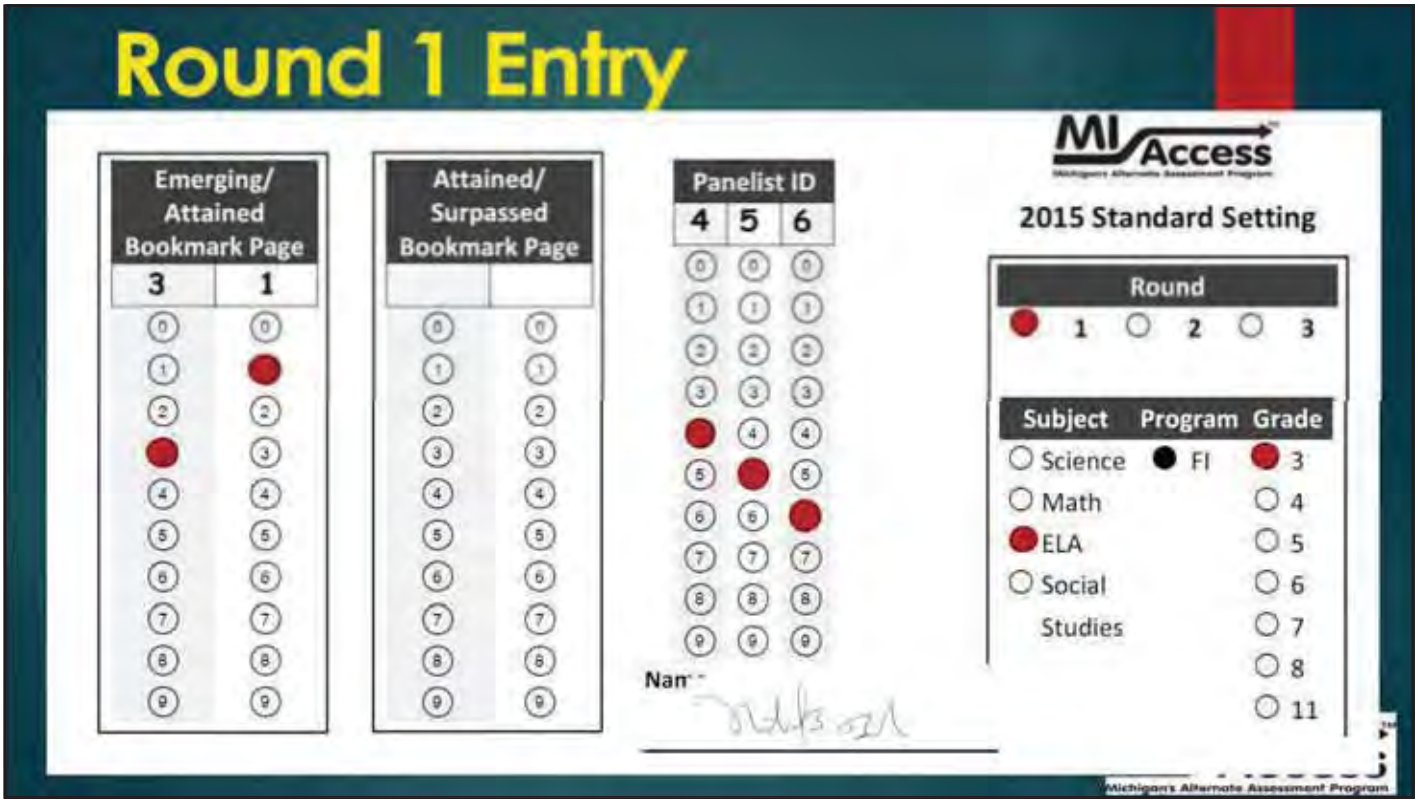
At the end of round 1, you will also have a clear understanding of how your cut scores separate the items in the ordered item booklet that you reviewed into the different performance categories. Remember it is the performance levels and your understanding of the just barely Attained and just barely Surpassed students that should guide your cut score recommendations.



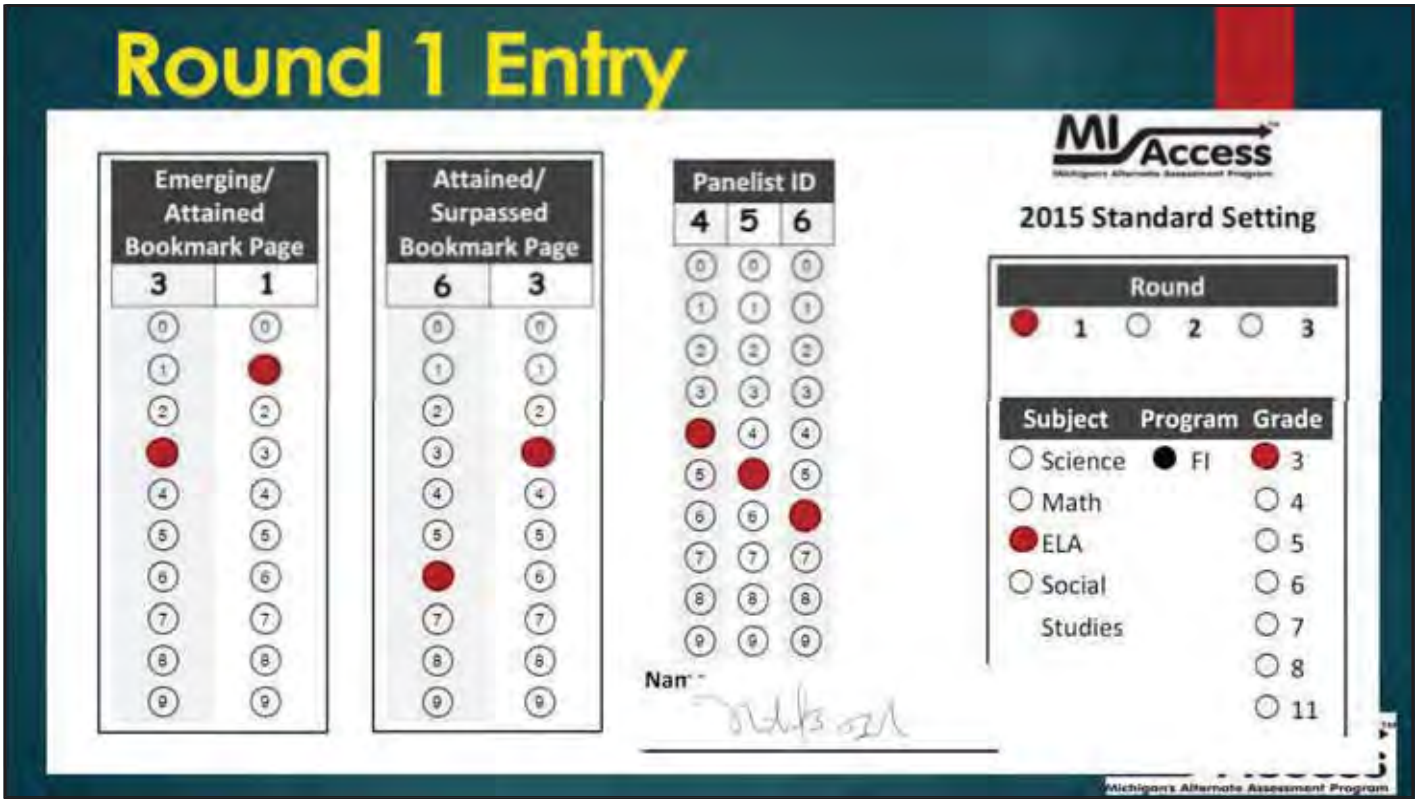
Most of you have booklets that span multiple grades. We want you to set bookmarks for each grade separately. Here's how you should do that…



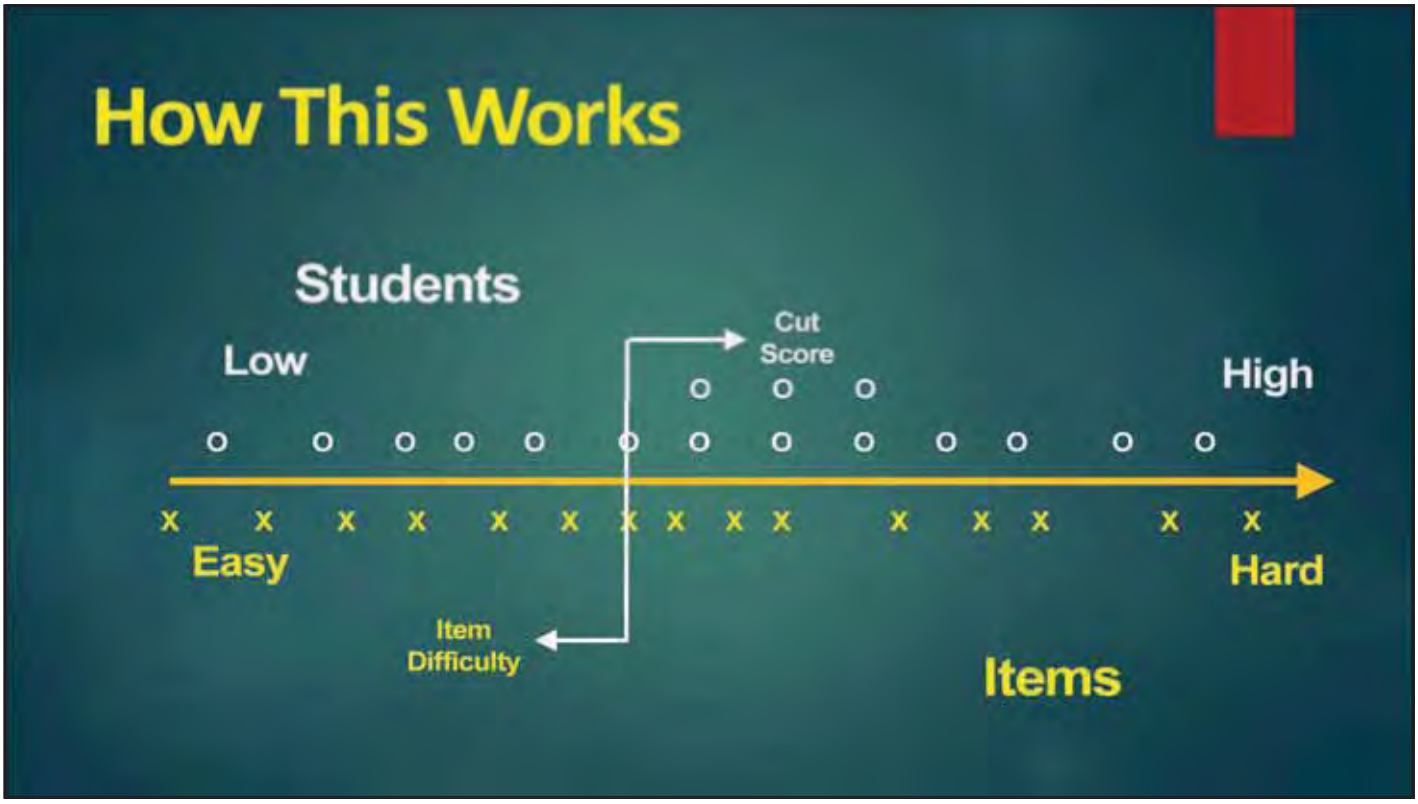
This is what the bookmark form will look like.
Fill in panelist ID Number, which will be on your panelist packet.
Let's say I am on the grades 3-4 ELA panel. I start with the third grade booklet and enter Round 1, ELA, Grade 3, and my ID: Panelist # 456.



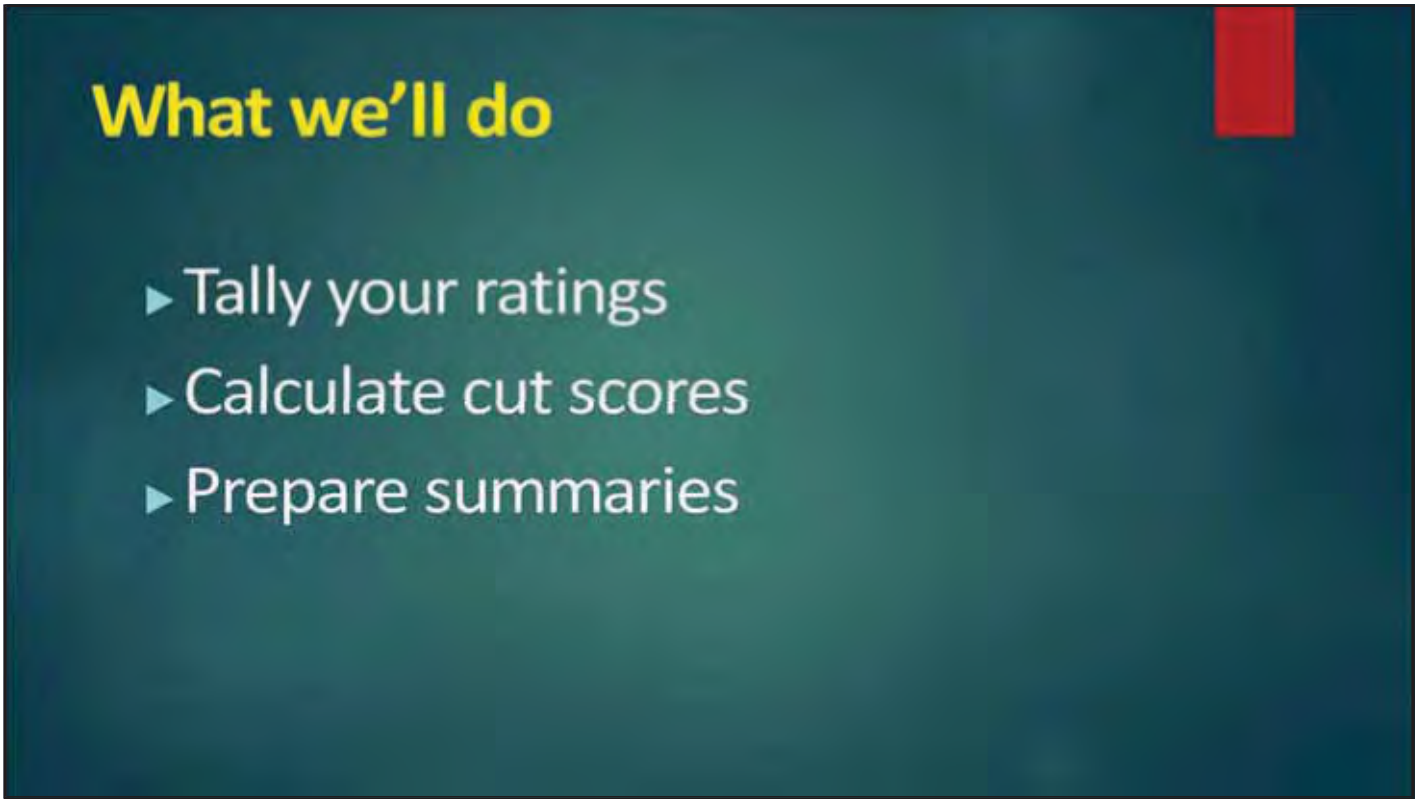
Then, I review the items in the booklet starting on page 1, looking for the hardest item a student just barely performing at the Attained level. When I find that item, and it could include one of the constructed item pages with a score of 1, 2, 3, or 4, I look at the items on the next couple of pages, just to make sure I'm in the right place. Then, I place a bookmark on the first page that I DON'T think a student just barely performing at the Attained level would have a 2/3 chance of answering correctly. Thus, for example, if I think the hardest item a student just barely performing at the Attained level is on page 29, I put my first bookmark on page 30. If page 30 in my ordered item booklet is blank, I put my bookmark on page 31.



I then proceed through the booklet looking for the hardest item a student just barely performing at the Surpassed level would have about a 2/3 chance of answering correctly. Again, that page may contain a multiple-choice item or a sample response that earned a score of 1, 2, 3, or 4. Now, let's say I find that hardest item on page 61. I check the next couple of pages to make sure I have picked the right one, and then I put my bookmark on the first page after 61 that has an item on it. In this case, it is on page 63.
So I enter two bookmarks: 31 and 63, and my completed Round 1 scan sheet looks like this.



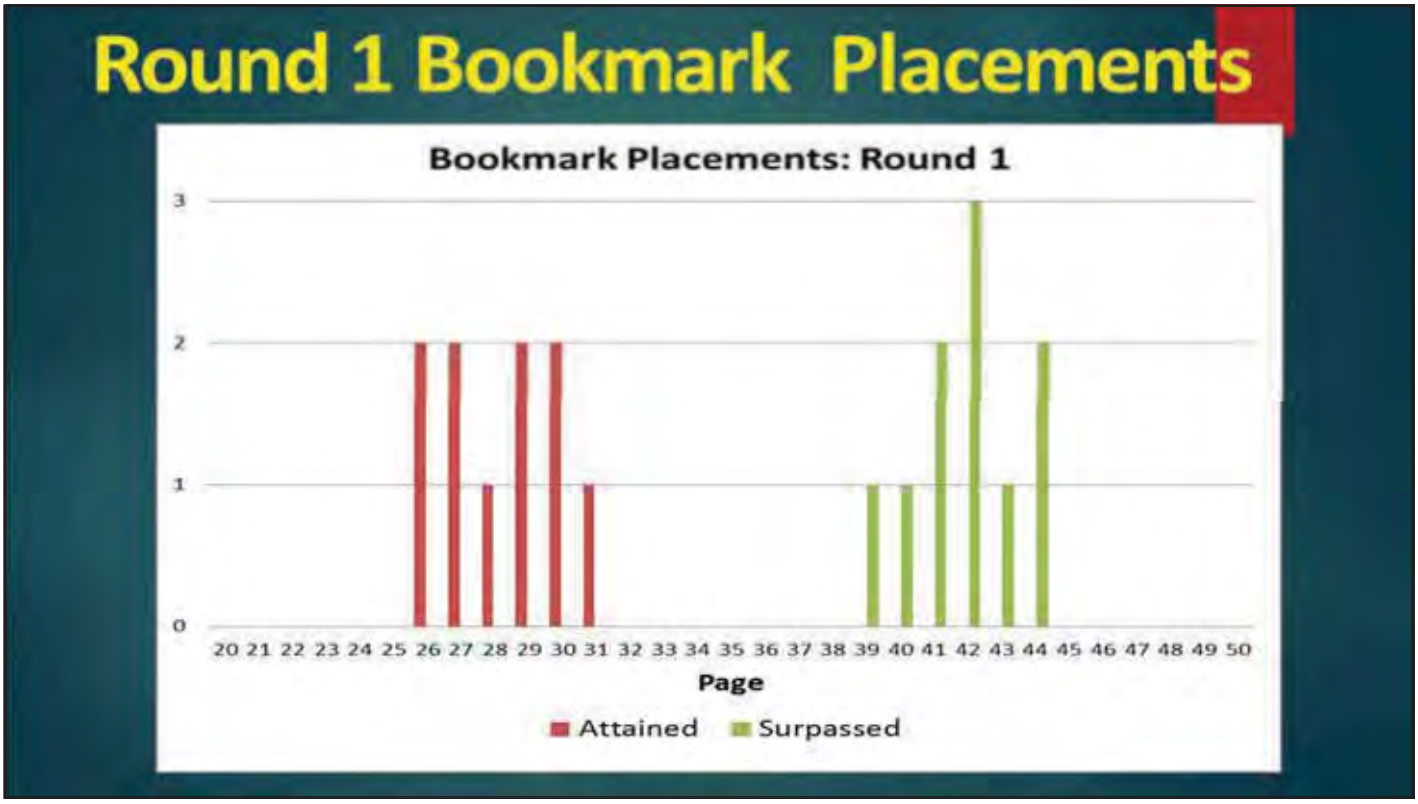
Now, we'll show you how your bookmarks help us find cut scores.
Recall that I pointed out that all the items and all the students can be placed on a continuum that shows item difficulty and student achievement level on the same scale. That continuum is represented by this orange line.
[Click] Below the line, we will show the items, arranged from easy to hard. Each X represents an item, and the ones to the left of the screen are easy, while those on the right are more difficult.
[Click] Above the line, we will show the students, arranged from lowest to highest ability. Each O represents one student. The most important point about what we now see is that if you look straight down from any student, you can see an item that that student will have a 2/3 chance of answering correctly [show a couple of examples]. This student would have a 2/3 chance of answering this item correctly, a greater chance of answering correctly to any item to the left of that item and a weaker chance of answering correctly to any item to the right of that item.
[Click] Now, let's say that you have been working your way through this booklet, starting with the easiest item and you get to a point where you say, "A threshold Level 2 student would not have a 2 out of 3 chance of answering this item correctly but would be able to answer all the ones before this one, so I'm going to put my bookmark right here.
[Click] In this instance, we have placed the bookmark on page 8 (count off left to right). However, the bookmarked page is the first item that the threshold Level 2 student will NOT be able to answer correctly, so page 7 is the last item he or she would be able to answer correctly. At the bottom of page 7, there is statistical information that tells us the scale score required for a student to have a 2/3 chance of answering this item correctly. That scale score becomes our cut score for Level 2.
Let's see if there are any questions before we move on.



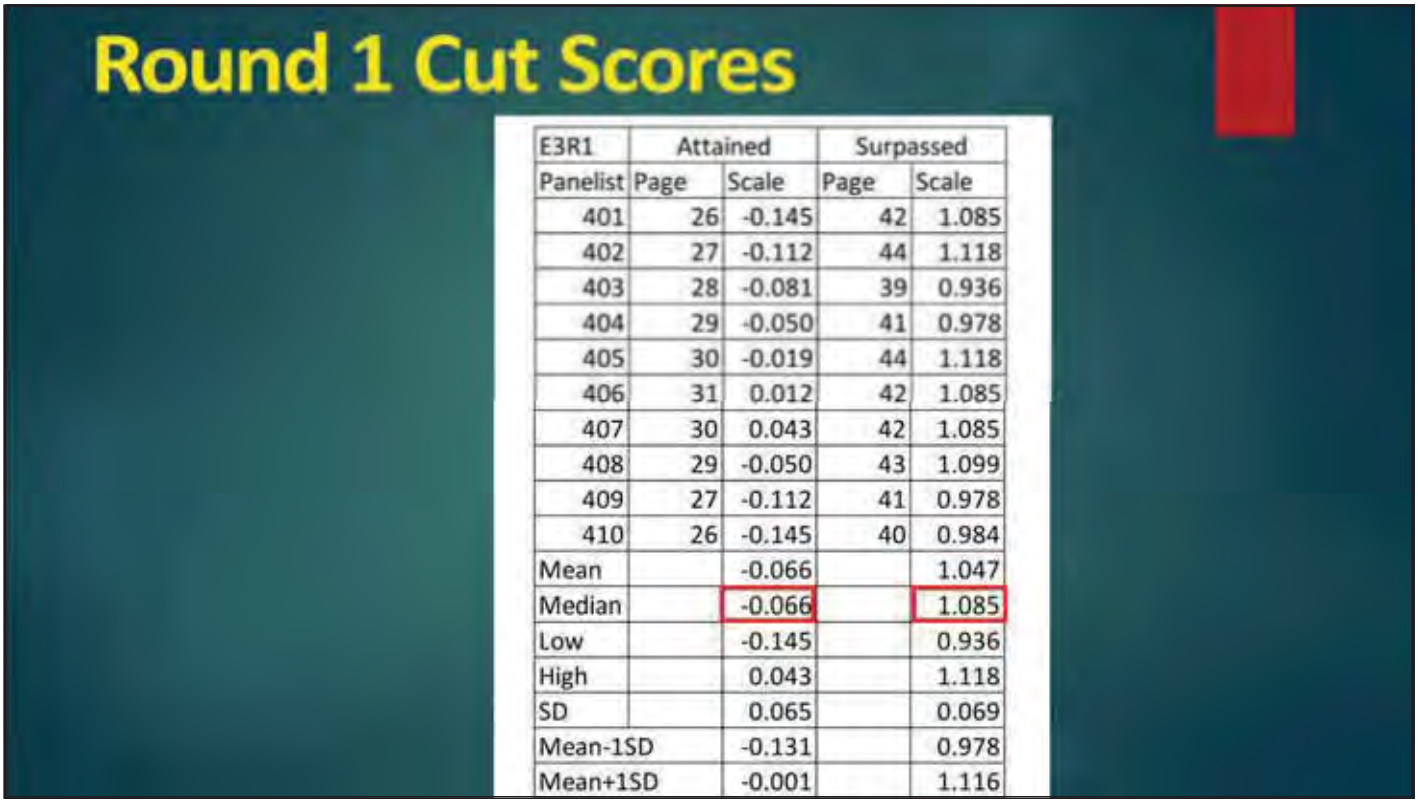
After you have completed your Rating form/construct map for Round 1, you will turn it in to your facilitator. They will check all your materials, make sure they have everything they are supposed to have, and give you your next assignment or dismiss you for lunch. Some of you will finish earlier than others. That's fine.

While you are having lunch, we will tally all the results, translate them into cut scores, do some more calculations, create tables and graphs, and have them ready for you to study and discuss in the afternoon.

Here's an example of what you will see.



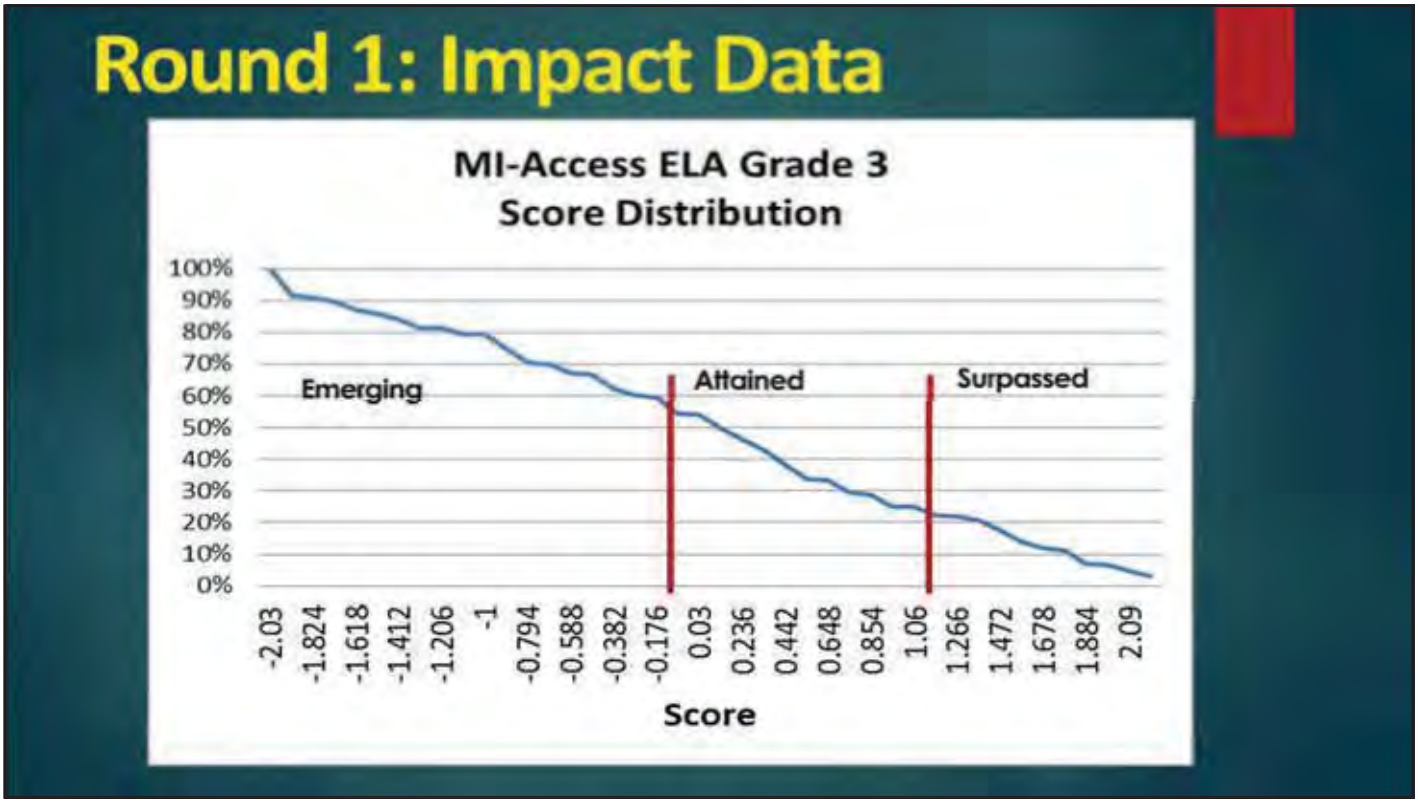
This panel had 10 panelists. You can see how they spread their Level 2 (Attained) and Level 1 (Surpassed) bookmarks: pages 26-31 for Level 2 and pages 39-44 for Level 3. We would want to have a discussion about this dispersion of bookmarks to find out how different panelists using the same PLDs and reviewing the same items arrived at different conclusions.



E3R1	Attained		Surpassed	
Panelist	Page	Scale	Page	Scale
401	26	-0.145	42	1.085
402	27	-0.112	44	1.118
403	28	-0.081	39	0.936
404	29	-0.050	41	0.978
405	30	-0.019	44	1.118
406	31	0.012	42	1.085
407	30	0.043	42	1.085
408	29	-0.050	43	1.099
409	27	-0.112	41	0.978
410	26	-0.145	40	0.984
Mean		-0.066		1.047
Median		-0.066		1.085
Low		-0.145		0.936
High		0.043		1.118
SD		0.065		0.069
Mean-1SD		-0.131		0.978
Mean+1SD		-0.001		1.116

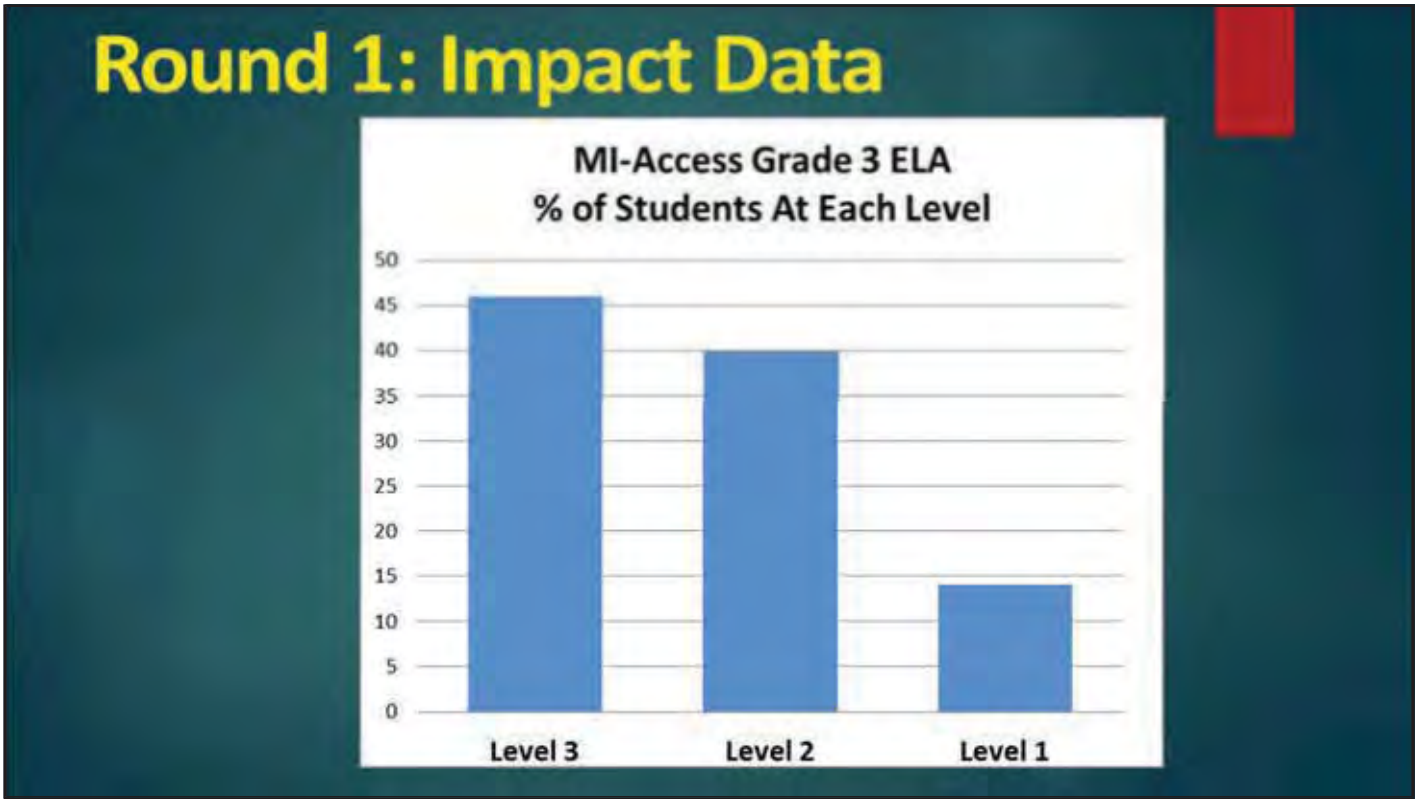
We will take those bookmarks and translate them in to scale score cuts, as shown here, using the process I just described. You will see not only the mean or median cut score for each level for each test but also the range. In this way, you can see that, just as bookmarks were dispersed, so too are your estimates of where the cut scores should be. You will have a chance to discuss these differences before Round 2.

I also point out that what you are seeing here is just for one grade. Once you begin, we will show you results for all grades in your grade band, and you can discuss not only differences of opinion within a given grade but overall differences across grades.



After Round 1 we will also be giving you impact data so you can see the consequences of your cut scores. Here are the consequences of the group's hypothetical cut scores. We will be giving this information to you after Round 1 so that you can ask yourself "Do these percentages seem realistic?"

If you see problems, you have something to consider as you make your Round 2 judgments. The impact data are sometimes called "reality check" data and we include them just to make sure everyone knows the real-world impact of what we are doing here. But we don't want the numbers and percentages to dictate everything we do. We want to make sure all decisions are firmly grounded in the PLDs and test content.

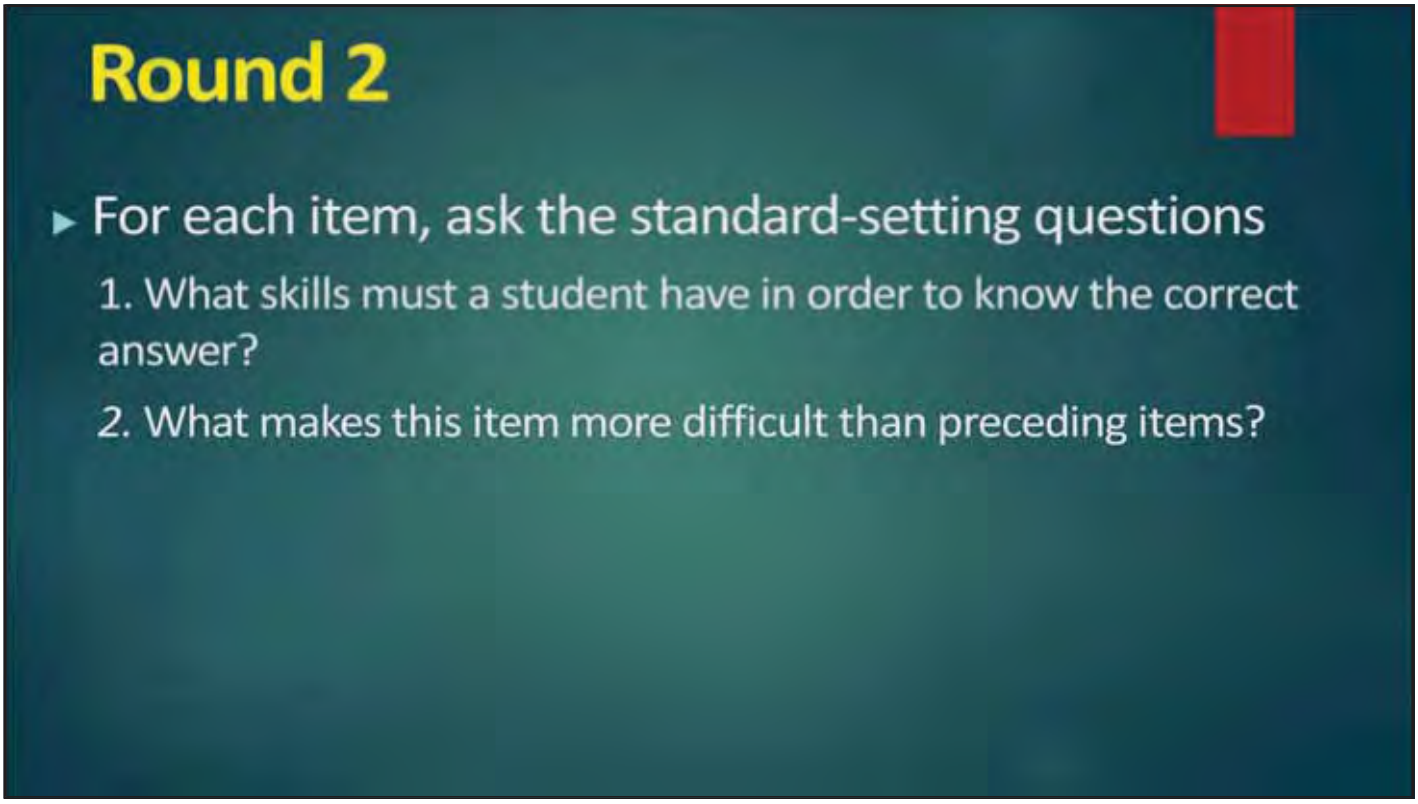


After Round 1 we will also be giving you impact data so you can see the consequences of your cut scores. Here are the consequences of the group's hypothetical cut scores. We will be giving this information to you after Round 1 so that you can ask yourself "Do these percentages seem realistic?"

If you see problems, you have something to consider as you make your Round 2 judgments. The impact data are sometimes called "reality check" data and we include them just to make sure everyone knows the real-world impact of what we are doing here. But we don't want the numbers and percentages to dictate everything we do. We want to make sure all decisions are firmly grounded in the PLDs and test content.

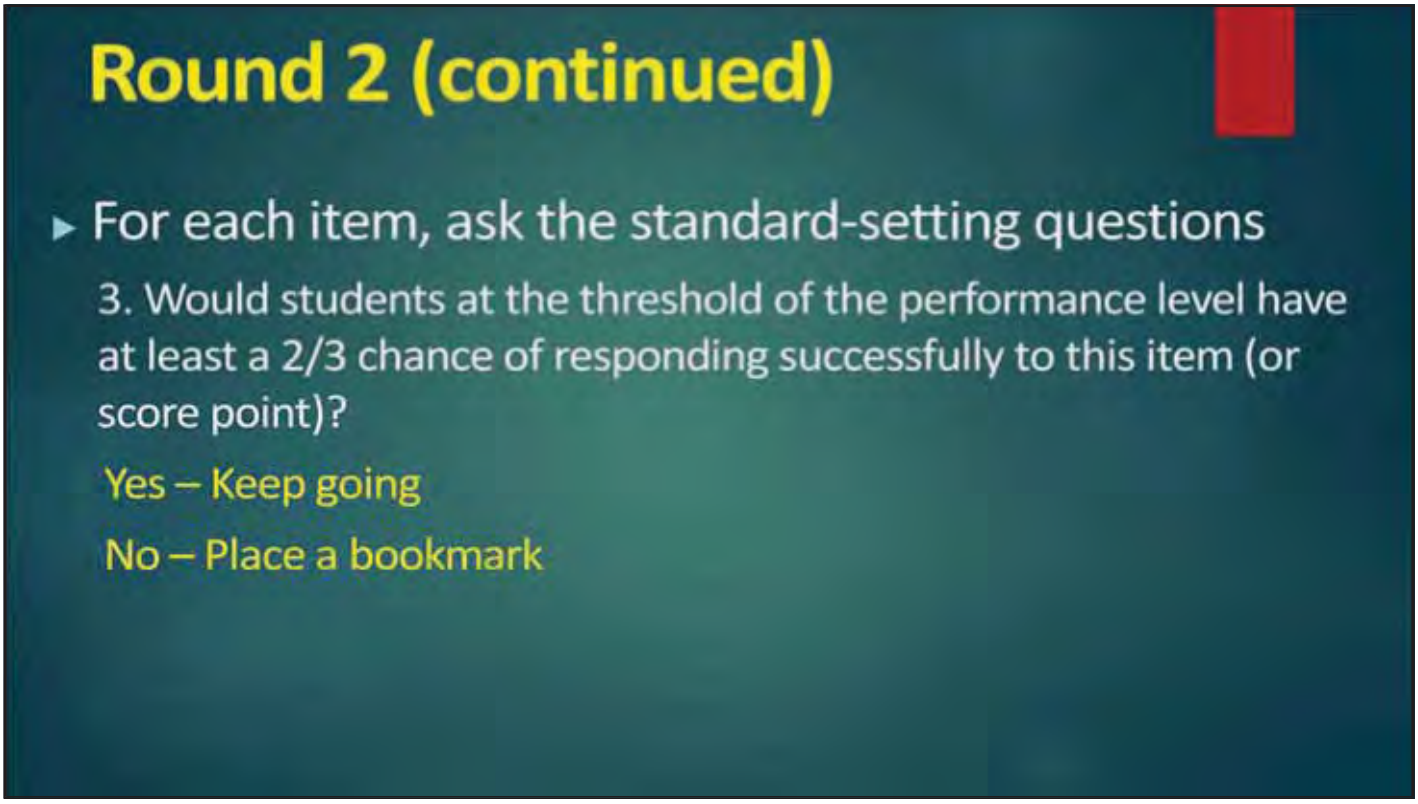
After you have taken a look at the results from Round 1, you will want to discuss these results with your panel. We will help get those discussions started and keep them moving along. We want to make sure everyone has a chance to speak. Topics for discussion include: how different people interpreted the PLDs; what kinds of strategies people used for placing their bookmarks; who's lenient, who's stringent, and why; who's using criteria other than the PLDs and test content; and who's having trouble.

If your judgment is different than the rest of the groups that's okay; we will discuss this and give you the opportunity to provide new recommendations in Round 2. This discussion is important for everyone to check in and understand how others arrived at their judgments.

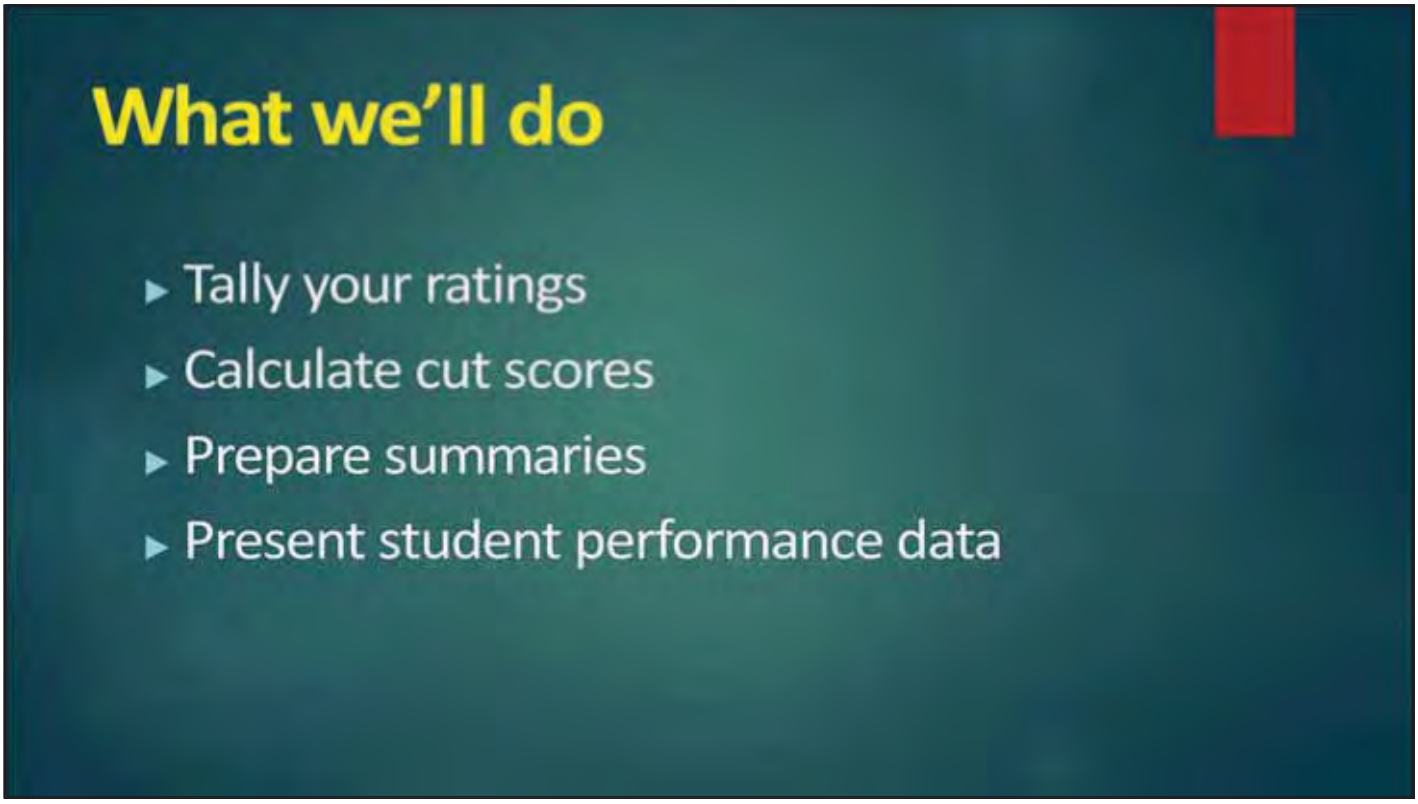


In the second round of standard setting, you will review the same ordered item booklet that you reviewed in Round 1, and you will ask yourself the same questions [Read questions]

However, you will not need to examine every item. At the end of Round 1, you effectively eliminated some early pages because no one set a bookmark there as well as some pages near the end of the OIB because no one set a bookmark there. In the example we just saw, no one set a Level 2 bookmark before page 26 or after page 44. You probably don't need to consider pages 1-25 or 45-50 in Round 2, although they will certainly be there. We find that Round 2 goes much faster than Round 1, partly because you have become more familiar with the procedure but also because you know you don't have to reexamine each and every page.

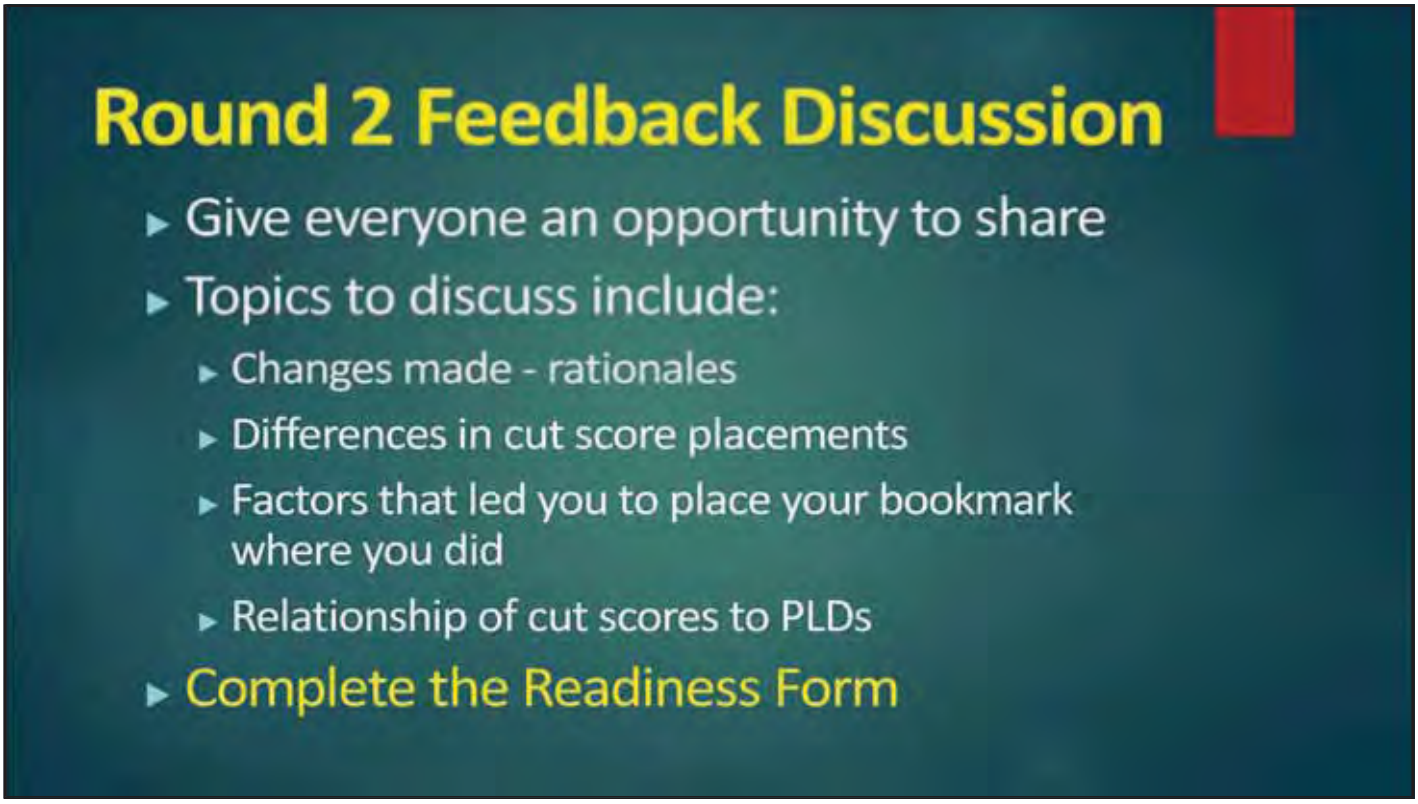


And, of course, you will end with this question. Start with Level 2 (Attained) and ask this question. If the answer is Yes, keep going. If the answer is No, stop and place a bookmark. Afterwards, start thinking about Level 3, asking the same questions.



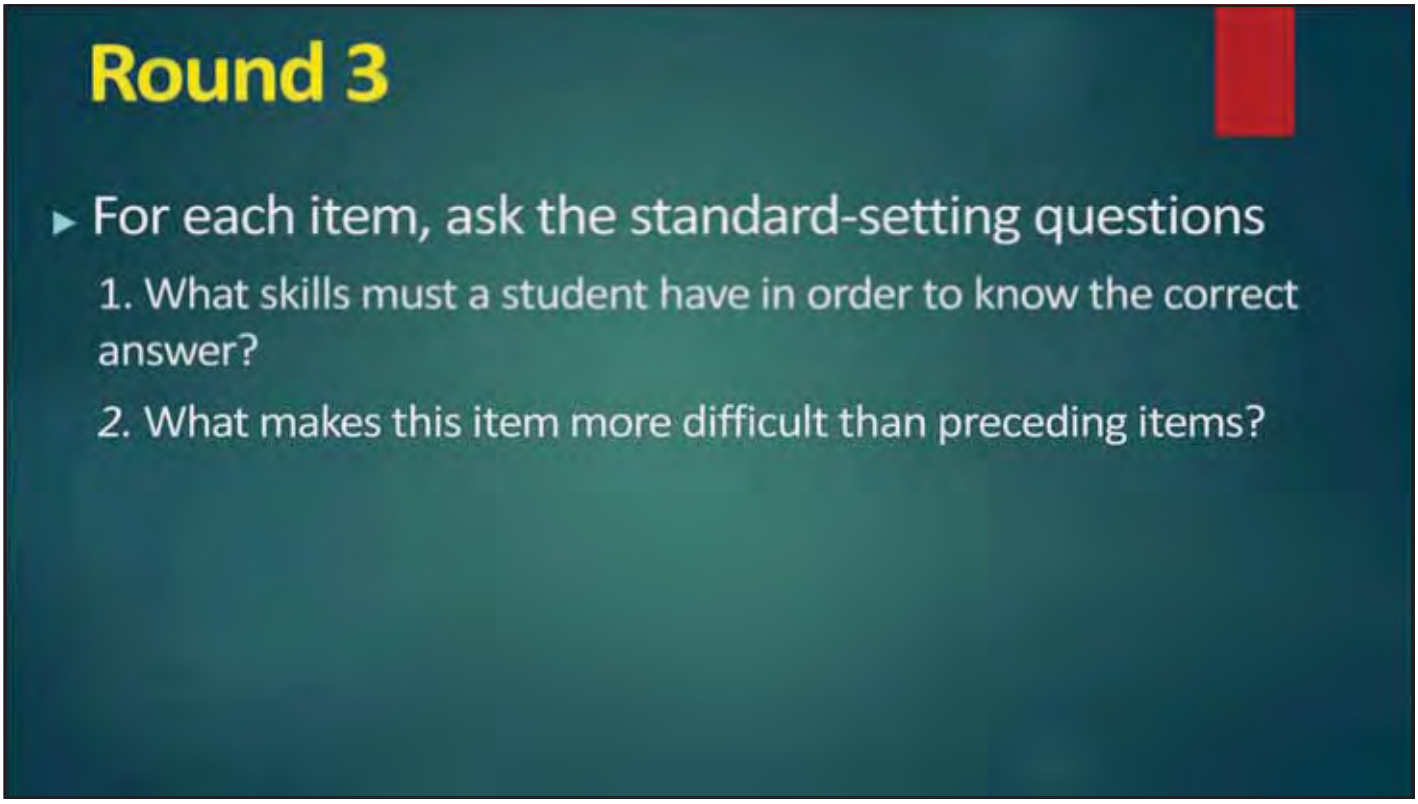
When you finish Round 2, make sure you have completed Round 2 of your rating form, and turn it in to your facilitator. Your facilitators will make sure they have all the materials they are supposed to have from you and give you your next assignment or dismiss you.

After Round 2, we will once again tally your ratings, calculate cut scores, and create tables and charts for you. We will also prepare tables and charts showing the impact of your Round 2 cut scores on MI-Access students. We'll talk more about this impact information in a moment after we look at the Round 2 feedback.

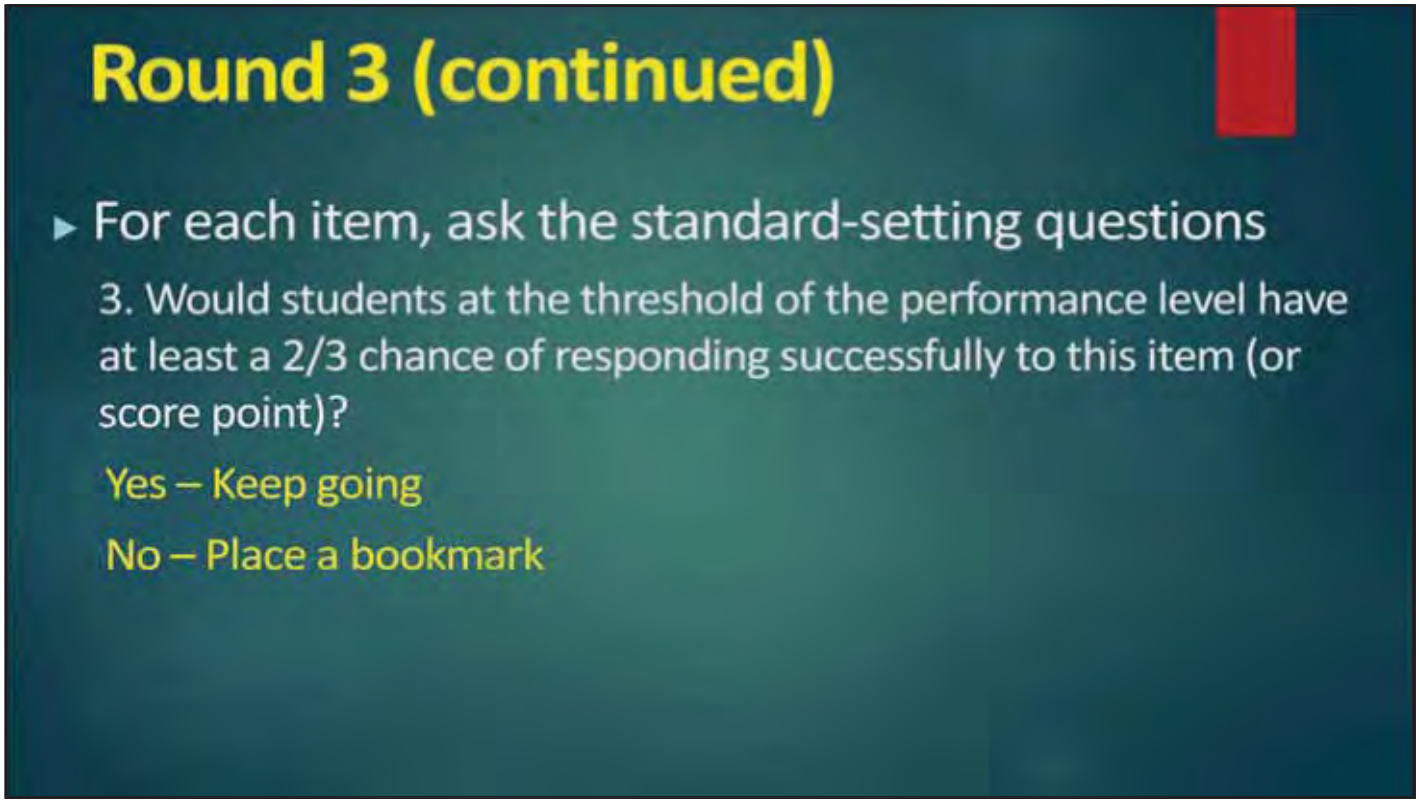


After you have taken a look at the results from Round 2, you will want to discuss these results with your panel. We will help get those discussions started and keep them moving along. We want to make sure everyone has a chance to speak. Topics for discussion include who made changes in cut scores and why, continuing differences in interpretation of PLDs, and response to the impact data.

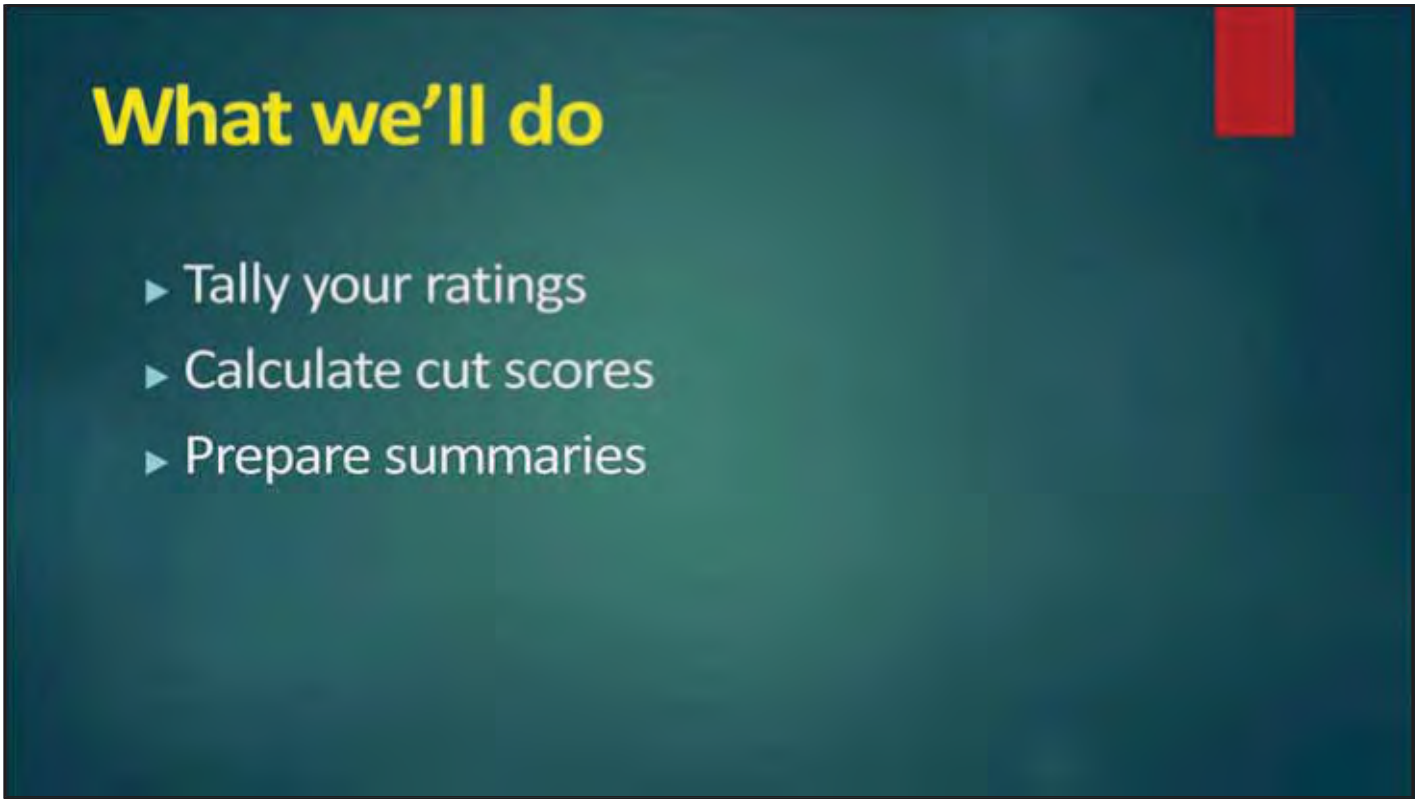
If your judgment is different than the rest of the groups that's okay; we will discuss this and give you the opportunity to provide new recommendations in Round 3. This discussion is important for everyone to check in and understand how others arrived at their judgments.



In the third round of standard setting, you will review the same ordered item booklets that you reviewed in Round 2, and you will ask yourself the same questions [Read questions] Proceed exactly as in Round 2.

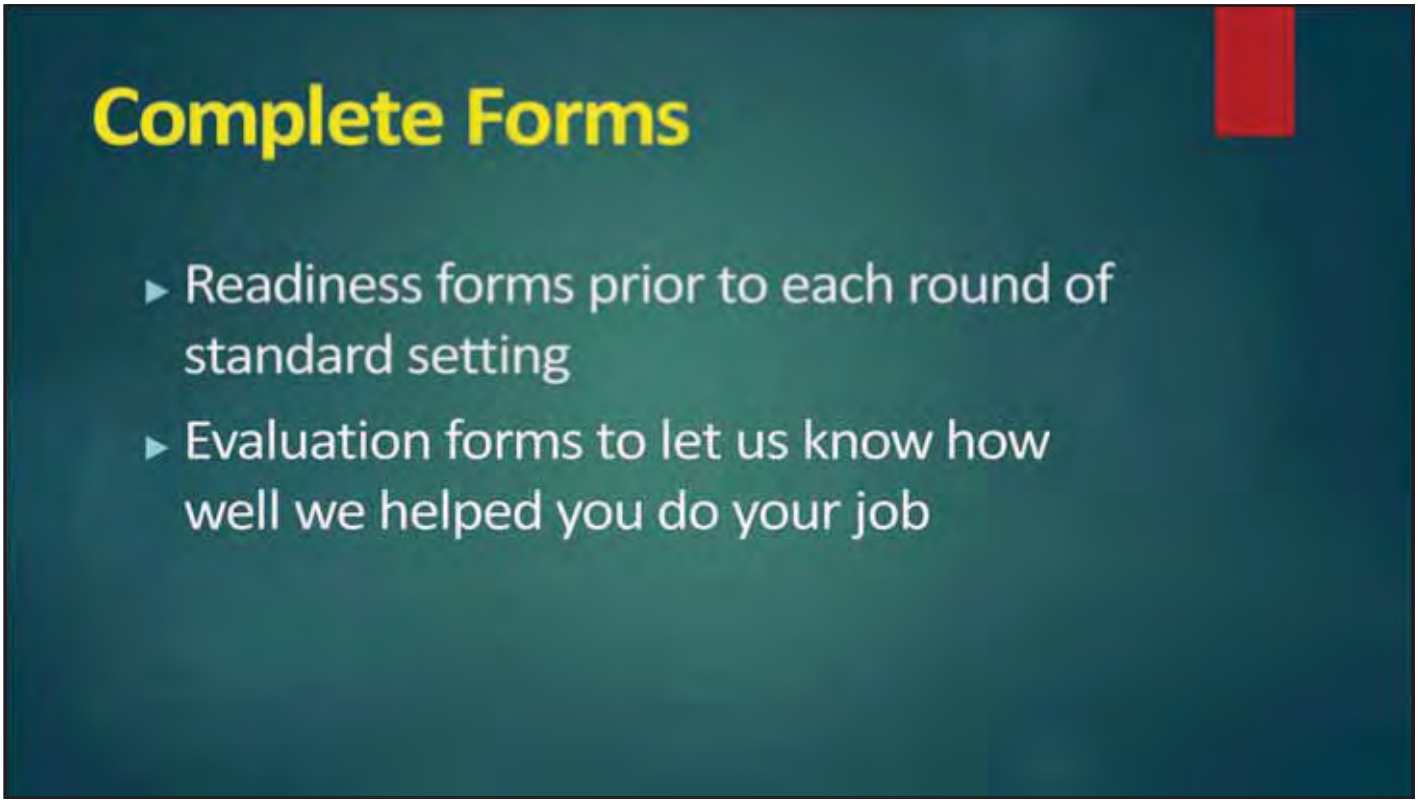


And, of course, you will end with this question. Start with Level 2 (Attained) and ask this question. If the answer is Yes, keep going. If the answer is No, stop and place a bookmark. Afterwards, start thinking about Level 3, asking the same questions.

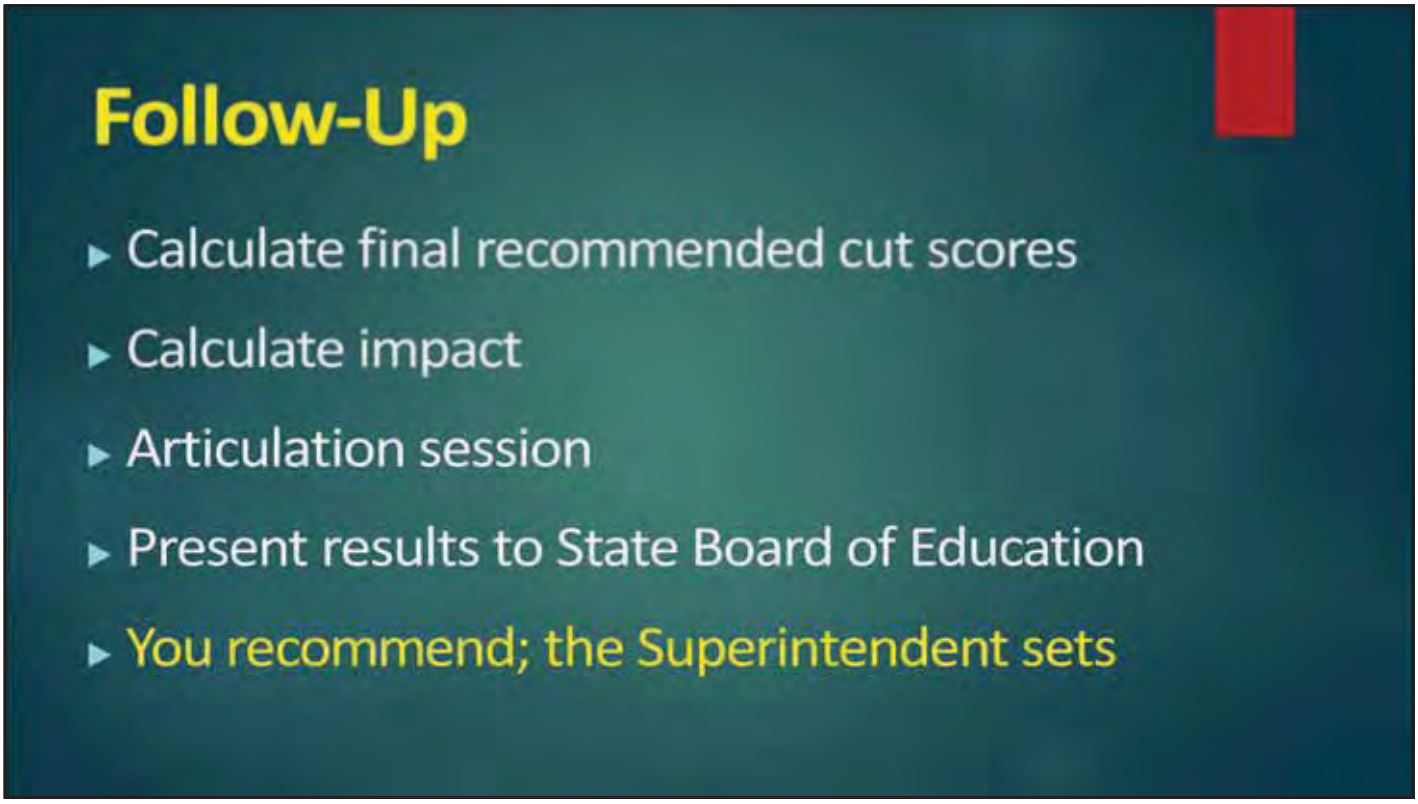


When you finish Round 3, make sure you have completed Round 2 of your rating form, and turn it in to your facilitator. Your facilitators will make sure they have all the materials they are supposed to have from you and give you your next assignment or dismiss you.

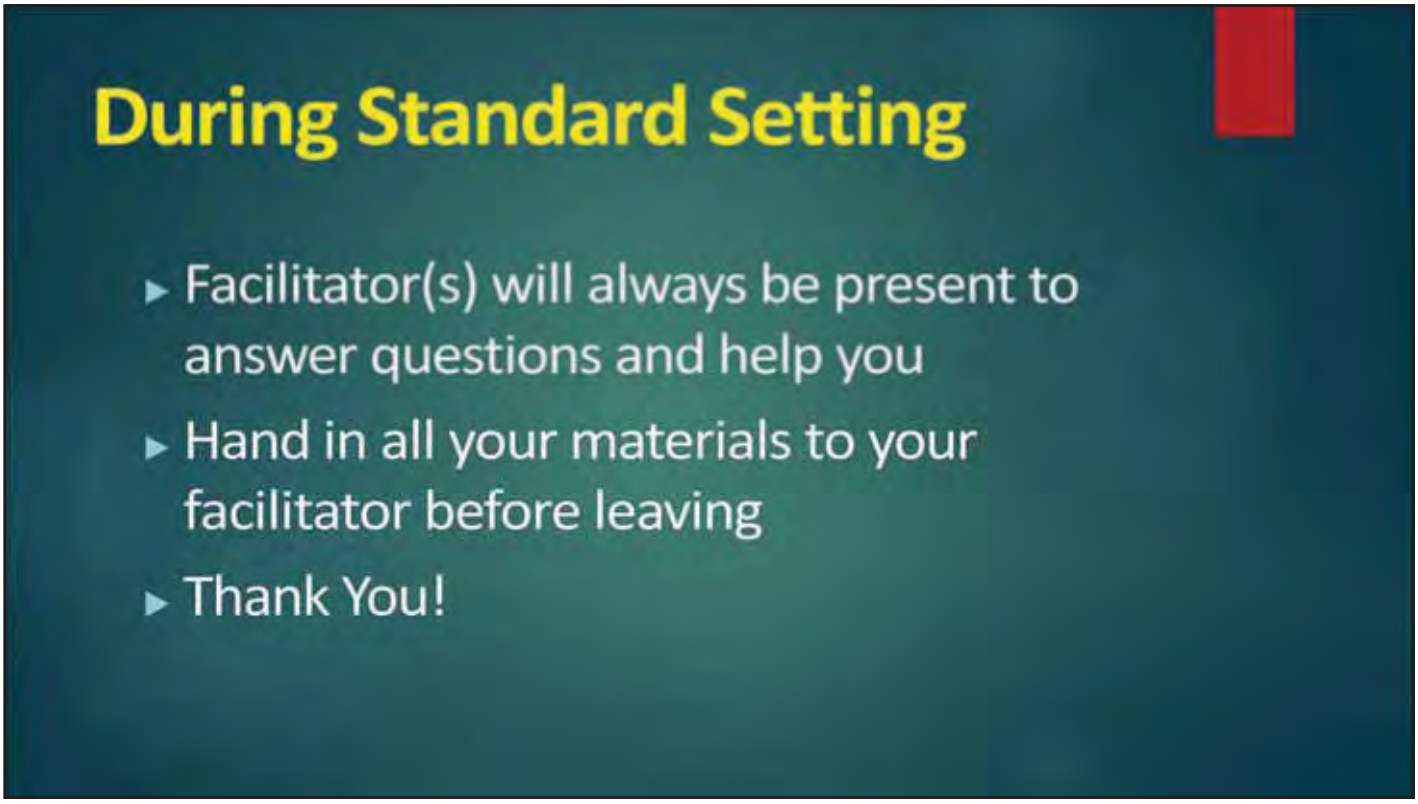
After Round 3, we will once again tally your ratings, calculate cut scores, and create tables and charts for you. We will also prepare tables and charts showing the impact of your Round 3 cut scores on MI-Access students.



We will ask you to provide us with feedback in an evaluation form of what you thought of the whole process. It is very important that you fill out this form and tell us how you arrived at your judgments and give us your opinions. You also get the chance to provide us with feedback throughout the process with different readiness and feedback forms. Our goal is that the process goes as smoothly as possible.

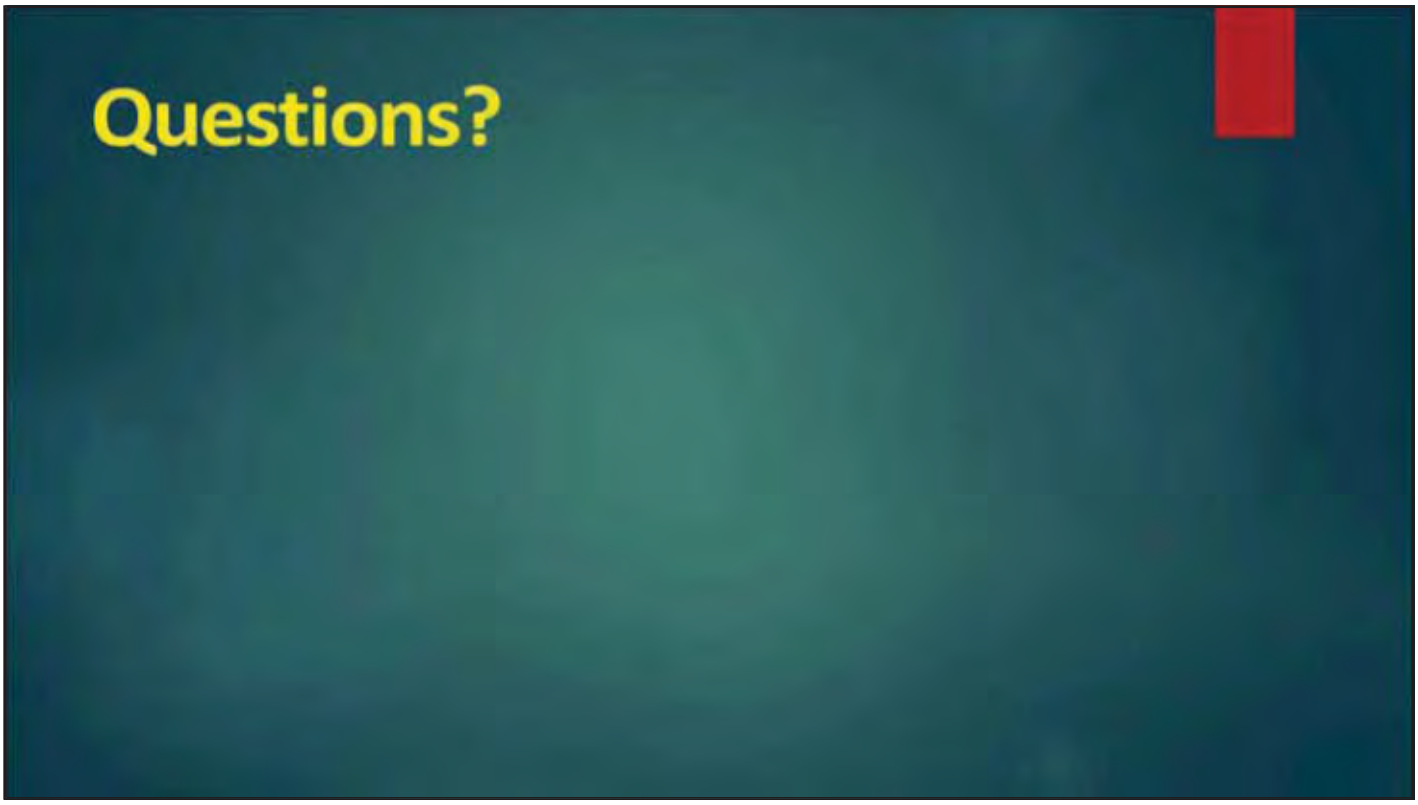


We will tally your cut scores and report the mean, median, and range for level (Attained, Surpassed). These recommendation will be reviewed during an articulation session before sending your recommendations along to the State Board of Education. We have noted before, but it bears repeating, that what we are doing here this week is establishing **recommended** cut scores. The State Board of Education has final responsibility and authority for actually setting those cut scores. Once the Board has taken action, we can enter the final cut scores into our score report programs and produce the score reports for your students.



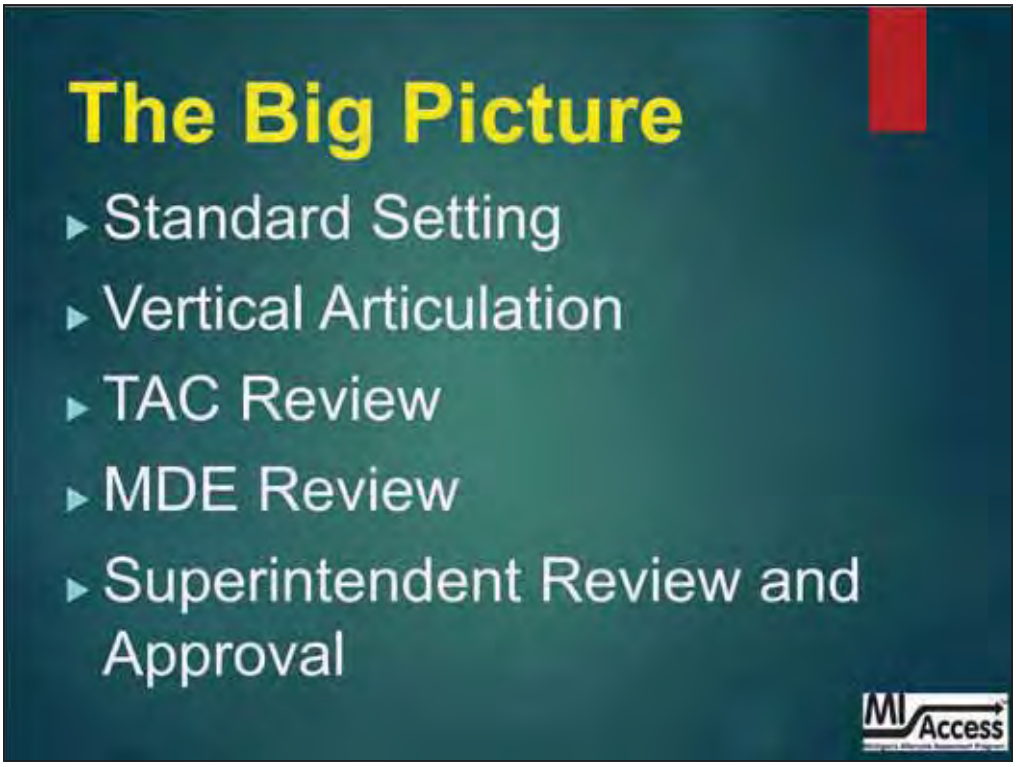
Some final notes….Facilitators and MDE staff will also be on hand to help you and answer your questions. MI staff will help guide you through each of the steps of standard setting. Once you finish any activity, you should make sure to hand in your materials to MI staff that will be in your room.

Most importantly, we hope you enjoy the standard setting process and we greatly appreciate your time and your willingness to participate.



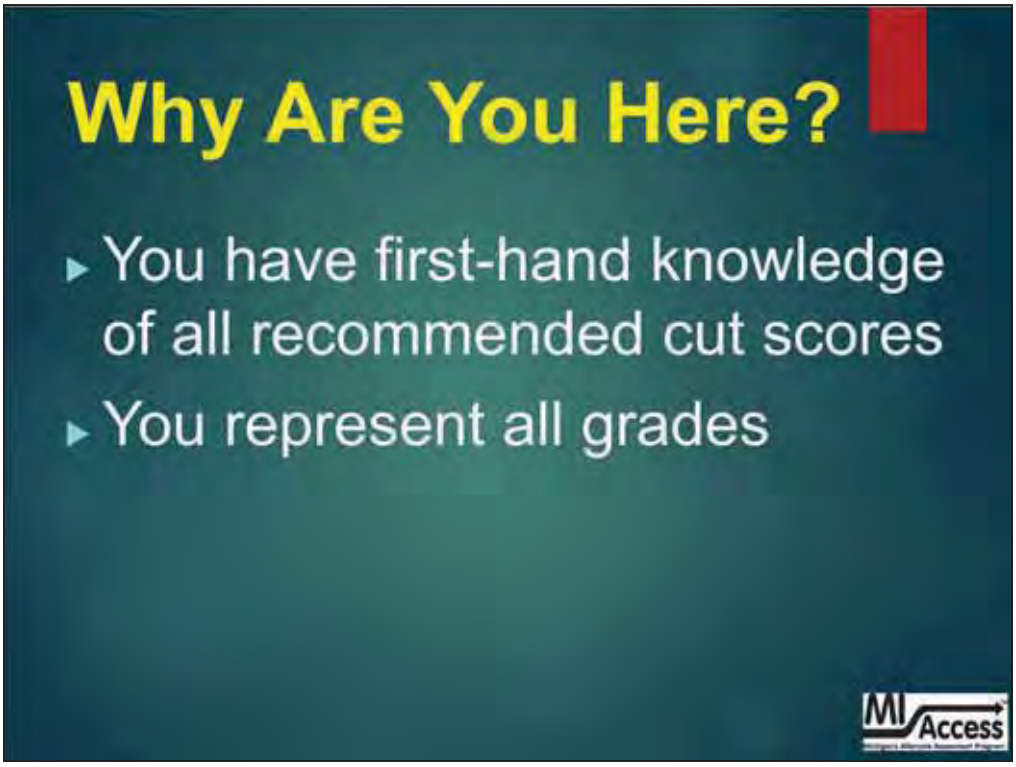
Do you have any questions about any part of the presentation before we tell you about your room assignments?
[Answer any procedural questions. For policy questions, defer to MDE. For subject or grade-band-specific questions, defer to facilitators, who will answer them once they have been dismissed from the large-group session.]

MI-Access
Vertical Articulation
MICHIGAN Department of Education
MI Access
Michigan's Alternate Assessment Program

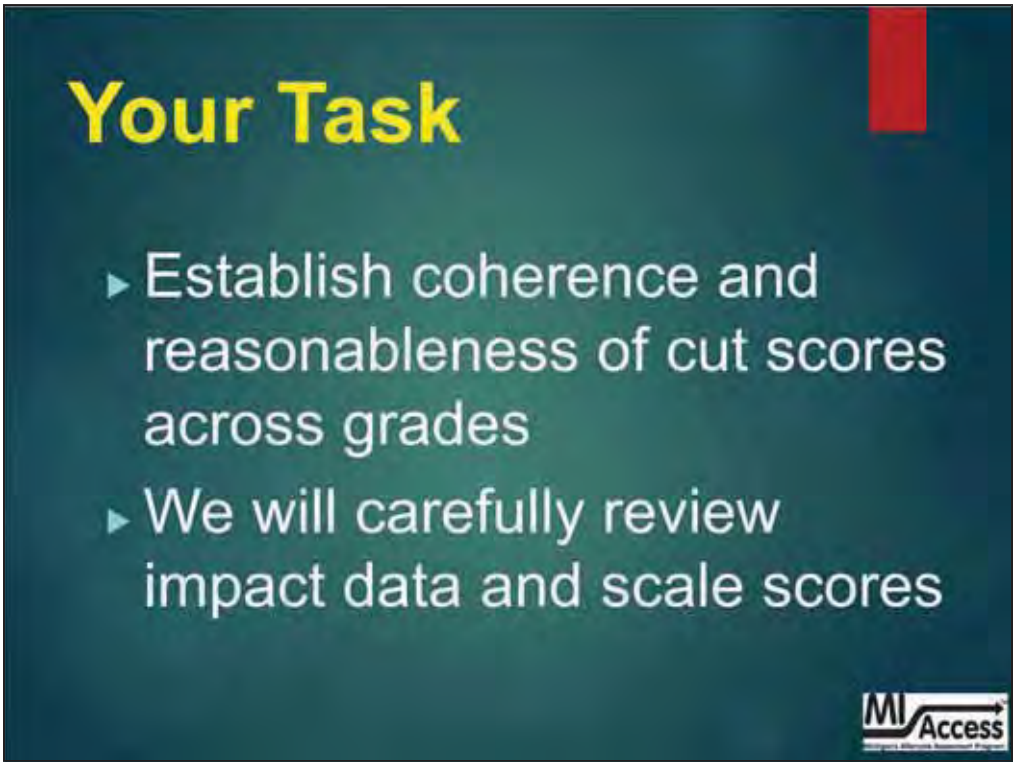


For the past three days, you have been involved in standard setting for a single grade or grade band. That's part of a larger enterprise of setting cut scores for all grades in such a way that when a superintendent or a school board or the general public looks at the full impact across grades, it seems reasonable. Therefore, now that we have recommended cut scores grade by grade, we want to take a more panoramic view of the results and see if we want to make any adjustments. Just as you discussed your individual cut scores with others at your table and then with others in your room and made certain adjustments, we can now expand that conversation to include panelists in other rooms considering other grades.

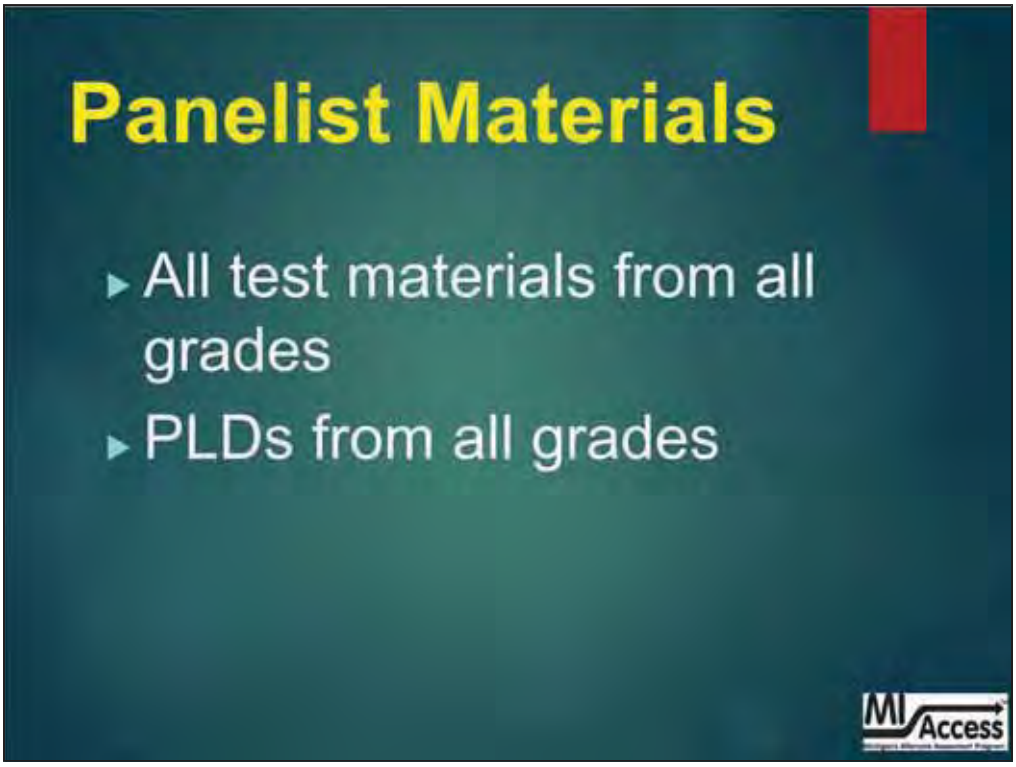
After we finish our work here today, the process still won't be finished. The technical advisory committee will examine what we did here this week to make sure we followed generally accepted practice and will make their recommendation to the Department as to whether or not any further adjustments are in order. The Department will then review all input and may make additional adjustments before making final recommendations to the State Board of Education. The Board, as I have mentioned previously this week, has the final authority to accept, modify, or reject the cut scores recommended to them.



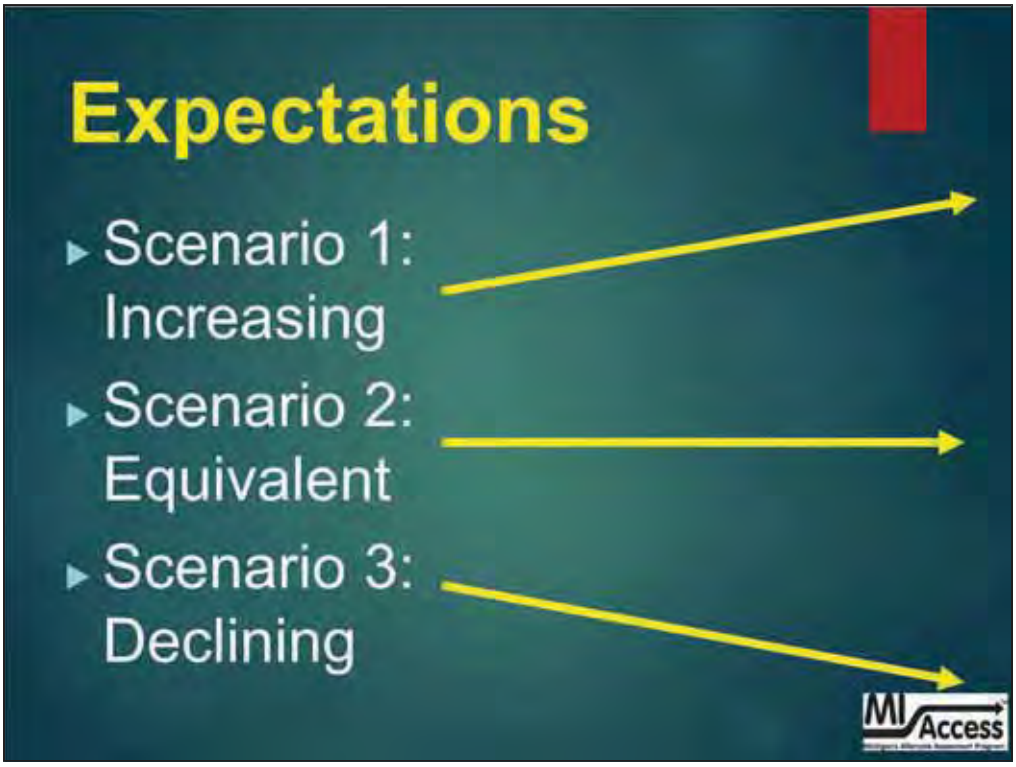
You were selected because you know the students and the tests. You also know what went on in the various breakout rooms over the past three days. We need your advice as we look over all the recommendations to see if there are any conflicts we need to resolve. I will explain momentarily what I mean by "conflict." the bottom line is that you are here because you are the most knowledgeable people in the state with regard to these students, these tests, and these recommendations.



In a few minutes, we are going to look at a chart showing all cut scores and their impacts on all grades. Your task is to tell us whether the results we have obtained so far are generally reasonable when we consider all grades or if we should make some changes.



During this activity, you will have access to all the materials you used over the past three days plus all the materials the other panels used.

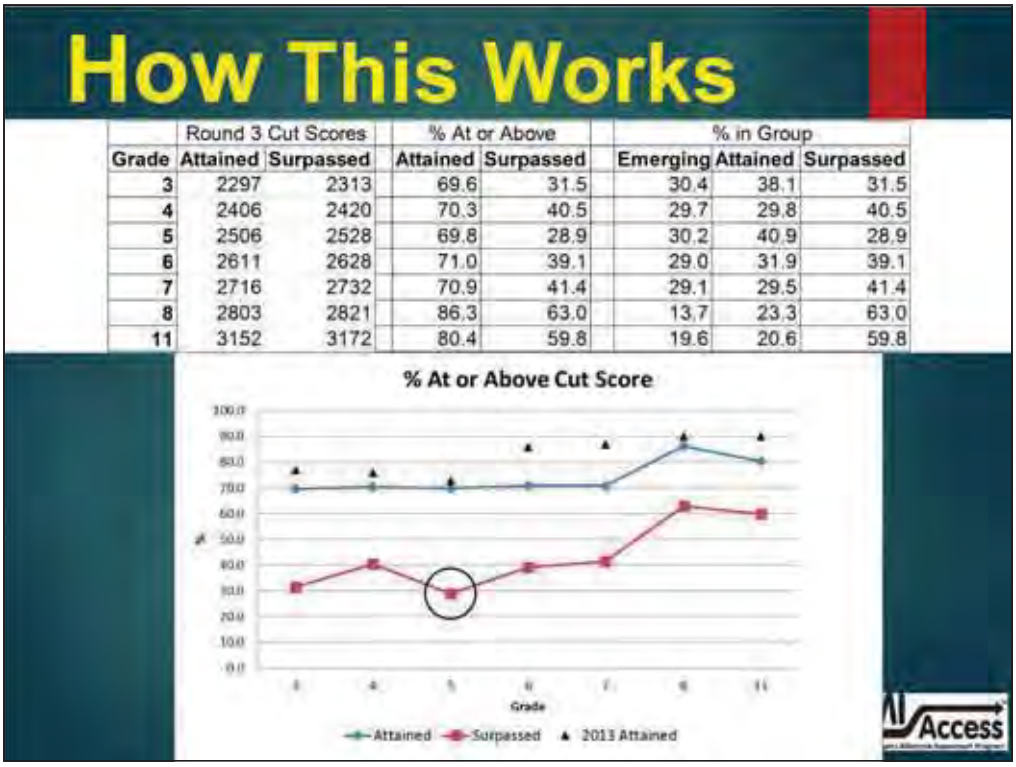


When I speak of the reasonableness of cut scores and their associated impacts, I am talking about expectations. Think about how schools and systems perform over time. There are typically three scenarios in terms of students at or above some cut score:

1. Student performance increases over time – each year, students do a little better than they did the year before; in any given year, a larger percentage of 8th graders attain the standards than they did as 7h graders the year before.
2. Student performance is fairly stable over time – each year looks pretty much like the year before, and within any given year, 8th graders, 7th graders, 6th graders and so on all achieve at about the same level.
3. Then there's generally declining performance – each year, students perform a little less well than they did the year before; 8th graders don't do quite as well as 7th graders, and 7th graders don't do quite as well as 6th graders.

There are variations on these three scenarios; for example, generally increasing except for 6th grade, which in our district is the first year of middle school, and everyone knows what happens then. Then there's generally declining except for 8th grade where we have our reading enrichment program. These tend to be school- or district-specific phenomena, and we are going to be looking at trends for over half the country.

What we don't expect is something like this – where the percentage of 4th graders at Level 2, for example, is much higher than that of 3rd graders. But then at 5th grade, they drop off again, only to rise at 6th and 7th grades, drop off again at 8th grade and high school. How would you explain this to parents? "Your 4th grader did quite well this year, but the likelihood that he or she will do well again next year is not so good."

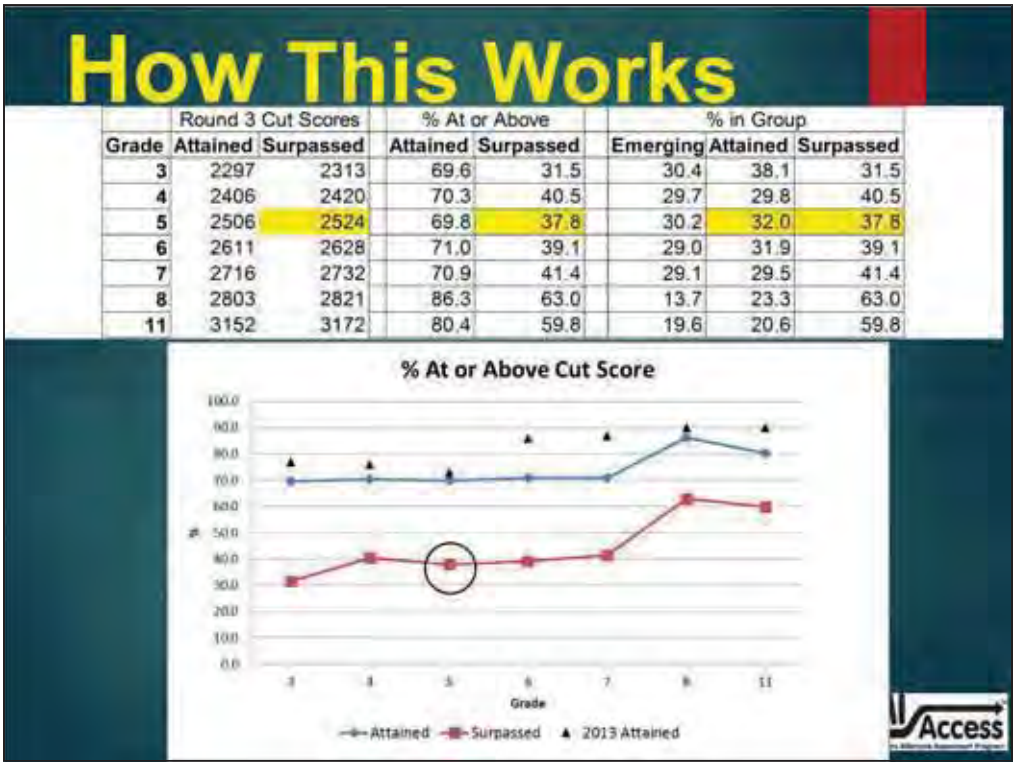


When we break into two groups – one for ELA and one for Math – you will see a set of tables and a graph like this. The first table will show all the final cut scores expressed in scale scores. The second table will show the percentages of students at or above Level 2 and Level 1, by grade, given the cut scores in the first table. This same information is also depicted in the down below. The third table will show the percentages of students in each of the three levels, based on the cut scores in the first table.

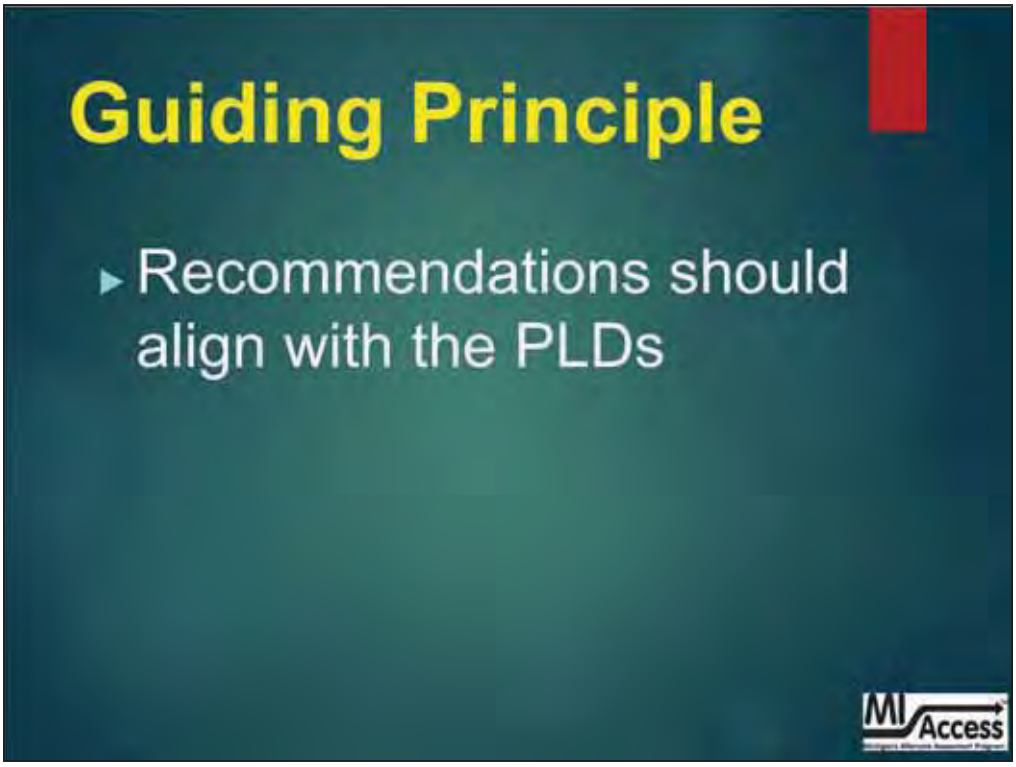
As we consider the information in these tables and this graph, we will be able to recommend a change in any cut score. When we change any cut score in the first table, the other tables and the graph automatically update. Let's work through an example.

The graph you see here generally conforms to the second scenario I mentioned earlier – generally flat. All grades show about 50% of students scoring at or above Level 2, except for 7th grade. Why are the results for grade 7 so different from all the others? Notice that in the first table, the Level 2 cut score is 3050. If we look at grades 5, 6, and 8, we see a much larger jump from grade 6 to grade 7 than for grade 5 to grade 6 and a very small jump from grade 7 to grade 8. Is there something about 7th graders, the test, the items that would make these differences seem reasonable?

In this situation, we would go back to the [tests/work samples – depending on procedure] and let everyone take a look. At some point, someone may suggest lowering the cut score for grade 7 to make the progression from grade 5 to grade 6 to grade 7 to grade 8 more even. Eventually, someone suggests moving the Grade 7 Level 2 cut score to 2965. Here's what happens…



Here, we have changed the Grade 7 Level 2 cut score from 3050 to 2965. You will notice that the percentages in the second and third tables have changed, and that the dip in the Level 2 graph has been reduced. It is not entirely gone, however. That was not really our objective. Our objective was to reset the cut score where it would seem more reasonable.

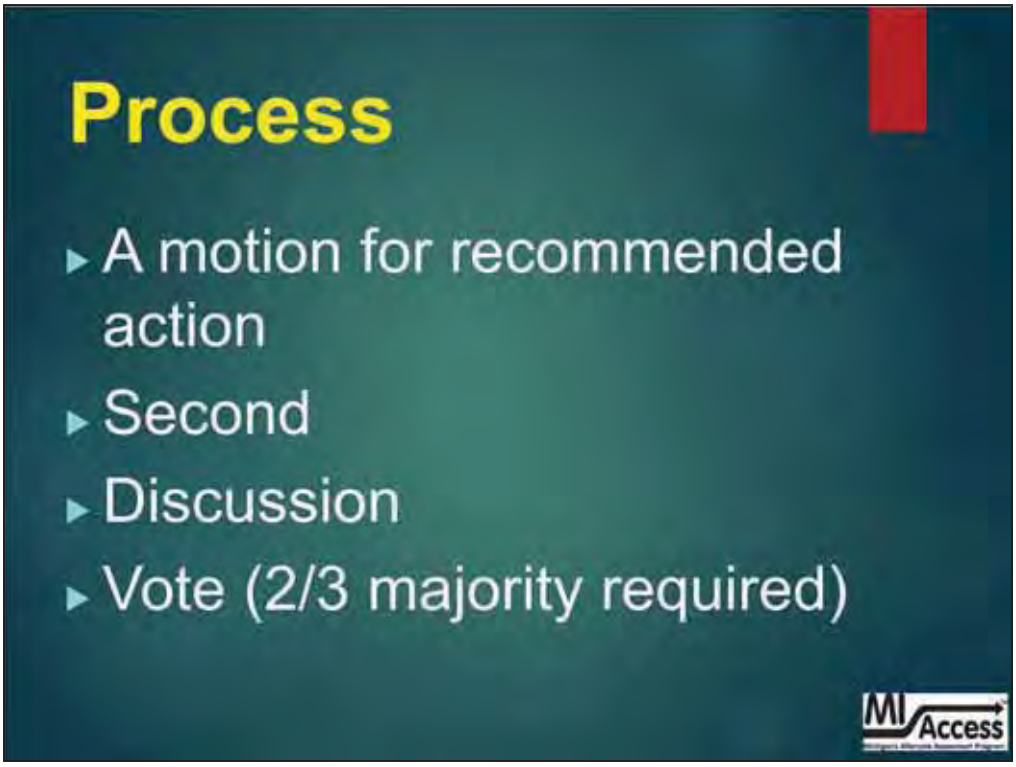


The scale scores for these tests are set up so that there is a general increase from grade 3 to grade 4, and so on up through grade 11. If the scaled cut score for Level 2 for grade 6 is higher than the cut score for Level 2 for grade 7, we may want to look into that. If the percentages of students at Level 2 and above varies significantly from grade to grade, we may also want to look into that. I'm not saying that neither of these two things should ever happen; I'm just saying they would be unexpected. We would first want to find out if the departures from expectation are justified or can be explained. If we can't explain them, then we consider changing something.

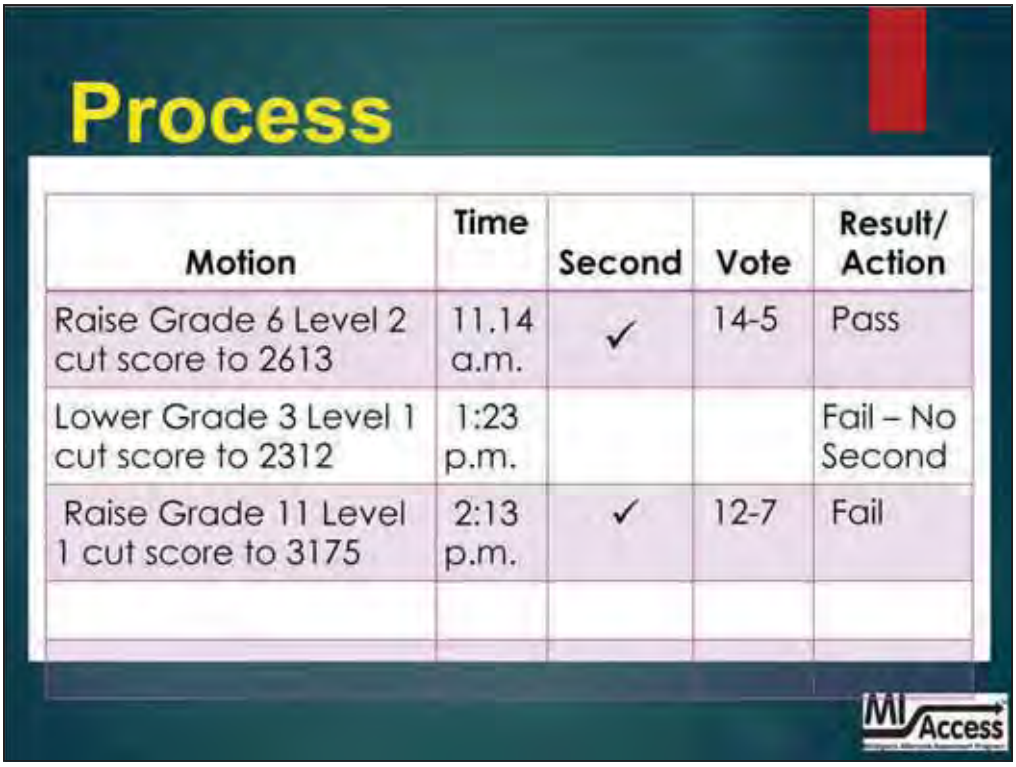
However, we don't want to move cut scores just to make the lines smooth. Every recommendation for a cut score change should be grounded in the PLDs

BoW Version: In changing a cut score, we are saying that a work sample that we previously thought was in one level is now in another level. Can you look at this work sample and the PLD and honestly say that this work sample belongs in this level? If the answer is Yes, then the change is justified. If the answer is No, the change is not justified. We must either accept the original cut score or find another substitute.

Bookmark Version: To change a cut score, we have to move our collective bookmark. Before accepting the change, we need to look at the page that most closely aligns with the new recommended cut score and ask ourselves if the item on the preceding page (remember what the bookmark signifies) aligns with the PLD for that level. If the answer is Yes, then the change is justified. If the answer is No, the change is not justified. We must either accept the original cut score or find another substitute.



We're going to do this democratically. To change any cut score, we will need a motion and a second. After the second, there will be discussion, not before. At some point, Craig or I will call for the vote. Since we are essentially overriding someone else's decision, we will need a 2/3 majority to pass the motion. We will have someone in each room recording the vote count and noting whether or not the motion passed. At the end, we will ask for a motion to accept all cut scores, changed plus remaining unchanged, and follow the same rules.



Here's an example.

The first motion passed by a vote of.

The second motion failed because a majority voted against it.

The third motion also failed. Even though a majority voted for it, it did not get a 2/3 majority.

We're going to do this democratically. To change any cut score, we will need a motion and a second. After the second, there will be discussion, not before. At some point, Craig or I will call for the vote. Since we are essentially overriding someone else's decision, we will need a 2/3 majority to pass the motion. We will have someone in each room recording the vote count and noting whether or not the motion passed. At the end, we will ask for a motion to accept all cut scores, changed plus remaining unchanged, and follow the same rules.

Appendix E.2 MI-Access Standard Setting (2017)

MI-Access Standard Setting
Participation and Supported Independence (June 19-22, 2017)
Functional Independence (July 10-12, 2017)

Measurement Incorporated
July 24, 2017

Table of Contents

List of Tables and Figures 3
Executive Summary 5
  Conclusions and Recommendations 9
Introduction 10
Planning and Implementation 11
  Participation and supported Independence 11
    Bodies of work 11
    Training materials 15
    Agenda 15
    Conduct of the meeting 15
    MDE preview 17
    Vertical articulation 17
  Functional Independence 18
    Training materials 18
    Agenda 19
    Conduct of the meeting 20
    MDE preview 21
    Vertical articulation 22
Results 23
  Round by Round Results 23
  Vertical Articulation 31
  Evaluation 34
    Sample comments 37
Conclusions and Recommendations 37
References 38

Appendix A Training Materials 39
Appendix B Evaluation Comments 63
Appendix C PowerPoint Presentations 71

List of Tables and Figures

Table #	Title	Page
ES-1	Standard Setting Panels	5
ES-2	Participation/Supported Independence Agenda	6
ES-3	Functional Independence Agenda	6
ES-4	Final Cut Scores and Impact	8
1	Standard Setting Panels	10
2	Detailed Agenda: Participation and Supported Independence	16
3	Detailed Agenda: Functional Independence	19
4	Round 1 Results for Participation	23
5	Round 2 Results for Participation	24
6	Round 1 Results for Supported Independence	25
7	Round 2 Results for Supported Independence	26
8	Round 1 Results for Functional Independence	27
9	Round 2 Results for Functional Independence	28
10	Round 3 Results for Functional Independence	29
11	Cut Scores and Impact for Participation – After Vertical Articulation	31
12	Cut Scores and Impact for Supported Independence – After Vertical Articulation	32
13	Cut Scores and Impact for Functional Independence – After Vertical Articulation	32
14	Evaluation Results for Participation	35
15	Evaluation Results for Supported Independence	35
16	Evaluation Results for Functional Independence	36
17	Evaluation Results for Vertical Articulation	36

Figure #	Title	Page
1	MI-Access Participation Scoring Document	12
2	MI-Access Supported Independence Scoring Document	12
3	Sample Body of Work	13
4	Body of Work Data Entry Sheet	14
5	Bookmark Rating Form	20
6	Final Round Impact: Participation	29
7	Final Round Impact: Supported Independence	30
8	Final Round Impact: Functional Independence	30
9	Impact After Vertical Articulation: Participation	33
10	Impact After Vertical Articulation: Supported Independence	33
11	Impact After Vertical Articulation: Functional Independence	34

Executive Summary

Measurement Incorporated (MI) assisted the Michigan Department of Education (MDE) in the conduct of standard setting for MI-Access Participation, Supported Independence, and Functional Independence (FI) for grades 3-8 plus high school, providing a lead facilitator, panel facilitators, and sufficient psychometric and clerical staff to conduct meetings. Participation and Supported Independence (P and SI) panels met the week of June 19-22, 2017, and the Functional Independence (FI) the week of July 10-12, 2017.

For all MI-Access assessments, the MDE provides three performance levels:

1 – Emerging toward the standard
2 – Attained the standard
3 – Surpassed the standard

MI staff conducted a Body of Work standard-setting procedure for the 8 P/SI panels with two rounds of rangefinding and no pinpointing, and a Bookmark procedure for the 8 FI panels, as recommended by the Technical Advisory Committee (TAC). Panels are described in Table ES-1. The agendas are summarized in Tables ES-2 and ES-3.

Table ES-1 Standard Setting Panels

Participation/Supported Independence		**Functional Independence**	
Panel	**Members**	**Panel**	**Members**
P/SI ELA 3-4	8	FI ELA 3-4	8
P/SI ELA 5-6	8	FI ELA 5-6	10
P/SI ELA 7-8	8	FI ELA 7-8	10
P/SI ELA High School	8	FI ELA High School	9
P/SI Math 3-4	8	FI Math 3-4	9
P/SI Math 5-6	8	FI Math 5-6	10
P/SI Math 7-8	8	FI Math 7-8	10
P/SI Math High School	8	FI Math High School	10

Table ES-2

Participation/Supported Independence Agenda

Date	A.M.	P.M.
June 19	General Training	Body of Work Training; Participation Round 1
June 20	Participation Round 2	Participation Vertical Articulation Training; Vertical Articulation; MDE Preview
June 21	Supported Independence Round 1	Supported Independence Round 2
June 22	Supported Independence Vertical Articulation Training; Vertical Articulation; MDE Preview	

Table ES-3

Functional Independence Agenda

Date	A.M.	P.M.
July 10	General Training	Bookmark Training; Begin Round 1
July 11	Complete Round 1	Review Round 1; Complete Round 2
July 12	Review Round 2; Complete Round 3	Vertical Articulation/Policy Brief

For each set of meetings, panelists received general instruction in the purpose of the meeting, followed by specific instruction on the tests and the Performance Level Descriptors (PLDs). Instruction on the tests included review of tests and manuals and a demonstration by MDE staff (through live role-play for P/SI and PowerPoint with links to the MDE website for FI) of administration procedures. PLD review consisted of facilitator-led discussion of PLDs with questions and answers.

MI staff provided an overview of the standard-setting procedure (Body of Work for P/SI and Bookmark for FI), followed by a short practice round to give all panelists an opportunity to practice the method before applying it. After a brief question-and-answer session, panelists indicated their readiness to proceed with Round 1. In Round 1, P/SI panelists reviewed ordered work samples, while FI panelists reviewed ordered item booklets (OIBs), lower grade first, higher grade second. The task for P/SI panelists was to sort 30 student work samples into three categories: Emerging, Attained, or Surpassed. For FI panelists, the task was to identify two pages in each OIB that would indicate the beginning of the Attained and Surpassed score regions, entering their ratings on scannable documents.

After Round 1, MI staff collected the scannable documents, scanned them, and calculated preliminary cut scores. Facilitators then shared Round 1 results with panelists, including the distribution of panelists' ratings or bookmarks and their individual and group cut scores. P/SI panelists also reviewed impact data – percentages of students classified at each level, based on the Round 1 cut scores – after Round 1. After a discussion of the results, panelists indicated their readiness for Round 2 and commenced, as in Round 1. At the end of Round 2, MI facilitators collected all materials and dismissed the panels. MI psychometricians then analyzed Round 2 data. For FI, facilitators shared Round 2 results with impact data. After discussion of these results, FI panelists completed Round 3 as they had done Rounds 1 and 2. MI staff then calculated final cut scores and impact.

After two rounds of ratings for P/SI panelists and three rounds of bookmark placements for FI panelists, the groups were reorganized into three large committees:

- ELA Vertical Articulation Committee
- Math Vertical Articulation Committee
- MDE Preview

The MDE preview sessions occurred at the same time as the vertical articulations. The purpose of the preview session was to discuss administration and scoring procedures, present information regarding forthcoming changes to the Science assessments, and obtain feedback from Michigan educators.

Vertical articulation committees were made up of representatives of each grade-level panel. Each facilitator appointed approximately half his or her panelists to the VAC and the other half to the MDE preview session. Because VACs for Participation and Supported Independence were conducted on separate days, those panelists who served on a VAC on Tuesday (Participation) went to the MDE preview session on Thursday, and *vice versa*. For FI, there was a single meeting of the VAC.

Vertical articulation began with an overview of the process, followed by a question-and-answer period. After the question-and-answer session, ELA and Math VACs separated into different rooms. During the remainder of the session, each VAC reviewed results (cut scores and impact from Round 2 for P/SI and from Round 3 for FI) across grade spans and recommended changes. To change any cut score, it was necessary to have a motion, second, discussion, and vote. Given that the changes were to override decisions made over two rounds of deliberation, a 2/3 majority was required to pass any motion.

Final results are presented in Table ES-4. Cut scores for P/SI are expressed in raw score terms, while cut scores for FI are expressed in logits. Changes brought about by vertical articulation are highlighted in yellow.

Table ES-4
Final Cut Scores and Impact

Test	Level 2 Cut	Level 3 Cut	% At Level 1	% At Level 2	% At Level 3
P ELA Grade 3	31	45	45.2	27.7	27.2
P ELA Grade 4	32	43	40.8	25.9	33.3
P ELA Grade 5	28	42	38.8	26.5	34.7
P ELA Grade 6	29	41	37.9	25.6	36.5
P ELA Grade 7	28	45	40.5	35.3	24.2
P ELA Grade 8	27	43	46.1	31.8	22.1
P ELA High School	34	46	38.4	26.4	35.2
P Math Grade 3	33	47	49.8	24.7	25.6
P Math Grade 4	32	47	47.5	32.3	20.2
P Math Grade 5	32	46	49.1	30.9	20.0
P Math Grade 6	31	44	46.7	26.8	26.5
P Math Grade 7	27	43	38.9	31.4	29.7
P Math Grade 8	28	43	39.4	31.7	29.0
P Math High School	31	46	40.9	30.5	28.6
SI ELA Grade 3	28	43	20.5	33.4	46.1
SI ELA Grade 4	31	44	20.6	26.7	52.8
SI ELA Grade 5	30	46	17.8	34.6	47.5
SI ELA Grade 6	31	46	16.0	30.8	53.2
SI ELA Grade 7	31	46	17.4	32.3	50.3
SI ELA Grade 8	33	45	20.2	23.8	56.0
SI ELA High School	35	46	30.5	23.6	46.0
SI Math Grade 3	35	47	45.7	26.4	27.9
SI Math Grade 4	34	45	31.4	29.3	39.3
SI Math Grade 5	31	46	27.3	35.1	37.7
SI Math Grade 6	32	44	37.3	30.6	32.1
SI Math Grade 7	30	45	29.1	44.5	26.4
SI Math Grade 8	30	46	23.9	41.1	35.0
SI Math High School	33	47	26.6	34.5	39.0
FI ELA Grade 3	0.525	1.65	26.8	38.6	34.6
FI ELA Grade 4	0.338	1.70	13.8	35.0	51.2
FI ELA Grade 5	0.384	1.53	13.9	28.4	57.7
FI ELA Grade 6	0.636	1.70	18.8	28.1	53.1

Test	Level 2 Cut	Level 3 Cut	% At Level 1	% At Level 2	% At Level 3
FI ELA Grade 7	0.098	0.96	9.0	15.8	75.2
FI ELA Grade 8	0.589	1.38	14.2	16.5	69.3
FI ELA High School	0.233	1.05	11.9	14.2	73.9
FI Math Grade 3	0.584	2.067	34.6	34.0	31.4
FI Math Grade 4	0.444	1.363	24.1	29.7	46.2
FI Math Grade 5	0.87	2.022	34.4	32.8	32.8
FI Math Grade 6	.517	1.351	38.3	32.6	29.1
FI Math Grade 7	0.199	1.404	38.8	35.0	26.2
FI Math Grade 8	0.367	1.39	29.7	34.8	35.5
FI Math High School	0.095	1.074	27.8	34.2	38.0

Panelists evaluated the process and their facilitators on eight critical-incident factors, each on a 5-point scale (Strongly Agree to Strongly Disagree). With regard to facilitators and process, 97-100 percent of panelists agreed with each statement. With regard to the final cut scores, well over 90 percent agreed with their accuracy and fairness. With regard to facilities and food, reaction was mixed, with 56 percent of agreeing that the facilities and food service helped to create a good working environment.

Conclusion and Recommendation

The process for arriving at cut scores was rigorous, consistent with best practices, conducted by highly competent practitioners, and monitored by a highly qualified outside observer. Panelists had high praise for the facilitators and expressed great confidence in the validity of the cut scores their panels set. The resulting cut scores and corresponding impacts were reasonably consistent across grades as well as with historical trends in Michigan for these populations. It is our recommendation that the cut scores be adopted without modification or adjustment.

Introduction

Measurement Incorporated (MI) assisted the Michigan Department of Education (MDE) in the conduct of standard setting for MI-Access Functional Independence (FI), Supported Independence (SI), and Participation (P) for grades 3-8 plus high school. Specifically, MI provided a lead facilitator, panel facilitators, and sufficient, psychometric, and clerical staff to conduct eight panel meetings the week of June 19-22, 2017, and eight panel meetings the week of July 10-12, 2017.

For all MI-Access assessments, the MDE provides three performance levels:

1 – Emerging toward the standard
2 – Attained the standard
3 – Surpassed the standard

For the P and SI standard-setting activities, MI staff conducted a Body of Work procedure with two rounds of rangefinding and no pinpointing for eight panels, as recommended by the Technical Advisory Committee (TAC). For the FI standard setting activity, MI staff conducted a Bookmark procedure with three rounds of bookmark placements, as recommended by the TAC. Panels are described in Table 1. Their demographic characteristics are summarized in Tables 2 and 3.

Table 1
Standard Setting Panels

Participation/Supported Independence		Functional Independence	
Panel	**Members**	**Panel**	**Members**
P/SI ELA 3-4	8	FI ELA 3-4	8
P/SI ELA 5-6	8	FI ELA 5-6	10
P/SI ELA 7-8	8	FI ELA 7-8	10
P/SI ELA High School	8	FI ELA High School	9
P/SI Math 3-4	8	FI Math 3-4	9
P/SI Math 5-6	8	FI Math 5-6	10
P/SI Math 7-8	8	FI Math 7-8	10
P/SI Math High School	8	FI Math High School	10

Planning and Implementation

MI submitted a detailed plan to MDE and modified it in response to comments from the TAC. The final version of the plan called for a four-day meeting for P/SI the week of June 19-22, 2017 for Participation and Supported Independence, and a three-day meeting the week of July 10-12, 2017 for Functional Independence. The plan called for application of a Body of Work procedure (Cizek & Bunch, 2007; Kingston & Tiemann, 2012) for the P and SI event, given that tests were composed primarily of performance tasks, and a Bookmark procedure (Cizek & Bunch, 2007; Lewis, Mitzel, Mercado, & Schulz, 2012) for the FI event, given the fact that the tests were almost entirely selected response and the items were scaled with the Rasch model. Details of the plan and its execution are provided below.

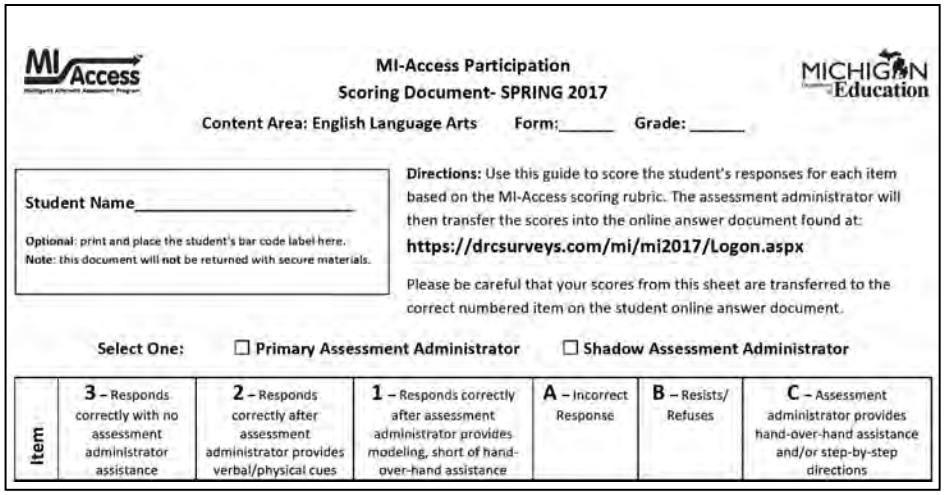
Participation and Supported Independence

The nature of the assessments for P and SI (portfolio) lends itself to the body of work procedure. This procedure requires panelists to sort work samples into categories based on performance levels. Panelists sort a preliminary collection of student work samples, ordered by total score, to identify regions in which cut scores might be located in a process known as rangefinding. After rangefinding, some of the original work samples may be removed and replaced by different work samples with scores within the regions identified during the rangefinding round. This subsequent round is often referred to as pinpointing. One or more pinpoint rounds may be employed. For this activity, there were two rounds of rangefinding and no pinpointing, per instructions from the Technical Advisory Committee (TAC). After the final round of item review, MI conducted a vertical articulation, engaging representatives of each grade level for each content area to examine all cut scores across all grades and recommend changes.

Planning. In planning for this set of panel activities, we made the following assumption:

Teachers score the P and SI assessments as students respond to them; therefore, there would be no need for additional scoring after online and paper documents are collected. However, there was a need for MDE verification of samples of teacher-rendered scores as a validity check. Time for that activity was built into the overall project schedule.

Bodies of work. The MI-Access Participation exam consists of a series of activities administered by a trained educator (primary administrator) with the assistance of a "shadow administrator." For Participation students, each task is scored on a 0-3 scale, as shown in Figure 1. For Supported Independence students, each task is scored on a 0-2 scale, as shown in Figure 2. In all instances, the total score for any student on any item is the sum of the scores entered by the two administrators. Thus, for Participation students, scores for each item can range from 0 to 6, while for Supported Independence students, scores for each item can range from 0 to 4.



MI Access

MI-Access Participation
Scoring Document- SPRING 2017

MICHIGAN Education

Content Area: English Language Arts **Form:_____** **Grade: _____**

Student Name_____

Optional: print and place the student's bar code label here.
Note: this document will **not** be returned with secure materials.

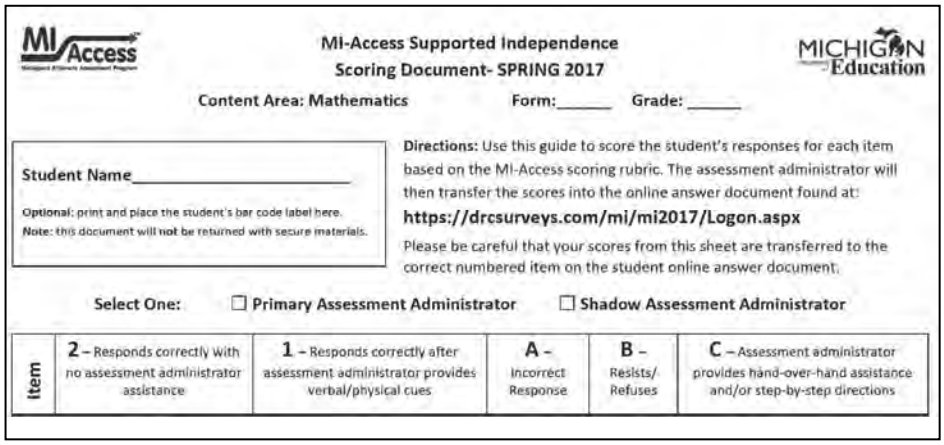
Directions: Use this guide to score the student's responses for each item based on the MI-Access scoring rubric. The assessment administrator will then transfer the scores into the online answer document found at:
https://drcsurveys.com/mi/mi2017/Logon.aspx

Please be careful that your scores from this sheet are transferred to the correct numbered item on the student online answer document.

Select One: ☐ **Primary Assessment Administrator** ☐ **Shadow Assessment Administrator**

Item	**3** – Responds correctly with no assessment administrator assistance	**2** – Responds correctly after assessment administrator provides verbal/physical cues	**1** – Responds correctly after assessment administrator provides modeling, short of hand-over-hand assistance	**A** – Incorrect Response	**B** – Resists/ Refuses	**C** – Assessment administrator provides hand-over-hand assistance and/or step-by-step directions

Figure 1. MI-Access Participation Scoring Document



MI Access

MI-Access Supported Independence
Scoring Document- SPRING 2017

MICHIGAN Education

Content Area: Mathematics **Form:_____** **Grade: _____**

Student Name_____

Optional: print and place the student's bar code label here.
Note: this document will **not** be returned with secure materials.

Directions: Use this guide to score the student's responses for each item based on the MI-Access scoring rubric. The assessment administrator will then transfer the scores into the online answer document found at:
https://drcsurveys.com/mi/mi2017/Logon.aspx

Please be careful that your scores from this sheet are transferred to the correct numbered item on the student online answer document.

Select One: ☐ **Primary Assessment Administrator** ☐ **Shadow Assessment Administrator**

Item	**2** – Responds correctly with no assessment administrator assistance	**1** – Responds correctly after assessment administrator provides verbal/physical cues	**A** – Incorrect Response	**B** – Resists/ Refuses	**C** – Assessment administrator provides hand-over-hand assistance and/or step-by-step directions

Figure 2. MI-Access Supported Independence Scoring Document

MI staff worked closely with MDE staff to identify scored documents from the spring 2017 administration such that scores from 0 to perfect or very nearly so for each grade or grade span were included among the work samples. MI staff then translated those scored documents into worksheets panelists used to evaluate the performance levels of the students whose work was represented thereon. Each worksheet included not only the scores for each item but the total score and the average score for each item. The purpose of the total score was to give panelists a clear indication of the total performance of the student; i.e., the body of work for that

student. The purpose of the average score for each item was to help panelists place item-level performance for each student in a larger context of how students statewide had performed on that item. A sample body of work is shown in Figure 3. After reviewing each work sample, panelists entered their evaluation on a form similar to that shown in Figure 4. The large circles in Figure 4 are scanner alignment marks.

Item Number	Student Score	Average Item Score
1	4	3.15
2	6	3.85
4	0	2.18
5	0	3.44
7	0	3.25
8	0	3.24
10	0	3.02
11	0	3.76
13	4	4.50
14	0	3.80

Work Sample: 1

Total Score: 14

Barcode: 3288674553

Program: MI-Access P

Subject: ELA

Grade: 3

Figure 3. Sample Body of Work

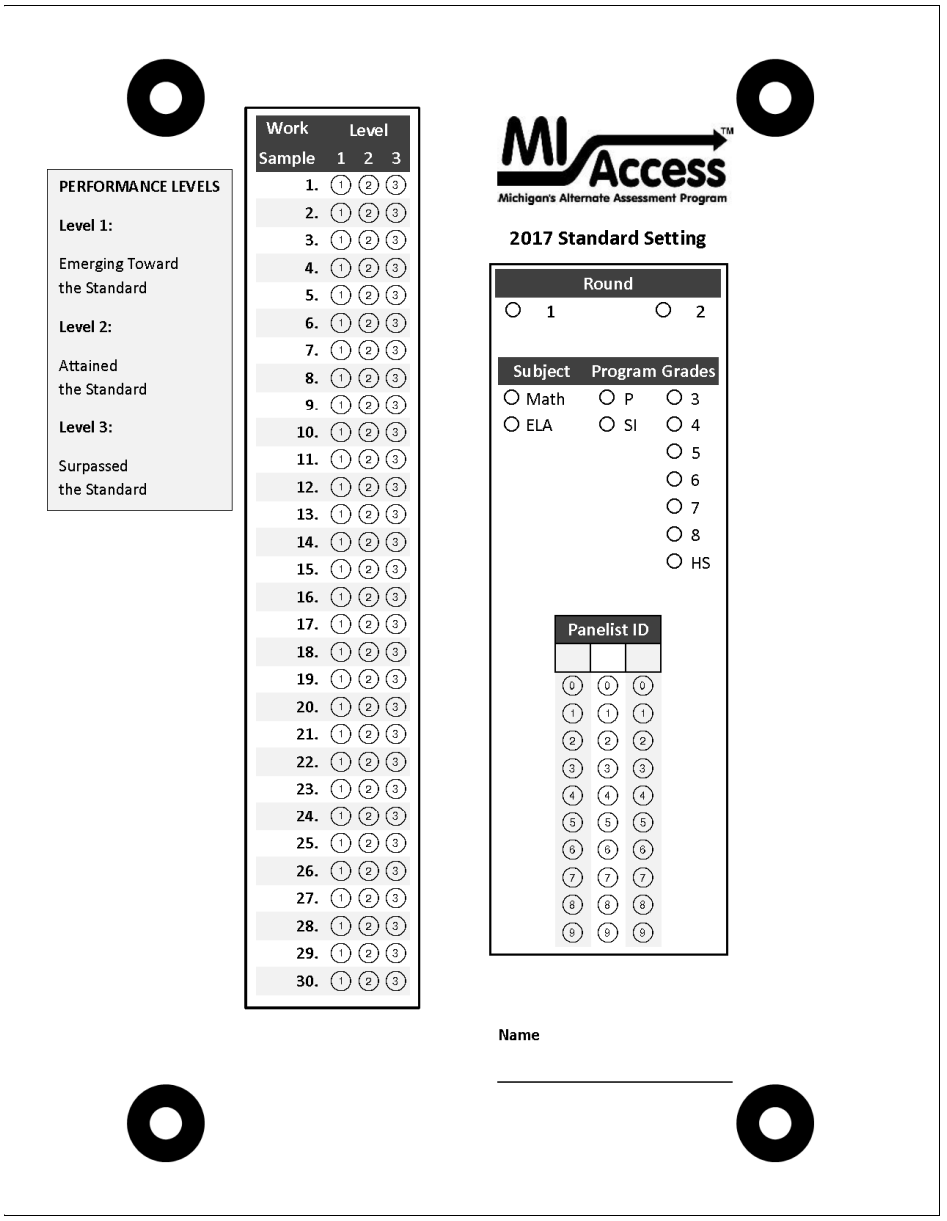


Figure 4. Body of Work Data Entry Sheet

Training materials. MI prepared materials for an opening session that included the goals and tasks of the session as well as a PowerPoint presentation on the body of work procedure. In addition, panel facilitators prepared grade/subject-specific materials that helped panelists understand the nature of the tests and factors affecting performance. Performance level descriptors (PLDs) were developed by MDE test development and curriculum teams. All training materials and forms were submitted to the MDE for review and approval prior to implementation. These materials are listed below and included in full in Appendix A. PowerPoint presentations are included in Appendix C.

- Overview (PowerPoint)
- Body of Work (PowerPoint)
- Facilitator Script
- Body of Work Practice Round Form
- Body of Work Entry Form – Round 1
- Body of Work Entry Form – Round 2
- Readiness Form
- Process Evaluation Form

Agenda. Table 2 (next page) shows the day-by-day agenda for the four-day event for P/SI.

Conduct of the meeting. Dr. Bunch provided an overview of the four days and gave the panels their charge (see Appendix A). In addition, John Jacquith from MDE provided background information with respect to the development, administration, and scoring of the P/SI assessments. Afterwards, panelists dispersed to their breakout rooms to review the tests under the direction of the facilitators. These same facilitators also led the panelists in a review of the PLDs. After lunch on July 10, Dr. Bunch provided an overview of the Body of Work procedure (see Appendix C). Panelists then dispersed to their breakout rooms for Body of Work practice with a small set of P Scoring Documents. Following this practice round, facilitators answered questions, and determined readiness to begin Round 1 by administering and reviewing the Round 1 Readiness Form (see Appendix A). Dr. Bunch and MDE staff circulated among the eight panel rooms throughout each day to observe and answer questions. An external evaluator, Dr. Adam Wyse, also observed the various activities and sat in

Panelists worked in small groups of 3-4 within a room of 7-9. They consulted with others at their table during each round. One panel (Science grades 4 and 7) had two different sets of tests to review (four tests in all); therefore, their schedule was a bit different from those of high school.

Table 2
Detailed Agenda: Participation and Supported Independence

Day/ Session	Panel 1: Math 3-4	Panel 2: Math 5-6	Panel 3: Math 7-8	Panel 4: Math HS	Panel 5: ELA 3-4	Panel 6: ELA 5-6	Panel 7: ELA 7-8	Panel 8: ELA HS
6/19 a.m.	Intro/Training							
	Review of PLDs and tests				Review of PLDs and tests			
6/19 p.m.	P Round 1	P Round 1	P Round 1	P Round 1	P Round 1	P Round 1	P Round 1	P Round 1
6/20 a.m.	Review P Round 1; P Round 2	Review P Round 1; P Round 2	Review P Round 1; P Round 2	Review P Round 1; P Round 2	Review P Round 1; P Round 2	Review P Round 1; P Round 2	Review P Round 1; P Round 2	Review P Round 1; P Round 2
6/20 p.m.	Articulation Training							
	Articulation				Articulation			
6/21 a.m.	Review of PLDs and tests				Review of PLDs and tests			
	SI Round 1	SI Round 1	SI Round 1	SI Round 1	SI Round 1	SI Round 1	SI Round 1	SI Round 1
6/21 p.m.	Review SI Round 1; SI Round 2	Review SI Round 1; SI Round 2	Review SI Round 1; SI Round 2	Review SI Round 1; SI Round 2	Review SI Round 1; SI Round 2	Review SI Round 1; SI Round 2	Review SI Round 1; SI Round 2	Review SI Round 1; SI Round 2
6/22 a.m.	Articulation/MDE Preview							
6/22 p.m.	Articulation; Wrap-Up							

Each panel completed both rounds for P before beginning with SI. Review materials consisted of a packet of 30 completed Scoring Documents (see Figures 1 and 2) arranged from lowest to highest score. Their task was to assign each completed Scoring Document to one of the following three levels using the PLDs:

1 – Emerging 2 – Attained 3 - Surpassed

Panelists were free to discuss any Scoring Document with others at their tables, but the entries they made had to be their own, not that of the table. They entered their ratings on forms similar to that shown in Figure 4.

After Round 1, MI staff analyzed the ratings and identified regions where cut scores might be, using the standard rangefinding procedure associated with Body of Work (cf. Cizek & Bunch, 2007, Ch. 9). In Round 2, panelists rated the Scoring Documents as in Round 1, assigning each to

one of the three performance levels, using the PLDs. As they completed Round 2, they turned in their completed rating sheets, and MI staff calculated cut scores for Levels 2 and 3, using logistic regression as described in Cizek & Bunch (2007, Ch. 9).

On the final day of the meeting, the ELA and Math panels were divided into two groups: one for vertical articulation and another for a preview session with MDE staff. Half of the panelists participated in MDE's preview session and the other half in the vertical articulation. Dr. Bunch provided an introduction to vertical articulation (see Appendix A) and gave the panelists their charge. They then divided by subject and conducted separate vertical articulations for ELA and Math.

MDE preview. The MDE preview was divided into three parts:

- Part 1: Overview of the changes to alternate assessments under ESSA, in particular, the 1% cap imposed at the state level on participation. The assessment selection guidance document for participation that should be used by IEP teams was distributed and reviewed.

- Part 2: New content expectations in science. The participants received an overview of Michigan's content expectations in science that were adopted by the state board of education in November of 2015. Table groups responded to specific questions regarding these new content standards and students with the most significant cognitive impairments. This feedback will compiled and used in the development of the process for creating alternate content expectations in science aligned to the current Michigan science content expectations. This process will start this fall and will involve our item development vendor and a series of educator panels.

- Part 3: A review of key test administration issues for MI-Access (specific to the levels represented in the room: P/SI or FI). These issues were identified by commonly seen incident reports this past year, as well as issues or questions raised during standard setting regarding the administration of MI-Access.

Vertical articulation. The vertical articulation facilitators (Drs. Bunch and Deville) presented displays of data depicting the Round 2 results in terms of cut scores, percent of students at or above each cut score, and percent of students in each category (Emerging, Attained, and Surpassed). Panelists also had access to all test materials they had used during the two rounds of standard setting. The processes for changing any cut score were as follows:

- Motion from the floor to make a specific cut score change (e.g., change the Math 6-8 Participation Level 2 cut score from 24 to 26)
- Second to the motion
- Discussion
- Vote

For the vote, a 2/3 majority was required for passage inasmuch as the action effectively overrode the work of two rounds of panel activity. As panelists recommended changes, the facilitator would enter the new cut score, and the remaining tables and graphic on the display would update so that panelists could see the immediate impact of the change. The two facilitators kept the discussion focused on the PLDs and the relationship between the new cut score and the performance level.

Functional Independence

Test booklets for Functional Independence are similar to those for M-STEP (i.e., a combination of selected and constructed response items for FI ELA and selected response items only for FI Math). The numbers of students taking the FI assessments are sufficient to calibrate the items using item response theory. Therefore, the Bookmark procedure (Lewis, Mitzel, Mercado, & Schulz, 2012) was not only appropriate but clearly indicated. In this procedure, panelists review test items from easiest to most difficult and identify points in the ordered item booklet where students at the threshold of a given performance level would cease to have a reasonable chance of answering correctly. This level is typically 50 or 67 percent, depending on the nature of the test and the judgment of the technical advisory committee (TAC). For this particular application, the threshold level was set at 67 percent. After the third round of item review, MI conducted a vertical articulation, engaging representatives of each grade level for each content area to examine all cut scores across all grades and recommend changes.

MDE calibrated for each selected-response item and each score point for each constructed-response item a theta value associated with a fixed probability (.67) of answering each selected-response item correctly or achieving that particular score or better on each constructed-response item. These theta estimates were then used to order selected-response items and score points of constructed-response items from easiest to most difficult in order to construct an ordered item booklet (OIB) for each assessment. MDE conducted all necessary item calibrations and constructed all OIBs based on input and requirements developed with MI. MI staff reviewed the item calibrations and the OIBs prior to on-site standard setting. MI staff prepared training materials and made copies of the OIBs.

Training materials. MI prepared materials for an opening session that included the goals and tasks of the session as well as a PowerPoint presentation on the Bookmark procedure. In addition, panel facilitators prepared grade/subject-specific materials that helped panelists understand the nature of the tests and factors affecting performance. All training materials were submitted to the MDE for review and approval prior to implementation. Training materials are listed below and included in Appendix A. PowerPoint presentations are included in Appendix C.

- Overview (PowerPoint)

- Bookmark Training (PowerPoint)
- Facilitator Script
- Bookmark Practice Round Form
- Bookmark Entry Form – Round 1
- Readiness Form
- Process Evaluation Form

Agenda. Table 3 shows the day-by-day agenda for the three-day event.

Table 3
Detailed Agenda: Functional Independence

Day/ Session	Panel 9: Math 3-4	Panel 10: Math 5-6	Panel 11: Math 7-8	Panel 12: Math HS	Panel 13: ELA 3-4	Panel 14: ELA 5-6	Panel 15: ELA 7-8	Panel 16: ELA HS
7/10 a.m.	Introduction/Training							
7/10 p.m.	Round 1 Grade 3	Round 1 Grade 5	Round 1 Grade 7	Round 1 HS	Round 1 Grade 3	Round 1 Grade 5	Round 1 Grade 7	Round 1 HS
7/11 a.m.	Round 1 Grade 4; Review Round 1 Grade 3; Round 2 Grade 3	Round 1 Grade 6; Review Round 1 Grade 5; Round 2 Grade 5	Round 1 Grade 8; Review Round 1 Grade 7; Round 2 Grade 7	Review Round 1 HS; Round 2 HS	Round 1 Grade 4; Review Round 1 Grade 3; Round 2 Grade 3	Round 1 Grade 6; Review Round 1 Grade 5; Round 2 Grade 5	Round 1 Grade 8; Review Round 1 Grade 7; Round 2 Grade 7	Review Round 1 HS; Round 2 HS
7/11 p.m.	Review Round 1 Grade 4; Round 2 Grade 4	Review Round 1 Grade 6; Round 2 Grade 6	Review Round 1 Grade 8; Round 2 Grade 8	Review Round 2 HS; Round 3 HS	Review Round 1 Grade 4; Round 2 Grade 4	Review Round 1 Grade 6; Round 2 Grade 6	Review Round 1 Grade 8; Round 2 Grade 8	Review Round 2 HS; Round 3 HS
7/12 a.m.	Review Round 2 Grade 4; Round 3 Grade 4	Review Round 2 Grade 6; Round 3 Grade 6	Review Round 2 Grade 8; Round 3 Grade 8		Review Round 2 Grade 4; Round 3 Grade 4	Review Round 2 Grade 6; Round 3 Grade 6	Review Round 2 Grade 8; Round 3 Grade 8	
7/12 p.m.	Articulation Training/MDE Preview							
	Articulation; Wrap-Up; Evaluation MDE Preview				Articulation; Wrap-Up; Evaluation MDE Preview			

Conduct of the meeting. Dr. Bunch provided an overview of the four days and gave the panels their charge (see Appendix B). In addition, John Jacquith from MDE provided background information with respect to the development, administration, and scoring of the FI assessments. Afterwards, panelists dispersed to their breakout rooms to review the FI tests under the direction of the facilitators listed in Table 3. These same facilitators also led the panelists in a review of the PLDs. After lunch on July 10, Dr. Bunch provided an overview of the Bookmark procedure (see Appendix B). Panelists then dispersed to their breakout rooms for Bookmark practice with a small set of items. Following this practice round, facilitators answered questions, and determined readiness to begin Round 1 by administering and reviewing the Round 1 Readiness Form (see Appendix B). Dr. Bunch, MDE staff, and Dr. Wyse circulated among the eight panel rooms throughout each day to observe and answer questions.

Panelists worked in small groups of 3-4 within a room of 8-10. They consulted with others at their table during each round. Panelists proceeded through three rounds of Bookmark item rating with feedback and discussion between rounds. As they worked their way through their ordered item booklets, they entered their bookmarks on scannable documents like the one shown in Figure 5.

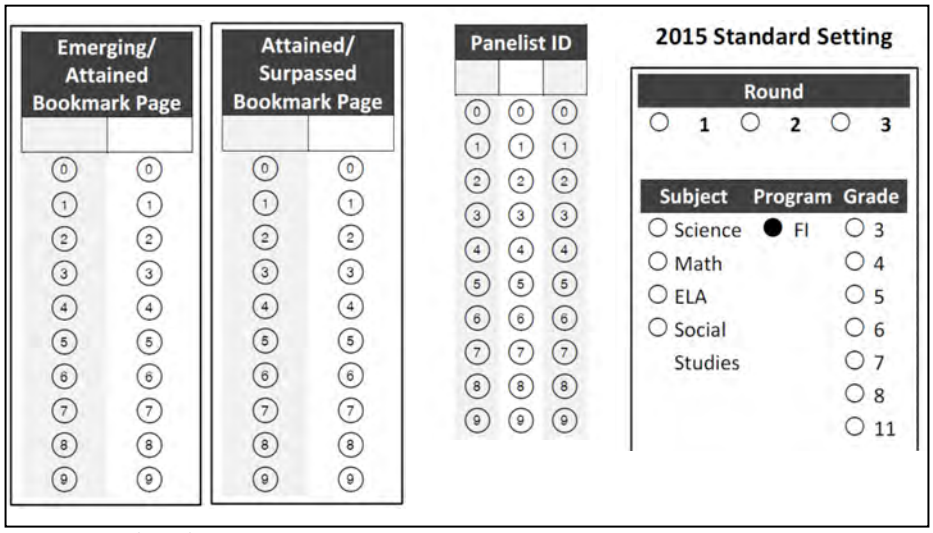


Figure 5. Bookmark Item Rating Form

As panelists completed their Bookmark item ratings, MI staff gathered them and processed the results. Because panelists worked in two-grade bands, they completed Round 1 for both grades before proceeding to Round 2 for either grade.

As panelists completed a round of Bookmark rating, MI staff collected the forms and processed them as described in Cizek & Bunch (2007, Ch. 10). Prior to Round 2, facilitators shared results of Round 1, facilitated a discussion of results in terms of dispersion of bookmarks as well as median cut score, and shared impact data. After discussion of Round 1 results, panelists completed the Readiness Form, indicating readiness to begin Round 2. They completed Round 2 as they had completed Round 1, working in small groups and entering two bookmarks. At the end of the round, MI staff collected the scannable forms, processed them as in Round 1, and prepared results to present to panelists. During the discussion of Round 2 results, facilitators shared the same types of information they had shared after Round 1 but also revealed impact data (i.e., the percentages of students who would be classified at each level as a result of the Round 2 cut scores) as additional context.

At the close of the Round 2 discussion, panelists indicated their readiness to begin Round 3. They proceeded through Round 3 as they had in Rounds 1 and 2, entering two bookmarks on their scannable forms. MI facilitators collected the forms and processed them as in Rounds 1 and 2. As noted in Table 3, the high school panels were scheduled to complete Round 3 on the afternoon of July 11, while all other panels were to complete Round 3 the morning of July 12. The other three ELA panels (grades 3-4, 5-6, and 7-8) also completed both Rounds 2 and 3 the afternoon of July 11. Inspection of their readiness forms, data entry forms, and other materials showed that they had indeed followed all directions and entered valid ratings on their bookmark forms. After Round 3, ELA and Math panels divided into two groups: vertical articulation and MDE process preview. Procedures for forming and leading the groups were the same as that described above for the P/SI panels.

MDE Preview. The MDE preview was divided into three parts:

- Part 1: Overview of the changes to alternate assessments under ESSA, in particular, the 1% cap imposed at the state level on participation. The assessment selection guidance document for participation that should be used by IEP teams was distributed and reviewed.

- Part 2: New content expectations in science. The participants received an overview of Michigan's content expectations in science that were adopted by the state board of education in November of 2015. Table groups responded to specific questions regarding these new content standards and students with the most significant cognitive impairments. This feedback will compiled and used in the development of the process for creating alternate content expectations in science aligned to the current Michigan science content expectations. This process will start this fall and will involve our item development vendor and a series of educator panels.

- Part 3: A review of key test administration issues for MI-Access (specific to the levels represented in the room: P/SI or FI). These issues were identified by commonly seen

incident reports this past year, as well as issues or questions raised during standard setting regarding the administration of MI-Access.

Vertical articulation. Vertical articulation began with an overview of the process, followed by a question-and-answer period. During this phase of the process, both ELA and Math VACs met together. After the question-and-answer session, ELA and Math VACs separated into different rooms. During the remainder of the day, each VAC reviewed results (cut scores and impact) across grade spans and recommended changes. Changes were effected by a motion, second, discussion and vote. Given that the changes were to override cut scores set over three rounds of deliberation, a 2/3 majority was required to pass any motion.

Results

Round-by-Round Results

Tables 4-10 show the round-by-round results of the all standard setting activities. Figures 6-8 show the impacts of the Round 3 cut scores.

Table 4

Round 1 Results for Participation

Test	Cut Scores		% in Level		
	Attained	Surpassed	Emerging	Attained	Surpassed
ELA Grade 3	31	47	45.2	32.5	22.4
ELA Grade 4	32	46	40.8	30.8	28.4
ELA Grade 5	28	42	38.8	26.5	34.7
ELA Grade 6	28	40	34.0	25.4	40.7
ELA Grade 7	29	45	44.2	31.6	24.2
ELA Grade 8	29	40	51.1	18.7	30.3
ELA High School	35	47	41.0	26.4	32.7
Math Grade 3	31	46	46.5	25.1	28.3
Math Grade 4	31	48	47.0	33.3	19.7
Math Grade 5	31	46	48.8	31.2	20.0
Math Grade 6	32	43	46.7	26.5	26.8
Math Grade 7	26	45	35.1	38.3	26.6
Math Grade 8	31	44	48.9	22.1	29.0
Math High School	29	44	34.6	33.3	32.1

Table 5

Round 2 Results for Participation

Test	Cut Scores		% in Level		
	Attained	**Surpassed**	**Emerging**	**Attained**	**Surpassed**
ELA Grade 3	31	45	45.1	27.3	27.6
ELA Grade 4	32	43	40.6	26.1	33.3
ELA Grade 5	28	42	38.8	26.6	34.6
ELA Grade 6	29	41	37.5	25.8	36.7
ELA Grade 7	28	45	40.3	35.4	24.4
ELA Grade 8	27	43	45.6	32.1	22.3
ELA High School	34	46	38.3	26.5	35.1
Math Grade 3	33	47	49.4	24.7	25.9
Math Grade 4	32	47	47.4	32.6	20.1
Math Grade 5	32	46	48.8	31.1	20.1
Math Grade 6	31	44	46.4	27.0	26.7
Math Grade 7	27	43	38.4	31.7	29.9
Math Grade 8	28	43	38.7	32.0	29.3
Math High School	31	46	40.8	30.7	28.5

Table 6
Round 1 Results for Supported Independence

Test	Cut Scores		% in Level		
	Attained	Surpassed	Emerging	Attained	Surpassed
ELA Grade 3	28	43	20.5	33.4	46.1
ELA Grade 4	32	48	20.9	38.1	40.9
ELA Grade 5	30	43	17.8	28.3	53.9
ELA Grade 6	30	43	13.9	23.7	62.4
ELA Grade 7	29	44	14.4	29.4	56.2
ELA Grade 8	32	44	18.3	21.1	60.6
ELA High School	32	46	24.1	29.9	46.0
Math Grade 3	35	47	45.7	26.4	27.9
Math Grade 4	34	45	30.0	29.3	39.3
Math Grade 5	30	44	24.2	31.4	44.5
Math Grade 6	31	44	37.3	30.6	32.1
Math Grade 7	30	46	29.1	45.3	25.6
Math Grade 8	29	46	23.7	41.3	35.0
Math High School	32	46	23.0	34.8	42.1

Table 7
Round 2 Results for Supported Independence

Test	Cut Scores		% in Level		
	Attained	Surpassed	Emerging	Attained	Surpassed
ELA Grade 3	28	43	20.5	33.4	46.1
ELA Grade 4	31	44	20.6	26.7	52.8
ELA Grade 5	30	46	17.8	34.6	47.5
ELA Grade 6	31	44	16.0	22.2	61.8
ELA Grade 7	31	46	17.4	32.3	50.3
ELA Grade 8	33	45	20.2	23.8	56.0
ELA High School	35	46	30.5	23.6	46.0
Math Grade 3	35	47	45.7	26.4	27.9
Math Grade 4	34	45	31.4	29.3	39.3
Math Grade 5	31	46	27.3	35.1	37.7
Math Grade 6	32	44	37.3	30.6	32.1
Math Grade 7	30	45	29.1	44.5	26.4
Math Grade 8	30	46	23.9	41.1	35.0
Math High School	33	47	26.6	34.5	39.0

Table 8
Round 1 Results for Functional Independence

Test	Cut Scores		% in Level		
	Attained	Surpassed	Emerging	Attained	Surpassed
ELA Grade 3	0.360	1.341	22.3	35.7	42.0
ELA Grade 4	0.365	1.661	17.2	31.6	51.2
ELA Grade 5	0.193	1.107	11.2	19.7	69.1
ELA Grade 6	0.636	1.678	18.8	28.1	53.1
ELA Grade 7	-0.215	0.685	6.0	14.8	79.2
ELA Grade 8	0.589	1.141	14.2	8.8	77.0
ELA High School	0.261	1.049	11.9	14.2	73.9
Math Grade 3	0.584	1.104	34.6	12.0	53.4
Math Grade 4	0.444	1.138	24.1	20.3	55.6
Math Grade 5	0.768	1.218	28.2	13.7	58.1
Math Grade 6	0.621	1.351	46.7	24.2	29.1
Math Grade 7	0.199	1.393	38.8	35.0	26.2
Math Grade 8	0.673	1.390	43.2	21.3	35.5
Math High School	0.213	0.568	33.6	13.1	53.3

Table 9
Round 2 Results for Functional Independence

Test	Cut Scores		% in Level		
	Attained	Surpassed	Emerging	Attained	Surpassed
ELA Grade 3	0.629	1.380	26.8	31.2	42.0
ELA Grade 4	0.041	1.661	9.8	39.0	51.2
ELA Grade 5	0.193	1.107	11.2	19.7	69.1
ELA Grade 6	0.636	1.697	18.8	28.1	53.1
ELA Grade 7	-0.215	0.955	6.0	18.8	75.2
ELA Grade 8	0.589	1.141	14.2	8.8	77.0
ELA High School	0.233	1.049	11.9	14.2	73.9
Math Grade 3	0.584	1.684	34.6	26.0	39.4
Math Grade 4	0.444	1.138	24.1	20.3	55.6
Math Grade 5	0.870	1.634	34.4	24.8	40.8
Math Grade 6	0.621	1.351	46.7	24.2	29.1
Math Grade 7	0.199	1.404	38.8	35.0	26.2
Math Grade 8	0.673	1.390	43.2	21.3	35.5
Math High School	0.095	1.071	27.8	34.2	38.0

Table 10
Round 3 Results for Functional Independence

Test	Cut Scores		% in Level		
	Attained	Surpassed	Emerging	Attained	Surpassed
ELA Grade 3	0.525	1.652	26.8	38.6	34.6
ELA Grade 4	0.338	1.661	13.8	35.0	51.2
ELA Grade 5	0.384	1.531	13.9	28.4	57.7
ELA Grade 6	0.636	1.697	18.8	28.1	53.1
ELA Grade 7	-0.207	0.955	6.0	18.8	75.2
ELA Grade 8	0.589	1.141	14.2	8.8	77.0
ELA High School	0.233	1.049	11.9	14.2	73.9
Math Grade 3	0.584	2.067	34.6	34.0	31.4
Math Grade 4	0.444	1.363	24.1	29.7	46.2
Math Grade 5	0.87	2.022	34.4	32.8	32.8
Math Grade 6	0.621	1.351	46.7	24.2	29.1
Math Grade 7	0.199	1.404	38.8	35.0	26.2
Math Grade 8	0.367	1.39	29.7	34.8	35.5
Math High School	0.095	1.074	27.8	34.2	38.0

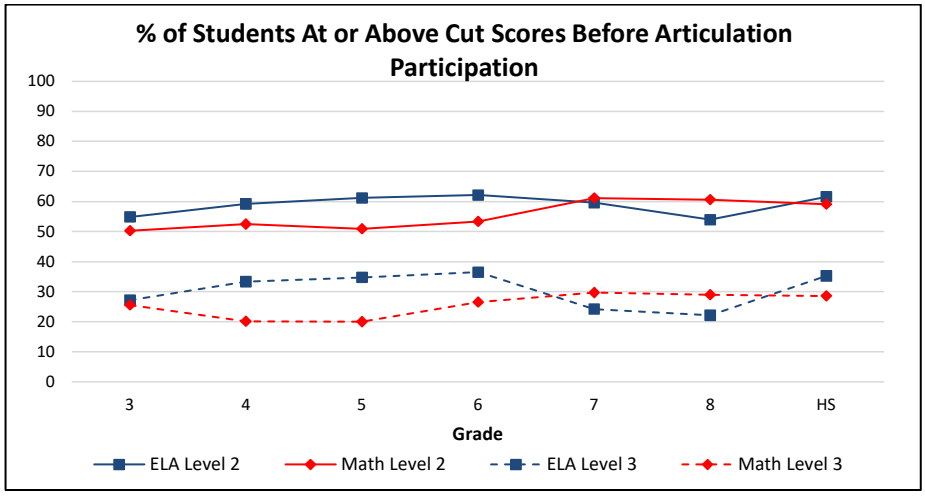


Figure 6. Final Round Impact: Participation

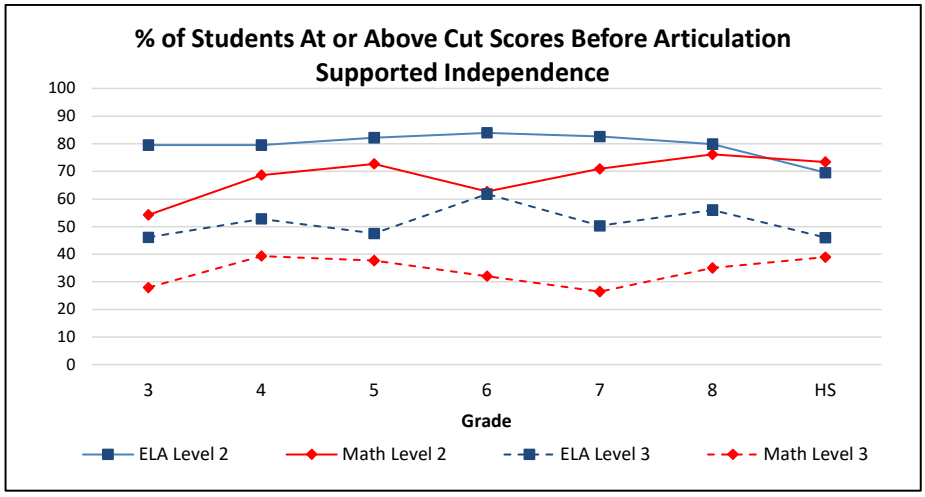


Figure 7. Final Round Impact: Supported Independence

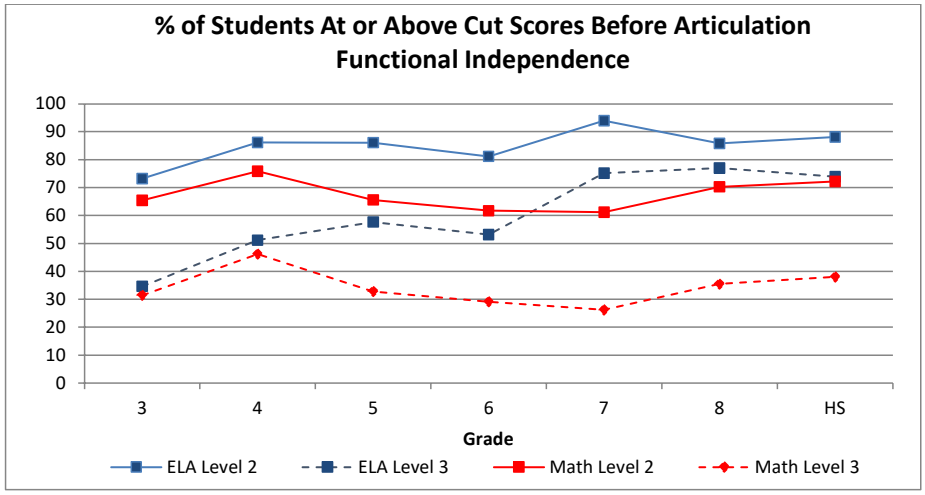


Figure 8. Final Round Impact: Functional Independence

Vertical Articulation

There were no cut score changes for Participation (i.e., Tables 9 and 13 are identical) There was one change for Supported Independence (ELA Grade 6 Level 3) There were three changes for Functional Independence (ELA Grade 7 Level 2, ELA Grade 8 Level 3, and Math Grade 6 Level 2). Results are depicted in Tables 11-13. Highlighted entries in these tables indicate changes, relative to Round 2 for P/SI or Round 3 for FI (Tables 8-10). Figures 9-11 show the impacts after vertical articulation.

Table 11
Cut Scores and Impact for Participation – After Vertical Articulation

Test	Cut Scores		% in Level		
	Attained	Surpassed	Emerging	Attained	Surpassed
ELA Grade 3	31	45	45.2	27.7	27.2
ELA Grade 4	32	43	40.8	25.9	33.3
ELA Grade 5	28	42	38.8	26.5	34.7
ELA Grade 6	29	41	37.9	25.6	36.5
ELA Grade 7	28	45	40.5	35.3	24.2
ELA Grade 8	27	43	46.1	31.8	22.1
ELA High School	34	46	38.4	26.4	35.2
Math Grade 3	33	47	49.8	24.7	25.6
Math Grade 4	32	47	47.5	32.3	20.2
Math Grade 5	32	46	49.1	30.9	20.0
Math Grade 6	31	44	46.7	26.8	26.5
Math Grade 7	27	43	38.9	31.4	29.7
Math Grade 8	28	43	39.4	31.7	29.0
Math High School	31	46	40.9	30.5	28.6

Table 12

Cut Scores and Impact for Supported Independence – After Vertical Articulation

Test	Cut Scores		% in Level		
	Attained	Surpassed	Emerging	Attained	Surpassed
ELA Grade 3	28	43	20.5	33.4	46.1
ELA Grade 4	31	44	20.6	26.7	52.8
ELA Grade 5	30	46	17.8	34.6	47.5
ELA Grade 6	31	46	16.0	30.8	53.2
ELA Grade 7	31	46	17.4	32.3	50.3
ELA Grade 8	33	45	20.2	23.8	56.0
ELA High School	35	46	30.5	23.6	46.0
Math Grade 3	35	47	45.7	26.4	27.9
Math Grade 4	34	45	31.4	29.3	39.3
Math Grade 5	31	46	27.3	35.1	37.7
Math Grade 6	32	44	37.3	30.6	32.1
Math Grade 7	30	45	29.1	44.5	26.4
Math Grade 8	30	46	23.9	41.1	35.0
Math High School	33	47	26.6	34.5	39.0

Table 13

Cut Scores and Impact for Functional Independence – After Vertical Articulation

Test	Cut Scores		% in Level		
	Attained	Surpassed	Emerging	Attained	Surpassed
ELA Grade 3	0.525	1.65	26.8	38.6	34.6
ELA Grade 4	0.338	1.70	13.8	35.0	51.2
ELA Grade 5	0.384	1.53	13.9	28.4	57.7
ELA Grade 6	0.636	1.70	18.8	28.1	53.1
ELA Grade 7	0.098	0.96	9.0	15.8	75.2
ELA Grade 8	0.589	1.38	14.2	16.5	69.3
ELA High School	0.233	1.05	11.9	14.2	73.9
Math Grade 3	0.584	2.067	34.6	34.0	31.4
Math Grade 4	0.444	1.363	24.1	29.7	46.2
Math Grade 5	0.87	2.022	34.4	32.8	32.8
Math Grade 6	.517	1.351	38.3	32.6	29.1
Math Grade 7	0.199	1.404	38.8	35.0	26.2
Math Grade 8	0.367	1.39	29.7	34.8	35.5
Math High School	0.095	1.074	27.8	34.2	38.0

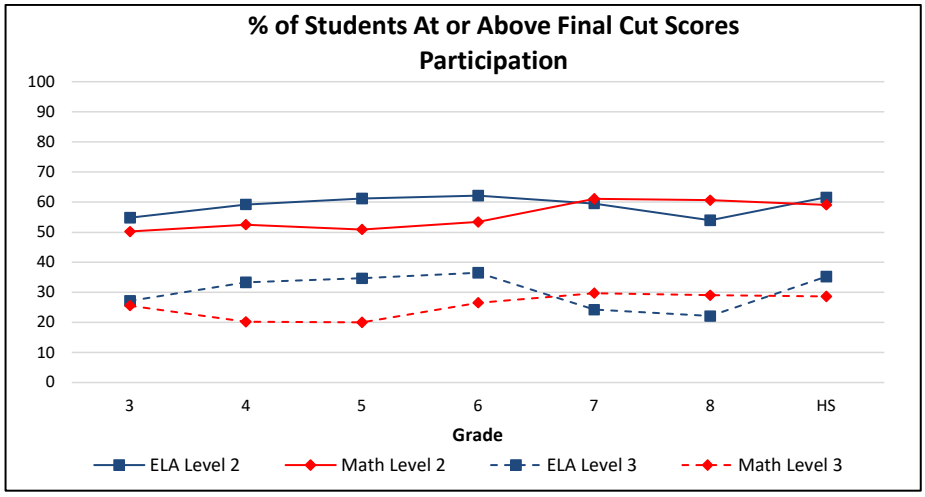


Figure 9. Impact After Vertical Articulation: Participation

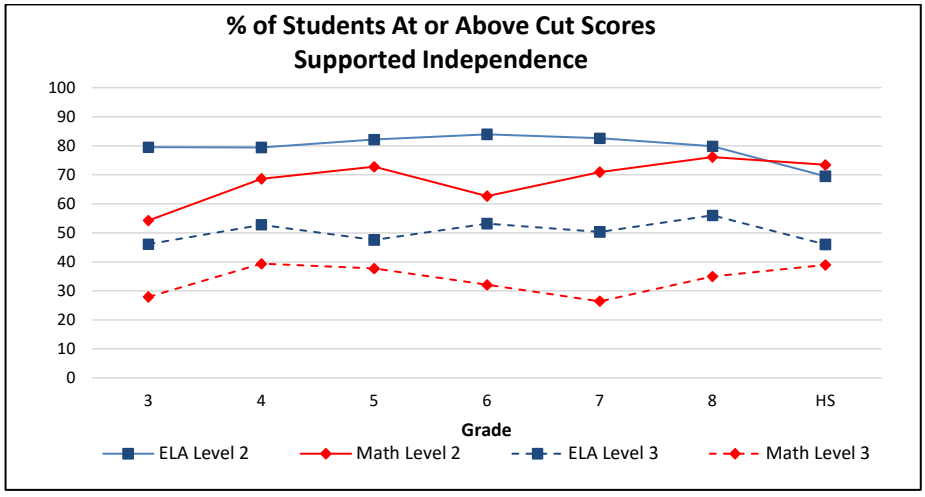


Figure 10. Impact After Vertical Articulation: Supported Independence

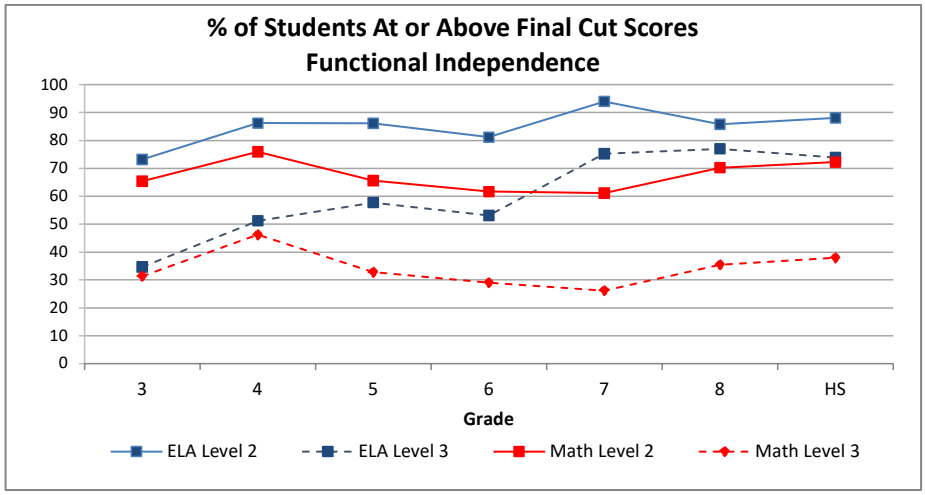


Figure 11. Impact After Vertical Articulation: Functional Independence

Evaluation

Panelists evaluated the process and their facilitators on eight critical-incident factors, each on a 5-point scale (Strongly Agree to Strongly Disagree). With regard to facilitators and process, 98-100 percent of panelists agreed with each statement. With regard to facilities and food, reaction was mixed, with 43 percent of FI panelists and 53 percent of FI panelists agreeing that the facilities and food service helped to create a good working environment. Results are summarized in Tables 14-17.

Table 14

Evaluation Results for Participation

[SA = Strongly Agree; A = Agree; N = Neutral; D = Disagree; SD = Strongly Disagree]

Statement	SA% + A%
Overall, the facilities and food service helped to create a good working environment.	58%
Overall, the training in the standard-setting purpose and methods was clear.	97%
Overall, I am confident that I was able to apply the standard setting methods appropriately.	100%
Overall, the standard setting procedures allowed me to use my experience and expertise to recommend cut scores for the tests.	100%
Overall, the facilitator helped to ensure that everyone was able to contribute to the group discussions and that no one unfairly dominated the discussions.	97%
Overall, I was able to understand and use the feedback provided (e.g., other participants' ratings, impact data).	100%
I believe that the final group-recommended cut score fairly represents the minimal level of performance for students at the **Attained** level.	97%
I believe that the final group-recommended cut score fairly represents the minimal level of performance for students at the **Surpassed** level.	98%

Table 15

Evaluation Results for Supported Independence

[SA = Strongly Agree; A = Agree; N = Neutral; D = Disagree; SD = Strongly Disagree]

Statement	SA% + A%
Overall, the facilities and food service helped to create a good working environment.	48%
Overall, the training in the standard-setting purpose and methods was clear.	100%
Overall, I am confident that I was able to apply the standard setting methods appropriately.	100%
Overall, the standard setting procedures allowed me to use my experience and expertise to recommend cut scores for the tests.	100%
Overall, the facilitator helped to ensure that everyone was able to contribute to the group discussions and that no one unfairly dominated the discussions.	100%
Overall, I was able to understand and use the feedback provided (e.g., other participants' ratings, impact data).	100%
I believe that the final group-recommended cut score fairly represents the minimal level of performance for students at the **Attained** level.	100%
I believe that the final group-recommended cut score fairly represents the minimal level of performance for students at the **Surpassed** level.	98%

Table 16

Evaluation Results for Functional Independence

[SA = Strongly Agree; A = Agree; N = Neutral; D = Disagree; SD = Strongly Disagree]

Statement	SA% + A%
Overall, the facilities and food service helped to create a good working environment.	62%
Overall, the training in the standard-setting purpose and methods was clear.	99%
Overall, I am confident that I was able to apply the standard setting methods appropriately.	99%
Overall, the standard setting procedures allowed me to use my experience and expertise to recommend cut scores for the tests.	99%
Overall, the facilitator helped to ensure that everyone was able to contribute to the group discussions and that no one unfairly dominated the discussions.	100%
Overall, I was able to understand and use the feedback provided (e.g., other participants' ratings, impact data).	99%
I believe that the final group-recommended cut score fairly represents the minimal level of performance for students at the **Attained** level.	95%
I believe that the final group-recommended cut score fairly represents the minimal level of performance for students at the **Surpassed** level.	91%

Table 17

Evaluation Results for Vertical Articulation

[SA = Strongly Agree; A = Agree; N = Neutral; D = Disagree; SD = Strongly Disagree]

Statement	SA% + A%
Overall, the training for this task was clear.	100%
The tables and graphs helped me keep track of the cut scores and the impact of the decisions we were making.	100%
The facilitator was effective in guiding discussion and keeping it moving toward a decision.	100%
The facilitator helped to ensure that everyone was able to contribute to the group discussions and that no one unfairly dominated the discussions.	100%
I had access to information I needed to make decisions about cut scores.	97%
Overall, I am confident that I was able to participate in this activity appropriately.	100%
The process was fair.	100%
Overall, I was able to understand and use the feedback provided (e.g., other participants' ratings, impact data).	100%
I believe that the final, articulated cut scores fairly represent expectations across grades at the **Attained** level.	100%
I believe that the final, articulated cut scores fairly represent expectations across grades at the **Surpassed** level.	100%

Sample comments. In general, panelists were enthusiastic about the facilities and the facilitators but less so about the meals choices. Most comments on the Final Evaluation forms were about food. That issue aside, panelists were very favorably disposed toward all aspects of the experience. Here are a few comments from each session. All comments are included in Appendix B.

- From Participation
  - Color coded paper to help keep packet more organized; Use less paper more digital!
  - Snack in the afternoon would be good. Water available in each room or nearby?
  - I felt that this was a great experience. It was interesting to go through this process.
  - It was good to hear others point of view. Lidia was great at facilitator.
  - Smooth - Best standard setting panel I have attended
- From Supported Independence
  - Digital!! Less Paper!!
  - This was a great experience and opportunity to understand the test more.
  - I had a great time being able to talk others and share resources. Everything was well organized. Great job!
- From Functional Independence
  - There is something wrong with a process where on 11th grader has an easier time getting a surpassed score than someone in another grade. These scores across the ELA grades are radically different and will not, in my opinion, reflect accurate results. I would recommend the same group looking at each ELA test.
  - Overall, I think the process is great! However, I think rounds 2 and 3 need less time than round 1. There is still too much down time.
  - Good experience and a great opportunity to participate with the MDE. Good job!
  - Rooms were cold. Training was well explained. Group encouraged discussion, sharing, and collaboration.

There were many expressions of thanks to MDE for listening and for paying attention to the needs of this population of students. Panelists found the experience very rewarding and expressed their gratitude for the support they receive from MDE.

Conclusions and Recommendations

The process for arriving at cut scores was rigorous, consistent with best practices, conducted by highly competent practitioners, and monitored by a highly qualified outside observer. Panelists had high praise for the facilitators and expressed great confidence in the validity of the cut scores their panels set. The resulting cut scores and corresponding impacts were reasonably consistent across grades as well as with historical trends in Michigan for these populations. It is our recommendation that the cut scores be adopted without modification or adjustment.

References

Cizek, G. J. & Bunch, M. B. (2007). *Standard Setting, Establishing and Evaluating Performance Standards on Tests*. Thousand Oaks, CA: Sage.

Kingston, N. M. & 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, NY: Routledge.

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.

Appendix A

Training Materials

- Facilitator Script for Participation and Supported Independence
- Facilitator Script for Functional Independence
- Practice Round Form for Participation and Supported Independence
- Practice Round Form for Functional Independence
- Readiness Form for Participation
- Readiness Form for Supported Independence
- Readiness Form for Functional Independence
- Process Evaluation Form for Participation
- Process Evaluation Form for Supported Independence
- Process Evaluation Form for Functional Independence
- Vertical Articulation Motions and Actions Form
- Vertical Articulation Evaluation Form

Facilitator Script for Participation and Supported Independence

Facilitator Script – Post-Overview Presentation

Goals: Introduce panelists to the Participation tests and PLDs.

Materials of Importance: Participation Tests; Participation PLDs; non-disclosure agreements.

Facilitator Outline:

1. Assist panelists with their table assignment.
2. Conduct group introduction (15-30 seconds per person).
3. Have panelists sign non-disclosure agreement and demographics form.
4. Allow panelists to become familiar with the Participation tests.
5. Dismiss panelists for 15-minute break at 10:30.
6. Lead panelists in a discussion of the Participation PLDs for the lower grade level followed by the upper grade level.
7. Dismiss panelists for 11:30 presentation of the Body of Work procedure.

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.
- Remind panelists that they have been exposed to the assessment tasks so that they can have first-hand experience of the types of items that students will be charged with completing, not so that they can critique the test development process.
- Ask panelists to discuss their impressions of the tasks. What did they think would have been easy or difficult for MI-Access students? What types of skills did they notice would be needed to successfully answer/perform the required tasks?
- Briefly remind panelists that PLDs are simply descriptions of students' abilities at a particular performance level; also point out that all their decisions concerning recommending cut scores must be firmly grounded in the PLDs.
- Ask panelists to carefully read the Participation PLDs for both grades and contemplate what it means to be Emerging, Attained, or Surpassed.
- Encourage panelists to imagine students they have known who might have fit the Emerging, Attained, or Surpassed 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.

Facilitator Script: Post-Body of Work Orientation Presentation

Goals: Introduce panelists to student work samples by leading them through the practice Body of Work (BoW) samples. Ensure all panelists are prepared to begin the Standard Setting process. Complete BoW Round 1 for the Participation test at the lower grade level first followed by the higher grade level. .

Materials of Importance: BoW practice samples; Readiness Form; Round 1 BoW samples for both grade levels of the Participation test.

Facilitator Outline:

1. Assist panelists through BoW Practice Round.
2. Have panelists complete Round 1 Readiness Form and begin Round 1 – Participation for the lower grade level first followed by the higher grade level.
3. Monitor Round 1 – Participation and be available to answer panelists' questions.
4. End Round 1 – Participation for both grades and collect and secure panelists' materials.
5. Bring the results of Round 1 – Participation to the data analysts.

Facilitator Talking Points:

- Conduct a room-wide discussion concerning the first two work samples in the BoW Practice samples.
- Have panelists complete the remaining four practice work samples with their tables.
- Encourage panelists to consult with the other people sitting at their tables during each round.
- Remind panelists that all their decisions concerning their placement of work samples into performance categories must be firmly grounded in the PLDs.
- Remind panelists to consider the following questions for each work sample:
    - What types of skills and abilities must a student possess to be capable of each work sample?
    - What skills and abilities make work samples progressively more challenging?
    - What performance level does each work sample best represent?
- Explain to panelists that it is ok for them to have reversals (e.g. work sample #4 is placed in Level 2 and work sample #5 is placed in Level 3) as they are sorting the work samples into categories. However, if they are consistently having an inordinate number of reversals encourage them to talk to you or to revisit their PLDs.
- Remind panelists to pace themselves. They have an hour and a half per grade level to sort all of their work samples into one of three performance categories.

- Encourage panelists to ask questions as they progress through Round 1. Emphasize that decisions must be based on the entire sample not just components and not on calculations.
- Before the panelists leave for the evening remind them of the next day's schedule and collect and secure all materials.

Facilitator Script: Review of Day 1; Participation Round 1 Discussion; Begin Round 2

Goals: Review Round 1 – Participation results for each grade level. Ensure all panelists are prepared to begin Round 2. Complete BoW Round 2 for the Participation test.

Materials of Importance: Round 1 – Participation results for each grade (tables, graphs, and impact data); Round 2 Readiness form; Round 2 – Participation work samples by grade.

Facilitator Outline:

1. Discuss Round 1 results for the Participation test for each grade.
2. Have panelists complete Round 2 Readiness Form and begin Round 2 – Participation for each grade.
3. Monitor Round 2 – Participation and be available to answer panelists' questions.
4. End Round 2 – Participation for each grade. Collect and secure panelists' materials.
5. Bring the results of Round 2 – Participation for each grade to the data analysts.

Facilitator Talking Points:

- Conduct a room-wide discussion concerning the Round 1 results for the Participation test by grade.
- Highlight the following topics:
    - What were the challenges panelists faced in Round 1?
    - What factors influenced panelists' decisions to rate certain work samples?
    - 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 work samples where the room was evenly divided in opinion (e.g. a work sample that half the room rated as Attained and the other half rated as Emerging).
- Review the Round 1 impact data for each grade.
- 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 three performance categories seem realistic?
- Explain to panelists the Round 2 process:
    - Work samples will be exactly the same as they were for Round 1. Carefully review the work samples in the relative vicinity of the Round 1 cut score.
    - Panelists should rate each work sample using the same process used in Round 1.
    - Panelists are free to discuss the work samples and PLDs with their tablemates, but not across tables.
- Ask if the panelists have any questions and ensure they are ready to begin Round 2.
- Before the panelist leave:
    - Select panelist to participate in the afternoon Articulation training and session.
    - Remind panelists of the schedule for the next day.
    - Collect and secure all their materials.

Goals: Introduce panelists to the Supported Independence (SI) tests and PLDs, Complete BoW Round 1 for the Supported Independence test for each grade.

Materials of Importance: BoW practice samples; Readiness Form; Round 1 BoW samples for both grade levels of the Supported Independence test.

Facilitator Outline:

1. Briefly review essential topics covered previously
2. Allow panelists to become familiar with the SI tests.
3. Lead panelists in a discussion of the Supported Independence PLDs by grade level.
4. Have panelists complete Round 1 Readiness Form and begin Round 1 – SI lower grade first.
5. Monitor Round 1 – Supported Independence and be available to answer panelists' questions.
6. End Round 1 – Supported Independence for both grades and dismiss panelists for lunch.
7. Bring the results of Round 1 – Supported Independence for each grade to the data analysts.

Facilitator Talking Points:

- Briefly review the following topics with the panelists:
    - PLDs:
        - PLDs are simply descriptions of students' abilities at a particular performance level.
        - All decisions in sorting student work samples must be firmly grounded in the PLDs.

        - The differences among Emerging, Attained, and Surpassed that the panel noted in the PLDs.
    - Body of Work procedure:
        - Each work sample represents the complete body of work for one student on the Supported independence test.
        - Work samples are ordered by total score with students receiving the lowest score appearing at the beginning of the set and students receiving the highest score appearing at the end of the set.
        - BoW procedure is characterized by utilizing the PLDs to place each student work sample into a performance category.
- Conduct a room-wide discussion concerning the first two work samples in the BoW Practice samples.
- Have panelists complete the remaining four practice work samples with their tables.
- Encourage panelists to consult with the other people sitting at their tables during each round.
- Remind panelists that all their decisions concerning their placement of work samples into performance categories must be firmly grounded in the PLDs.
- Before beginning Round 1 – Supported Independence, reiterate to panelists that they should consider the following questions for each work sample:
    - What types of skills and abilities must a student possess to be capable of each work sample?
    - What skills and abilities make work samples progressively more challenging?
    - What performance level does each work sample best represent?

Facilitator Script: Supported Independence Round 1 Discussion; Begin Round 2

Goals: Review Round 1 – Supported Independence results for each grade level. Ensure all panelists are prepared to begin Round 2. Complete BoW Round 2 for the Supported Independence.

Materials of Importance: Round 1 – Supported Independence results by grade (tables, graphs, and impact data); Round 2 Readiness form; Round 2 – Supported Independence work samples by grade.

Facilitator Outline:

1. Discuss Round 1 results for the Supported Independence test for each grade.
2. Complete Readiness Form and begin Round 2 – Supported Independence – lower grade first.

3. Monitor Round 2 – Supported Independence and be available to answer panelists' questions.
4. End Round 2 – Supported Independence and dismiss panelists for lunch.
5. Bring the results of Round 2 – Supported Independence (both grades) to the data analysts.

Facilitator Talking Points:

- Conduct a room-wide discussion concerning the Round 1 results for the SI test by grade.
- Highlight the following topics:
    - What were the challenges panelists faced in Round 1?
    - What factors influenced panelists' decisions to rate certain work samples?
    - 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 work samples where the room was evenly divided in opinion (e.g. a work sample that half the room rated as Attained and the other half rated as Emerging).
- Review the Round 1 impact data for each grade.
- 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 three performance categories seem realistic?
- Reiterate to panelists the Round 2 process:
    - Work samples will be exactly the same as they were for Round 1. Carefully review the work samples in the relative vicinity of the Round 1 cut score.
    - Panelists should rate each work sample using the same process used in Round 1.
    - Panelists are free to discuss the work samples and PLDs with their tablemates, but not across tables.
- Ask if the panelists have any questions and ensure they are ready to begin Round 2.
- Before the panelist leave:
    - Select panelist to participate in the Articulation session.
    - Collect and secure all their materials.

Facilitator Script for Functional Independence

Facilitator Script – Post-Overview Presentation

Goals: Introduce panelists to the Functional Independence (FI) tests and to the PLDs.

Materials of Importance: FI Tests; FI PLDs; non-disclosure agreements; demographics form.

Facilitator Outline:

1. Assist panelists with their 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 FI tests.
5. Dismiss panelists for 15-minute break at 10:30. [Stagger across the 8 rooms by 2-3 minutes]
6. Lead panelists in a discussion of the PLDs.
7. Dismiss panelists for 11:30 presentation of the Bookmark procedure.

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.
- Remind panelists that they are taking/studying the tests so that they can have first-hand experience of the types of items that students will be charged with completing, not so that they can critique the item/test development process.
- Ask panelists to discuss their impressions of the test content. What did they think would have been easy or difficult for MI-Access students? What types of skills did they notice would be needed to successfully answer the items on the test?
- Briefly remind panelists that PLDs are simply descriptions of students' abilities at a particular performance level; also point out that all their decisions concerning setting cut scores must be firmly grounded in the PLDs.
- Ask panelists to read the PLDs carefully and to contemplate what it means to be Emerging, Attained, or Surpassed.
- Encourage panelists to imagine students they have known who might have fit the Emerging, Attained, or Surpassed 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*** Attained students and what differentiates them from the Emerging performance level. Then, ask the panelists to focus on the ***Just Barely*** Surpassed students and what differentiates them from the Attained performance level.

Facilitator Script: Post-Bookmark Orientation Presentation

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

Materials of Importance: Bookmark practice OIB; Round 1 Readiness Form; Round 1 OIB for the Functional Independence test; Round 1 Bookmark Rating Forms.

Facilitator Outline:

6. Assist panelists through Bookmark Practice Round.
7. Have panelists complete Round 1 Readiness Form and begin Round 1.
8. Monitor Round 1 and be available to answer panelists' questions.
9. Dismiss panelists for the evening and collect their secure materials.
10. Bring completed Bookmark Rating Forms to data analysts.

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 ***Just Barely*** Attained students would have a less than 67% 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 Emerging and Attained performance levels.
- Discuss the results of the Practice Round with the group. Note the range of pages where panelists set their bookmarks.

- 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 ***Just Barely*** Attained students have a 67% chance of answering the item correctly?
    - After placing the Attained bookmark, would ***Just Barely*** Surpassed students have a 67% chance of answering the item correctly?
- Explain to panelists that once they identify an item that they think ***Just Barely*** Attained or Surpassed students have a less than 67% chance of answering correctly that they should take a look at the next few items in the Order-Item Booklet to confirm that they have reached the best page 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.
- 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. They have three hours to place their bookmarks.

Facilitator Script: Review of Day 1; Finalize Round 1

Goals: Complete Bookmark Round 1.

Materials of Importance: Round 1 Ordered-Item Booklet; Round 1 Bookmark Rating Forms.

Facilitator Outline:

8. Briefly review essential topics covered in Day 1.
    a. Performance Level Descriptors (PLDs).
    b. Bookmark procedure.
9. Monitor Round 1 and be available to answer panelists' questions.
10. End Round 1 and collect panelists' secure materials.
11. Bring completed Bookmark forms to data analysts.

Facilitator Talking Points:

- Briefly review the following topics with the panelists:
    - PLDs:

        - PLDs are simply descriptions of students' abilities at a particular performance level.
        - All decisions in bookmark placement must be firmly grounded in the PLDs.
        - The differences among Emerging, Attained, and Surpassed that the panel noted in the PLDs.
    - Bookmark procedure:
        - 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.
        - The bookmark procedure we will use is characterized by the placement of two bookmarks on the first items in the OIB that the ***Just Barely*** Attained or ***Just Barely*** Surpassed students would have a less than 67% chance of answering correctly.
- Before beginning Round 1 for the next grade level, reiterate to panelists that they should consider the following questions for each item in the OIB:
    - What do you know about students who correctly answer this item?
    - What makes items progressively more challenging?
    - Would ***Just Barely*** Attained students have a 67% chance of answering the item correctly?
    - After placing the Attained bookmark, would ***Just Barely*** Surpassed students have a 67% chance of answering the item correctly?

Facilitator Script: Round 1 Discussion; Begin Round 2

Goals: Review Round 1 results. Ensure all panelists are prepared to begin Round 2.

Materials of Importance: Round 1 results (tables, graphs,); Round 2 Readiness Form; Round 2 Ordered-Item Booklet; Round 2 Bookmark Rating Forms.

Facilitator Outline:

6. Discuss Round 1 results.
7. Have panelists complete Round 2 Readiness Form and begin Round 2.
8. Monitor Round 2 and be available to answer panelists' questions.
9. End Round 2 and collect panelists' secure materials.
10. Bring completed Bookmark Rating Forms to data analysts.

Facilitator Talking Points:

- 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. If there was a wide divergence of opinions 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 Attained bookmark recommended in Round 1. Similarly, the last page in the OIB that they will review for Round 2 will be the highest recommended Surpassed bookmark.
    - Panelists should place their 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*** Attained students have a 67% chance of answering the item correctly?
    - After placing the Attained bookmark, would ***Just Barely*** Surpassed students have a 67% chance of answering the item correctly?
- Ask if the panelists have any questions and ensure they are ready to begin Round 2.

Facilitator Script: Round 2 Discussion; Begin Round 3

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

Materials of Importance: Round 2 results (tables, graphs, and impact data); Round 3 Readiness Form; Bookmark Rating Forms; Ordered-Item Booklet.

Facilitator Outline:

1. Distribute then discuss Round 2 results.
2. Have panelists complete Round 3 Readiness Form and begin Round 3.

3. Monitor Round 3 and be available to answer panelists' questions.
4. End Round 3 and collect panelists' secure materials.
5. Bring completed Bookmark Rating Forms to data analysts.

Facilitator Talking Points:

- 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. If there was a wide divergence of opinions specifically ask panelists from both ends of the spectrum to explain their reasoning.
- 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 three performance categories seem realistic?
    - How do the percentages compare across grades?
- Explain to panelists the Round 3 process:
    - Round 3 will be more targeted. Panelists will start Round 3 on the lowest recommended Attained bookmark recommended in Round 2. Similarly, the last page in the OIB that they will review for Round 3 will be the highest recommended Surpassed bookmark from Round 2.
    - Panelists should place their bookmarks using the same process employed in Rounds 1 and 2.
    - 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*** Attained students have a 67% chance of answering the item correctly?
    - After placing the Attained bookmark, would ***Just Barely*** Surpassed students have a 67% chance of answering the item correctly?
- Ask if the panelists have any questions and ensure they are ready to begin Round 3.

Facilitator Script: Review; Wrap-up

Goals: Review and Revise Round 3 results for Functional Independence tests. Ensure panelists complete the Final Evaluation Form. Inform panelists of Day 4 assignments.

Materials of Importance: Round 3 results (tables, graphs, and impact data) for the FI test; Final Evaluation Form.

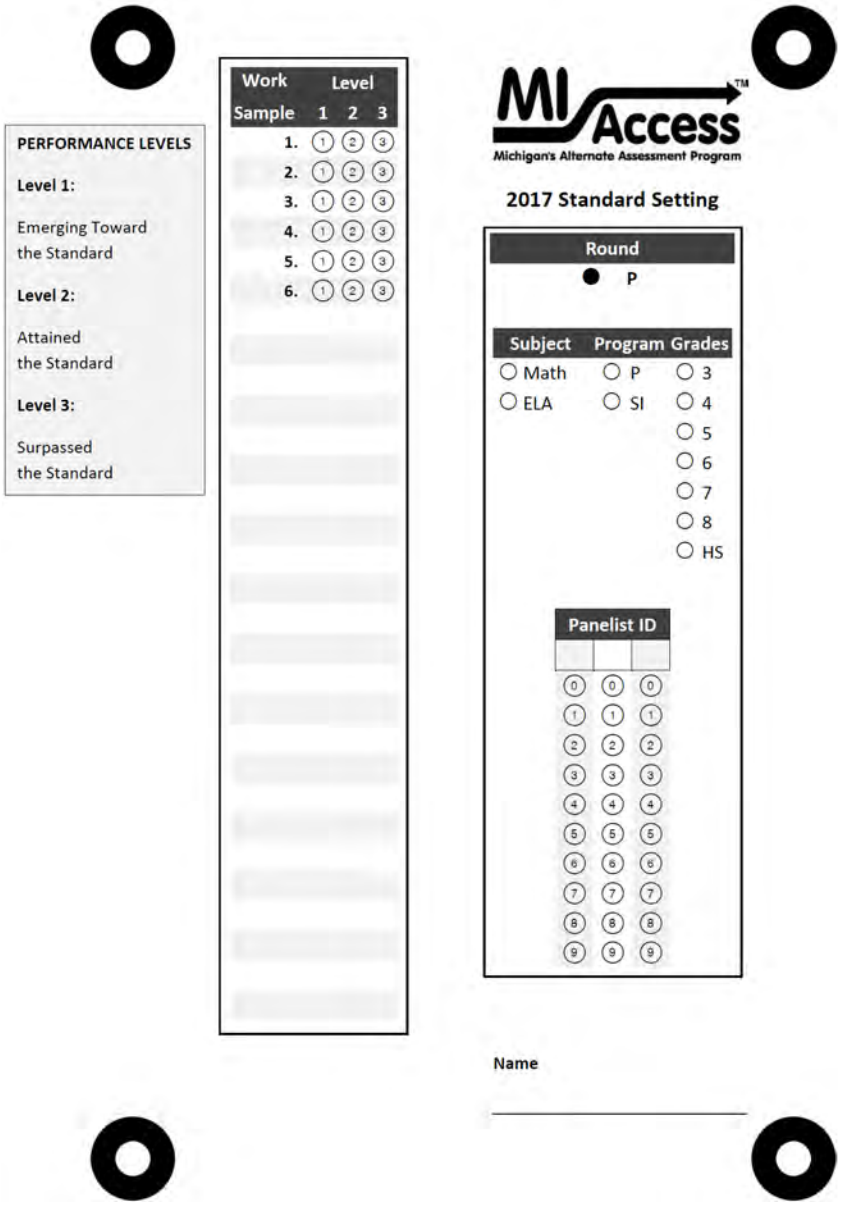
Facilitator Outline:

1. Review Round 3 results for FI tests.
2. Revise Round 3 results for FI tests.
3. Have panelists complete Final Evaluation Form.
4. Inform panelists of Day 4 room assignments.
5. Dismiss panelists for the evening.
6. Bring completed Bookmark Rating Forms to data analysts.

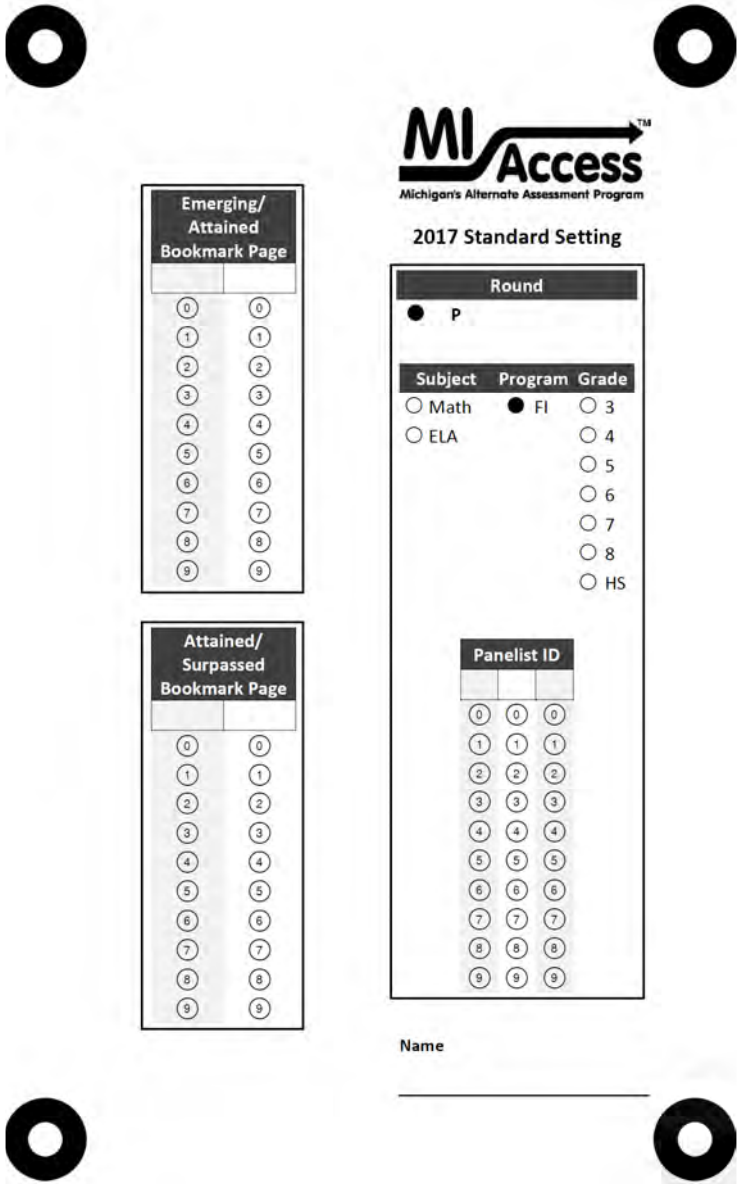
Facilitator Talking Points:

- Review the Round 3 results and impact data. Focus panelists' attention on whether the percentages of students in the three performance categories seem realistic. How do the percentages compare across grades?
- Explain to panelists the purpose of tomorrow's activities. The Vertical Articulation group will evaluate all cutscores and their impact across grades and make recommended changes based on the corresponding PLDs. The Wrap-Up Group will interact with MDE personnel in a question-answer-discuss session.
- Distribute Final Evaluation Forms and collect when all panelists have finished.
- Thank panelists for their work.

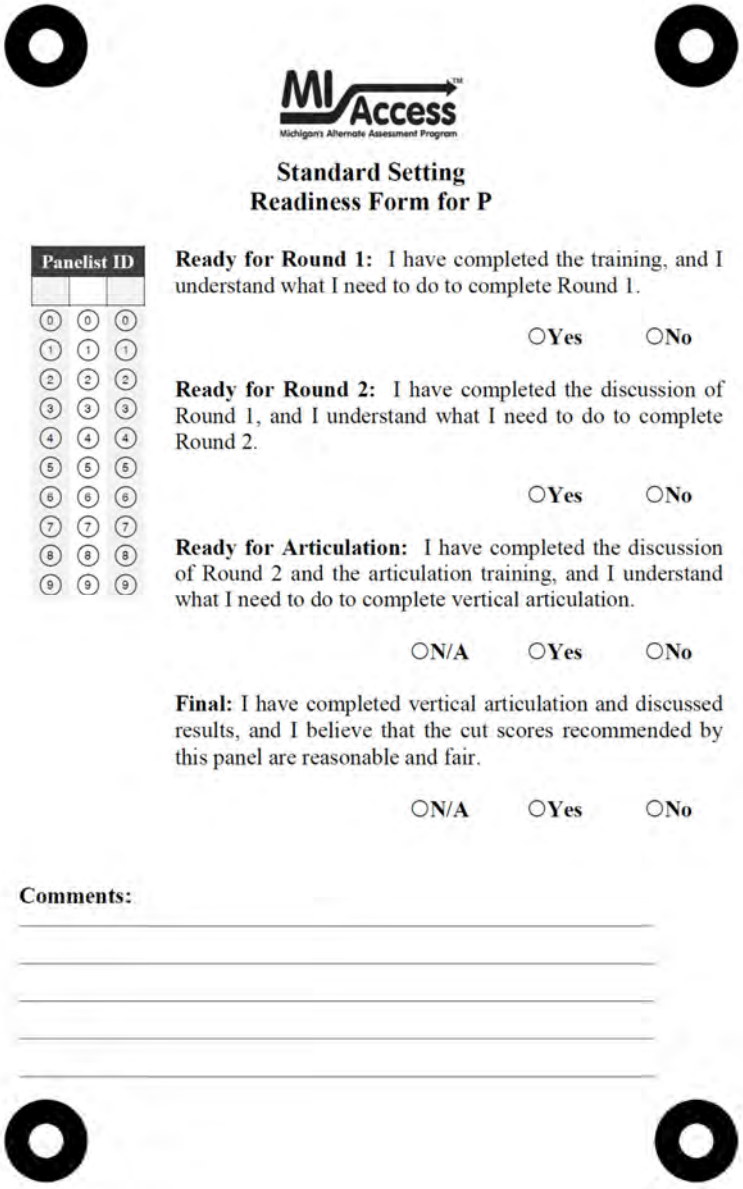
Practice Round Form for Participation and Supported Independence



Practice Round Form for Functional Independence



Readiness Form for Participation



MI Access
Michigan's Alternate Assessment Program

Standard Setting
Readiness Form for P

Panelist ID		
0	0	0
1	1	1
2	2	2
3	3	3
4	4	4
5	5	5
6	6	6
7	7	7
8	8	8
9	9	9

Ready for Round 1: I have completed the training, and I understand what I need to do to complete Round 1.

○**Yes** ○**No**

Ready for Round 2: I have completed the discussion of Round 1, and I understand what I need to do to complete Round 2.

○**Yes** ○**No**

Ready for Articulation: I have completed the discussion of Round 2 and the articulation training, and I understand what I need to do to complete vertical articulation.

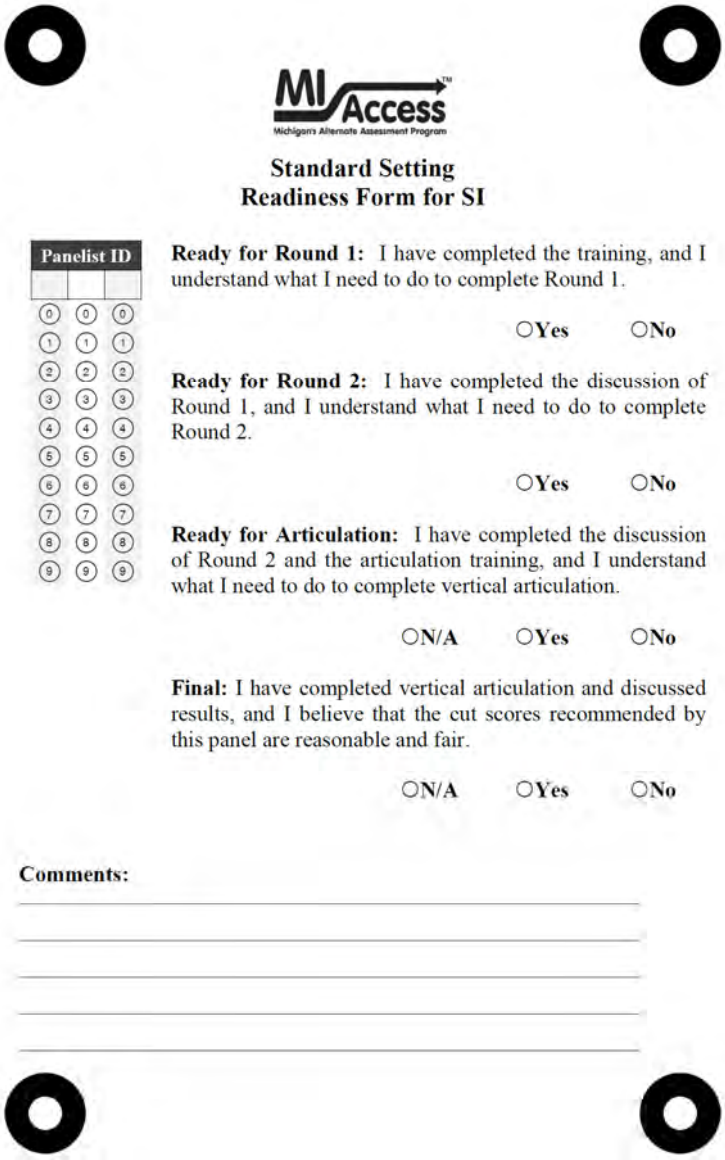
○**N/A** ○**Yes** ○**No**

Final: I have completed vertical articulation and discussed results, and I believe that the cut scores recommended by this panel are reasonable and fair.

○**N/A** ○**Yes** ○**No**

Comments:

Readiness Form for Supported Independence



MI Access
Michigan's Alternate Assessment Program

Standard Setting
Readiness Form for SI

Panelist ID		
0	0	0
1	1	1
2	2	2
3	3	3
4	4	4
5	5	5
6	6	6
7	7	7
8	8	8
9	9	9

Ready for Round 1: I have completed the training, and I understand what I need to do to complete Round 1.

○Yes ○No

Ready for Round 2: I have completed the discussion of Round 1, and I understand what I need to do to complete Round 2.

○Yes ○No

Ready for Articulation: I have completed the discussion of Round 2 and the articulation training, and I understand what I need to do to complete vertical articulation.

○N/A ○Yes ○No

Final: I have completed vertical articulation and discussed results, and I believe that the cut scores recommended by this panel are reasonable and fair.

○N/A ○Yes ○No

Comments:

Readiness Form for Functional Independence

MI Access
Michigan's Alternate Assessment Program

Standard Setting
Readiness Form for FI

Panelist ID		
0	0	0
1	1	1
2	2	2
3	3	3
4	4	4
5	5	5
6	6	6
7	7	7
8	8	8
9	9	9

Ready for Round 1: I have completed the training, and I understand what I need to do to complete Round 1.

○**Yes** ○**No**

Ready for Round 2: I have completed the discussion of Round 1, and I understand what I need to do to complete Round 2.

○**Yes** ○**No**

Ready for Round 3: I have completed the discussion of Round 2, and I understand what I need to do to complete Round 3.

○**Yes** ○**No**

Ready for Articulation: I have completed the discussion of Round 3 and the articulation training, and I understand what I need to do to complete vertical articulation.

○**N/A** ○**Yes** ○**No**

Final: I have completed vertical articulation and discussed results, and I believe that the cut scores recommended by this panel are reasonable and fair.

○**N/A** ○**Yes** ○**No**

Comments:

Process Evaluation Form for Participation

Standard Setting
Final Evaluation Form for P

Facilitator_____

Directions: Fill in a circle for each of the following statements corresponding to your opinion (Strongly Agree [SA], Agree [A], Neutral [N], Disagree [D], or Strongly Disagree [SD]). If you have any additional comments, please write them in the space provided at the end of this form.

	Statement	SA	A	N	D	SD
1	Overall, the facilities and food service helped to create a good working environment.	○	○	○	○	○
2	Overall, the training in the standard-setting purpose and methods was clear.	○	○	○	○	○
3	Overall, I am confident that I was able to apply the standard setting methods appropriately.	○	○	○	○	○
4	Overall, the standard setting procedures allowed me to use my experience and expertise to recommend cut scores for the tests.	○	○	○	○	○
5	Overall, the facilitator helped to ensure that everyone was able to contribute to the group discussions and that no one unfairly dominated the discussions.	○	○	○	○	○
6	Overall, I was able to understand and use the feedback provided (e.g., other participants' ratings, impact data).	○	○	○	○	○
7	I believe that the final group-recommended cut score fairly represents the minimal level of performance for students at the **Attained** level.	○	○	○	○	○
If you answered Disagree or Strongly Disagree to Question 7, do you believe the final group-recommended cut score for **Attained** is: ○ Too High, or ○Too Low						
8	I believe that the final group-recommended cut score fairly represents the minimal level of performance for students at the **Surpassed** level.	○	○	○	○	○
If you answered Disagree or Strongly Disagree to Question 8, do you believe the final group-recommended cut score for Advanced **Surpassed** is: ○Too High, or ○Too Low						

Comments:

_____

_____

_____

_____

_____

Thank you! When you have completed this form, please return it to your facilitator.

Process Evaluation Form for Supported Independence

Standard Setting
Final Evaluation Form for SI

Facilitator_____

Directions: Fill in a circle for each of the following statements corresponding to your opinion (Strongly Agree [SA], Agree [A], Neutral [N], Disagree [D], or Strongly Disagree [SD]). If you have any additional comments, please write them in the space provided at the end of this form.

	Statement	SA	A	N	D	SD
1	Overall, the facilities and food service helped to create a good working environment.	○	○	○	○	○
2	Overall, the training in the standard-setting purpose and methods was clear.	○	○	○	○	○
3	Overall, I am confident that I was able to apply the standard setting methods appropriately.	○	○	○	○	○
4	Overall, the standard setting procedures allowed me to use my experience and expertise to recommend cut scores for the tests.	○	○	○	○	○
5	Overall, the facilitator helped to ensure that everyone was able to contribute to the group discussions and that no one unfairly dominated the discussions.	○	○	○	○	○
6	Overall, I was able to understand and use the feedback provided (e.g., other participants' ratings, impact data).	○	○	○	○	○
7	I believe that the final group-recommended cut score fairly represents the minimal level of performance for students at the **Attained** level.	○	○	○	○	○
If you answered Disagree or Strongly Disagree to Question 7, do you believe the final group-recommended cut score for **Attained** is: ○ Too High, or ○Too Low						
8	I believe that the final group-recommended cut score fairly represents the minimal level of performance for students at the **Surpassed** level.	○	○	○	○	○
If you answered Disagree or Strongly Disagree to Question 8, do you believe the final group-recommended cut score for Advanced **Surpassed** is: ○Too High, or ○Too Low						

Comments:

_____

_____

_____

_____

_____

Thank you! When you have completed this form, please return it to your facilitator.

Process Evaluation Form for Functional Independence

MI Access
Michigan's Alternate Assessment Program

Standard Setting
Final Evaluation Form for FI

Facilitator_____

Directions: Fill in a circle for each of the following statements corresponding to your opinion (Strongly Agree [SA], Agree [A], Neutral [N], Disagree [D], or Strongly Disagree [SD]). If you have any additional comments, please write them in the space provided at the end of this form.

	Statement	SA	A	N	D	SD
1	Overall, the facilities and food service helped to create a good working environment.	○	○	○	○	○
2	Overall, the training in the standard-setting purpose and methods was clear.	○	○	○	○	○
3	Overall, I am confident that I was able to apply the standard setting methods appropriately.	○	○	○	○	○
4	Overall, the standard setting procedures allowed me to use my experience and expertise to recommend cut scores for the tests.	○	○	○	○	○
5	Overall, the facilitator helped to ensure that everyone was able to contribute to the group discussions and that no one unfairly dominated the discussions.	○	○	○	○	○
6	Overall, I was able to understand and use the feedback provided (e.g., other participants' ratings, impact data).	○	○	○	○	○
7	I believe that the final group-recommended cut score fairly represents the minimal level of performance for students at the **Attained** level.	○	○	○	○	○
If you answered Disagree or Strongly Disagree to Question 7, do you believe the final group-recommended cut score for **Attained** is: ○ Too High, or ○Too Low						
8	I believe that the final group-recommended cut score fairly represents the minimal level of performance for students at the **Surpassed** level.	○	○	○	○	○
If you answered Disagree or Strongly Disagree to Question 8, do you believe the final group-recommended cut score for Advanced **Surpassed** is: ○Too High, or ○Too Low						

Comments:

_____

_____

_____

_____

_____

Thank you! When you have completed this form, please return it to your facilitator.

Vertical Articulation Motions and Actions Form

Motion	Time	Second	Vote	Result/ Action

Vertical Articulation Evaluation Form

Facilitator_____

Directions: Fill in a circle for each of the following statements corresponding to your opinion (Strongly Agree [SA], Agree [A], Neutral [N], Disagree [D], or Strongly Disagree [SD]). If you have any additional comments, please write them in the space provided at the end of this form.

	Statement	SA	A	N	D	SD
1	Overall, the training for this task was clear.	○	○	○	○	○
2	The tables and graphs helped me keep track of the cut scores and the impact of the decisions we were making.	○	○	○	○	○
3	The facilitator was effective in guiding discussion and keeping it moving toward a decision.	○	○	○	○	○
4	The facilitator helped to ensure that everyone was able to contribute to the group discussions and that no one unfairly dominated the discussions.	○	○	○	○	○
5	I had access to information I needed to make decisions about cut scores.	○	○	○	○	○
6	Overall, I am confident that I was able to participate in this activity appropriately.	○	○	○	○	○
7	The process was fair.	○	○	○	○	○
8	Overall, I was able to understand and use the feedback provided (e.g., other participants' ratings, impact data).	○	○	○	○	○
9	I believe that the final, articulated cut scores fairly represent expectations across grades at the **Attained** level.	○	○	○	○	○
10	I believe that the final, articulated cut scores fairly represent expectations across grades at the **Surpassed** level.	○	○	○	○	○

Comments:

_____

_____

_____

_____

_____

Thank you! When you have completed this form, please return it to your facilitator.

Appendix B

Evaluation Comments

- Evaluation Comments From Participation Panelists
- Evaluation Comments From Supported Independence Panelists
- Evaluation Comments From Functional Independence Panelists
- Evaluation Comments From Vertical Articulation Committee Members

Evaluation Comments From Participation Panelists

- Dan was great!
- Healthier breakfast choice. Closer hotel
- Color coded paper to help keep packet more organized; Use less paper more digital!
- Dan is an amazing facilitator. The day 2 option for lunch was much better. More meals like that would be amazing. It would be nice to have water or snack in a downstairs location as much is spent then upstairs
- Color coded paper for SI and P or by grade but not all white. Provide paperclips for organization. Condition codes on the zero scores would be helpful too. Thx.
- While the facilities were appropriate the lack of non-sugary items for breakfast and the lack of beverages provided during the sessions was unacceptable. Water could have been provided at the tables during each session. Also snacks during the midday.
- Facility kept too cold. More fruit/vegies at meals. Bagels or bread for peanut butter/yogurt at breakfast for protein source
- Craig did a nice job of running our group. I appreciate how he did it.
- Snack in the afternoon would be good. Water available in each room or nearby?
- I felt that this was a great experience. It was interesting to go through this process.
- It was good to hear others point of view. Lidia was great at facilitator.
- Nice job! The first two days were informative and enjoyable.
- I really enjoyed completing this setting and the MDE presentation yesterday! I hope to participate in future committees!
- Lidia did an excellent job as a facilitator and ensured that all panelists were represented equally.
- Loved the second day food choice better than the first. Buffet/Sandwiches are better for a conference. More choices for drinks/snacks.
- Winnie did an excellent job of keeping us on track! Meals could be improved.
- Winnie was excellent facilitator. The food for lunch was not very appetizing.
- Facilities were okay. Food 1st day was not good. Chicken tough - no fruit available 2nd day (make your own sandwich) much better still no fruit.
- Smooth - Best standard setting panel I have attended
- The PLDs for "Participation" population seem a bit skewed/high. I believe students who should be taking the participation level assessment, should score at emerging.. If these students fit in the "attained" PLDs they should be taking SI level - But "doubling" anything (surpassed) is a higher level skill anyway, let alone ordinal terms to identify positions in patterns.
- The food is awful :(
- Food - Monday dirty and without lettuce. Tuesday was better. Heavy processed carbs @ breakfast. Fresh fruit, protein would be nice.
- Excellent, positive facilitator! Great to work with Jennie.
- The A/C was too cold and the food was poor. I also think that the hotel is too far away (added traffic, stress, etc.) Jennie is an amazing facilitator!
- Lansing center was great. The food was not the best. It was rather bland and unappealing. The direction given by the facilitator was clear and kept the group on focus.
- Great leadership - great team!

- No cookies, bad breakfast and lunch. No coffee out in the afternoon. No pop. Jennie was awesome!
- The purpose and expectations for this event were clear and well executed. The building that we are in is extremely cold and the food served was subpar and unappetizing. My facilitator was great and knowledgeable.
- The Lansing center was way too cold! The food that was served was (for the most part) un-healthy options. Two years ago there were cookies + refreshments in the afternoon. Also the Lansing center should offer free wifi to people who are at a conference instead of making people pay.
- The temperature was consistently set too low for comfortability. Concentration became more about how to stay warm than how to best apply our skills. It should be mentioned in emails prior - especially for people travelling, that jackets, sweatshirts, and blankets may be needed to stay warm. In the month of June no one expects the facilities temperatures /or the room below set to 58 degree. Whether it was blowing in cold air from outside or not make your people comfortable. They are happier and willing to take more time. It's pretty bad when people are eager to leave just to get warm.
- Rooms to cold. Food OK
- I thank you for inviting me to join this process.
- The only "problem" is the rooms are extremely cold.
- Our room was very cold. Wifi would be helpful. Packets separated by P + SI would be helpful. Or order of presentation
- Room was a little cold which led to distractibility
- Too many papers. Would have been easier if color coded. Facilities very cold
- With the PLD chart as a reminder, focus, this process allowed me to recommend and select scores are a level I felt confident
- Sara insured that we were engaged in the process. She pushed our thinking + encouraged us to step outside of our comfort zone. She also allowed for discussion which allowed us to hear one another point of view.
- Working with other professionals was very interesting and made me reflect on my own style of teaching; I felt I have a better understanding of these tests. Thank you Sara.
- Overall this has been excellent! Only problem for me is the room is very cold.

Evaluation Comments From Supported Independence Panelists

- Prefer hotel located next to Lansing Center. Simple lunches would be better: salad bar, taco bar, soup, sandwiches. Breakfast variety is needed: yogurt, fruit, boiled eggs, etc.
- Other than Tuesday, the lunch options were horrible. Breakfast was all carbs, could there be heathier options? Also, the commute to the hotel was too much with morning traffic. No matter how early I left (1 hour before 8:30 start), I was unable to avoid the dense traffic. It was really frustrating to have to travel far for lodging. Also, it would have been nice to have water/drinks provided in individual rooms at the lower level. Dan was a great facilitator, very knowledgeable and good at having discussions. Thank you!
- Great group and Dan was very good!
- Digital!! Less Paper!!

- Again my only complaint was the quality of the food served.
- Room too cold. More fruits and veggies in lunch. Protein source for breakfast.
- Craig did a great job.
- Facilities: cold. Food: Not very good.
- Small binder for material would be great.
- Lidia was a terrific facilitator and was very professional, courteous and respectful!
- I definitely felt more comfortable the 2nd day and beyond. Lidia did well to keep our group on track and focused. It was a great experience!
- Overall, a fantastic experience! Would love to participate again!
- Overall, this was a good experience. I enjoyed looking at the test and comparing results of students. Well done.
- I enjoyed working on the standard setting this week! The process and information from the MDE session was very helpful! Thank you for the opportunity to participate!
- She was great.
- Winnie was an excellent facilitator. Time allocated to finish was excessive. Food was awful!
- Food choices were not done according to my diet. That is the only complaint of the whole presentation/days. Winnie was excellent.
- This was a great experience and opportunity to understand the test more.
- Everything was well organized and facilitated nicely. The only thing I would comment on is having more lunch options for participants.
- Winnie was an outstanding evaluator. The food was not very good.
- Thank you for allowing me to participate.
- Food was not very good. Diets were not followed. I liked the selection of sandwich so that all diet needs were met.
- Kellogg Center next time? Food service was horrible!
- Enjoyed this experience and appreciate the opportunity to be a part of the procedure.
- Maybe at future events, this could be held at the Kellogg Center? The rooms were dark and not conducive to a great working environment. Thank you!
- The collaborative efforts of the educators made this experience valuable, interesting and fun. Kellogg Center would provide a better environment and food. If it must be here - maybe give everyone $5 for lunch daily to go out? Food here is terrible. That would have to be cheaper in the long run.
- Amy was great.
- Our breakout room was cold enough to have to bring in coats and blankets. Nothing snack were at breaks. Coffee/tea did not stay in room for morning breaks. Food - not healthy.
- Jennie was a great facilitator. She encouraged everyone to participate and considered everyone's perspective. The Lansing Center was freezing, bad limited internet access. The food was disgusting and it was located much too far from the hotel we stayed in. Also, the way our many documents were just given to us in no order was completely unorganized. The documents need to be organized in a binder, color located, etc.
- The Lansing Center had too many hidden costs for the MDE this year. Internet access extra $10 per person. Coffee/drinks only out at a certain time, otherwise it was extra. NO drinks set out in the afternoon. The temperature was extremely frigid. How can the department expect the best when conditions are poor? We literally had to wrap

ourselves in blankets, extra sweatshirts, coats and such AND STILL BE COLD!!! That is ridiculous! There has to be alternate venue where you would get more value for your money. A hotel, the Kellogg Center, MSU campus? Good Lord, any other place would benefit the MDE. I believe we did our best with the standard setting, under the conditions we were under.

- Having to drive 10 miles to and from the hotel was very inconvenient. Food and amenities available at the Lansing Center was poor. Jennie was a great facilitator and Dan rocks. Thanks for the treats.
- A/C too cold. Need wifi. Hotel too far away (added stress, traffic, time). Food poor - not diverse. Jennie was compassionate, open-minded, funny, and personal. A wonderful facilitator.
- The Lansing Center was not at all accommodating!
- The food was subpar and the air conditioning was too cold.
- Too cold. Lunch was more like dinner (too much)
- Veggie food was awful!
- The rooms are always cold.
- It was way too cold on Tuesday and Wednesday. The food was not good. Corey Palermo was amazing! John Jaquith kept us on track. I always love seeing and having meaningful conversations with Jennie!
- Post lunch for the day for individuals to decide if they will stay or dine out.
- I had a great time being able to talk others and share resources. Everything was well organized. Great job!
- Training itself was great. Suggestions: tell people to dress warm or turn the AC down; tell people what is on the menu so we are prepared. When doing lunch, please provide lunch type food. Lunch was too heavy. Sara Sliver was fantastic. She was very knowledgeable and I enjoyed her.
- Cold temp. Food was just ok. Afternoon snack would help. Coffee should be available all day. Staff rushed me to drink up coffee and ice tea at lunch.

Evaluation Comments From Functional Independence Panelists

- There is something wrong with a process where on 11th grader has an easier time getting a surpassed score than someone in another grade. These scores across the ELA grades are radically different and will not, in my opinion, reflect accurate results. I would recommend the same group looking at each ELA test.
- Food cold/lukewarm. Good sandwiches.
- It would be great if fruit and or yogurt could have been made available at breakfast.
- This process was very interesting to me! I enjoyed the process and would love to do this again.
- Very informative. Great group of educators varied experiences, local of schools. Breakfast could/should have been more substantial. Would be helpful to have possible carpool info from people from our area.
- Karen did a great job! After setting scores 2 years ago, I believe that changes that are represented are a result of the change from paper pencil to technology.
- Breakfast food needed to be more than sweets.
- Would appreciate lighter/healthy lunch and protein choice in breakfast. Yesterday lunch was good. Disappointed to hear other groups had so much down time.

- I believe it would be helpful to provide the previous grade level PLD for reference. This would allow us to better understand prerequisite skills.
- Food was horrible. I am not 100% sure the gluten free items were gluten free. I ended up eating out and not covered, had to pay out of pocket. Parking was hard to find; directions could be much better. Way too much down time! Too many breaks!
- The overall standard setting experience was enlightening. The only issue is in regards to the food options that were available. There should have been more appropriate breakfast items. There should be healthier options.
- Directions were very repetitive; lots of extra time throughout the days; several groups didn't need to report until lunch on Wed. Food choices for breakfast time limited (no protein, fruit, juice, etc.)
- The breakfast was horrible! I would like to have fruit, cereal bar, protein bar and or yogurt! Way to much sugar! Like our students, I crashed around 10:00 am and needed to eat breakfast out so that wouldn't happen! Lunch was great 1 day also! Not happy. Facilities beautiful.
- The meals were once again poor nutritionally. These needs to be fruit and protein options for breakfast. Lunch should be less heavy.
- The Radisson rooms were not clean and seem extremely overpriced for amenities. Do not like hotel! Food would it be less expensive to serve box lunches and let us eat around? The conference food was not good. The work we did was great though and well organized informative.
- The food was good most of the time. I would like to see changes to breakfast. Possibly fruit and something healthier. More drink choices (and ice tea was great!)
- Minor suggestion: fruit, cereal and yogurt for breakfast. Maybe give a choice about lunch on registration sheet. Offer lunch at hotel or lunch on own with $8.50 reimbursement. If not stick to sandwich/soup/salad/wraps for lunch food plus fruit please.
- Overall, I think the process is great! However, I think rounds 2 and 3 need less time than round 1. There is still too much down time.
- Schedule should be more flexible. Breakfast should include fruit.
- Good experience and a great opportunity to participate with the MDE. Good job!
- I fell the food could have been a bit better quality. I also wish there was a dessert or sweet snack in the afternoon.
- Rooms were cold. Training was well explained. Group encouraged discussion, sharing, and collaboration.
- There are far too many students that are surpassed in the 4th grade.
- Facility was amazing. Food was lacking. Elementary needs to be 3 days with 2 grades. High school needs to be 2 days with 1 grade.
- Winnie was fantastic to work with! Excellent knowledge and facilitation. I also enjoyed Michael's presentation on day 1. The process was interesting and challenging. I would strongly prefer having some options for breakfast that includes protein and fruit instead of any sweets/carbs.
- I loved going through this process - would love to do it again.
- great session, looking forward to next year!
- I thought the entire process ran more smoothly than it did 2 years ago. My group (Math 3-4) was very engaged and Winnie helped keep us on point while respecting our opinions at the same time.
- I thought the entire process was great!

- Facilities awesome! Food not so much. Chris did a great job of facilitating our group. Sweet treat would be delightful. Candy/Chocolate on the tables please. Dessert at lunch. Fruit at breakfast.
- Chris was very nice. I enjoyed the process. The food was not that great but not a huge deal. Thanks for including me.
- Would like afternoon snack and drinks. Very nice conference!
- I found it interesting that the person running our group was not affiliated with MDE; nor had much knowledge on MI-Access standards or requirements. There is no one in the state who could have facilitated? Chris was a great facilitator but someone that does what we do would be helpful.
- This was an excellent training! The staff at Radisson were also amazing!
- Great job, Chris. It was a pleasure working with you!
- The food was not very good and didn't meet the nutritional standards. Fidgets at the tables would be helpful.
- Honestly, the food was not great. The experience overall was positive.
- Dan was a great facilitator. He made this process exciting, understandable and thought provoking.
- Food breakfast could have been substantial - not just pastries - something healthy. Lunch was not very good.
- Overall, the procedures and expectations were very clear. The process used seemed appropriate and well thought out. In regards to food service, the breakfast could be improved by offering items that are more sustainable.
- Dan did a very job keeping all participants engaged and on task. Nice job Dan! Also I loved the cookies.
- I found the experience very challenging and educational. I would certainly register again when it is held.
- This was a great experience. It was great to hear a lot of different perspectives. I am grateful to have participated in this group. Dan did an awesome job. He is a model facilitator. He is very skilled in the art of true unbiased facilitation. He is kind and easy to relate to. Thank you for providing a vegan meal option.
- I really enjoyed my experience here, and I would gladly do it again. This was my first time doing something like this and I thought it was very interesting.
- Dan was a wonderful facilitator. Well-read and knowledgeable about data! Food and beverage selections could be improved.
- Thank you for allowing me to be a part of this process. I found the whole thing to be quite fascinating.
- I found this process to be very interesting!

Evaluation Comments From Vertical Articulation Committee Members

- Well presented and very clear expectations.
- Nice job keeping things moving and appropriate.
- Copy of test questions for each participant would be help.
- Great experience!
- Very interesting process.
- Thank you for selecting me to be a part of this process. I found it very interesting.
- Great!

- Thank you!
- Craig needs an updated computer to help his presentation.
- Great job Craig! Thank you!
- Craig did a great job of facilitating and getting input from multiple people at multiple tables. He was very unbiased and a great facilitator. Overall, a very great experience.
- This whole process was very beneficial.
- We did great!

Appendix C

PowerPoint Presentations

- Overview for Participation/Supported Independence — Field Code Changed
- The Body of Work Procedure — Field Code Changed
- Vertical Articulation Training for Participation/Supported Independence — Field Code Changed
- Overview for Functional Independence — Field Code Changed
- The Bookmark Procedure — Field Code Changed
- Vertical Articulation Training for Functional Independence — Field Code Changed

Appendix F. MI-Access FI Raw to Scale Score Conversion Tables

Table F-1. ELA Grade 3 Raw to Scale Score Conversion Table

TotalRawScore	Theta	SETheta	SS	SESS	PL
0	-5.682	1.894	2200	31	1
1	-4.316	1.099	2220	18	1
2	-3.445	0.807	2234	13	1
3	-2.908	0.67	2243	11	1
4	-2.516	0.588	2250	10	1
5	-2.203	0.533	2255	9	1
6	-1.940	0.494	2259	8	1
7	-1.711	0.465	2263	8	1
8	-1.505	0.443	2267	7	1
9	-1.316	0.426	2270	7	1
10	-1.140	0.413	2273	7	1
11	-0.975	0.402	2275	7	1
12	-0.817	0.394	2278	7	1
13	-0.664	0.387	2280	6	1
14	-0.516	0.383	2283	6	1
15	-0.371	0.379	2285	6	1
16	-0.228	0.378	2288	6	1
17	-0.086	0.377	2290	6	1
18	0.057	0.378	2292	6	1
19	0.200	0.38	2295	6	1
20	0.346	0.384	2297	6	1
21	0.495	0.389	2300	6	2
22	0.648	0.395	2302	7	2
23	0.808	0.404	2305	7	2
24	0.976	0.415	2308	7	2
25	1.153	0.429	2311	7	2
26	1.345	0.446	2314	7	2
27	1.553	0.469	2317	8	2
28	1.786	0.498	2321	8	3
29	2.053	0.536	2326	9	3
30	2.368	0.59	2331	10	3
31	2.763	0.671	2337	11	3
32	3.298	0.804	2346	13	3
33	4.159	1.092	2361	18	3
34	5.510	1.887	2383	31	3

Table F-2. FI ELA Grade 4 Raw to Scale Score Conversion Table

TotalRawScore	Theta	SETheta	SS	SESS	PL
0	-5.452	1.908	2300	33	1
1	-4.052	1.118	2325	19	1
2	-3.151	0.819	2340	14	1
3	-2.600	0.678	2349	12	1
4	-2.199	0.593	2356	10	1
5	-1.882	0.537	2362	9	1
6	-1.615	0.497	2366	8	1
7	-1.383	0.468	2370	8	1
8	-1.174	0.446	2374	8	1
9	-0.983	0.429	2377	7	1
10	-0.806	0.415	2380	7	1
11	-0.638	0.404	2383	7	1
12	-0.478	0.396	2386	7	1
13	-0.324	0.39	2388	7	1
14	-0.174	0.385	2391	7	1
15	-0.027	0.382	2393	7	1
16	0.119	0.38	2396	6	1
17	0.263	0.379	2398	6	1
18	0.407	0.38	2401	6	2
19	0.552	0.382	2403	7	2
20	0.700	0.386	2406	7	2
21	0.850	0.391	2408	7	2
22	1.005	0.397	2411	7	2
23	1.167	0.406	2414	7	2
24	1.336	0.417	2417	7	2
25	1.515	0.431	2420	7	2
26	1.708	0.448	2423	8	3
27	1.918	0.47	2426	8	3
28	2.152	0.498	2430	9	3
29	2.417	0.535	2435	9	3
30	2.729	0.585	2440	10	3
31	3.112	0.658	2447	11	3
32	3.621	0.779	2456	13	3
33	4.423	1.054	2469	18	3
34	5.705	1.856	2491	32	3

Table F-3. FI ELA Grade 5 Raw to Scale Score Conversion Table

TotalRawScore	Theta	SETheta	SS	SESS	PL
0	-5.531	1.917	2400	33	1
1	-4.108	1.13	2420	20	1
2	-3.188	0.827	2436	14	1
3	-2.629	0.681	2446	12	1
4	-2.227	0.594	2453	10	1
5	-1.909	0.537	2459	9	1
6	-1.643	0.496	2463	9	1
7	-1.412	0.467	2467	8	1
8	-1.205	0.444	2471	8	1
9	-1.016	0.427	2474	7	1
10	-0.840	0.413	2477	7	1
11	-0.674	0.402	2480	7	1
12	-0.515	0.394	2483	7	1
13	-0.363	0.388	2486	7	1
14	-0.214	0.383	2488	7	1
15	-0.068	0.38	2491	7	1
16	0.075	0.379	2493	7	1
17	0.219	0.379	2496	7	1
18	0.362	0.38	2498	7	1
19	0.507	0.382	2501	7	2
20	0.655	0.387	2503	7	2
21	0.807	0.392	2506	7	2
22	0.964	0.4	2509	7	2
23	1.128	0.41	2512	7	2
24	1.301	0.423	2515	7	2
25	1.486	0.439	2518	8	2
26	1.687	0.459	2521	8	3
27	1.910	0.485	2525	8	3
28	2.162	0.519	2530	9	3
29	2.455	0.565	2535	10	3
30	2.808	0.627	2541	11	3
31	3.256	0.715	2549	12	3
32	3.861	0.849	2559	15	3
33	4.794	1.12	2576	20	3
34	6.178	1.895	2600	33	3

Table F-4. FI ELA Grade 6 Raw to Scale Score Conversion Table

TotalRawScore	Theta	SETheta	SS	SESS	PL
0	-5.429	1.911	2500	33	1
1	-4.021	1.122	2526	20	1
2	-3.112	0.822	2542	14	1
3	-2.557	0.68	2551	12	1
4	-2.155	0.595	2558	10	1
5	-1.835	0.539	2564	9	1
6	-1.567	0.499	2569	9	1
7	-1.332	0.471	2573	8	1
8	-1.121	0.449	2577	8	1
9	-0.927	0.432	2580	8	1
10	-0.746	0.419	2583	7	1
11	-0.575	0.409	2586	7	1
12	-0.411	0.401	2589	7	1
13	-0.252	0.395	2592	7	1
14	-0.098	0.391	2594	7	1
15	0.054	0.388	2597	7	1
16	0.204	0.387	2600	7	1
17	0.354	0.386	2602	7	1
18	0.503	0.387	2605	7	1
19	0.654	0.39	2608	7	2
20	0.807	0.393	2610	7	2
21	0.964	0.399	2613	7	2
22	1.125	0.406	2616	7	2
23	1.294	0.415	2619	7	2
24	1.470	0.426	2622	7	2
25	1.658	0.441	2625	8	2
26	1.860	0.459	2629	8	3
27	2.081	0.483	2633	8	3
28	2.329	0.514	2637	9	3
29	2.614	0.555	2642	10	3
30	2.953	0.611	2648	11	3
31	3.375	0.692	2655	12	3
32	3.939	0.821	2665	14	3
33	4.820	1.096	2680	19	3
34	6.170	1.883	2700	33	3

Table F-5. FI ELA Grade 7 Raw to Scale Score Conversion Table

TotalRawScore	Theta	SETheta	SS	SESS	PL
0	-4.948	1.842	2600	33	1
1	-3.702	1.029	2629	19	1
2	-2.950	0.748	2643	13	1
3	-2.485	0.627	2651	11	1
4	-2.137	0.557	2658	10	1
5	-1.854	0.51	2663	9	1
6	-1.611	0.478	2667	9	1
7	-1.395	0.453	2671	8	1
8	-1.198	0.435	2674	8	1
9	-1.016	0.42	2678	8	1
10	-0.844	0.409	2681	7	1
11	-0.680	0.4	2684	7	1
12	-0.522	0.394	2687	7	1
13	-0.369	0.389	2689	7	1
14	-0.219	0.386	2692	7	1
15	-0.071	0.384	2695	7	1
16	0.075	0.383	2697	7	1
17	0.222	0.383	2700	7	2
18	0.370	0.385	2703	7	2
19	0.519	0.388	2705	7	2
20	0.671	0.392	2708	7	2
21	0.827	0.398	2711	7	2
22	0.989	0.406	2714	7	3
23	1.157	0.415	2717	7	3
24	1.334	0.427	2720	8	3
25	1.523	0.442	2723	8	3
26	1.726	0.46	2727	8	3
27	1.948	0.484	2731	9	3
28	2.197	0.514	2736	9	3
29	2.481	0.554	2741	10	3
30	2.817	0.607	2747	11	3
31	3.230	0.682	2754	12	3
32	3.773	0.802	2764	14	3
33	4.611	1.069	2779	19	3
34	5.912	1.862	2800	33	3

Table F-6. FI ELA Grade 8 Raw to Scale Score Conversion Table

TotalRawScore	Theta	SETheta	SS	SESS	PL
0	-5.035	1.853	2700	33	1
1	-3.764	1.045	2730	19	1
2	-2.984	0.764	2744	14	1
3	-2.499	0.64	2752	11	1
4	-2.137	0.568	2759	10	1
5	-1.843	0.519	2764	9	1
6	-1.592	0.484	2768	9	1
7	-1.371	0.457	2772	8	1
8	-1.171	0.437	2776	8	1
9	-0.988	0.421	2779	7	1
10	-0.816	0.408	2782	7	1
11	-0.653	0.398	2785	7	1
12	-0.498	0.39	2788	7	1
13	-0.348	0.384	2790	7	1
14	-0.202	0.38	2793	7	1
15	-0.059	0.377	2795	7	1
16	0.083	0.376	2798	7	1
17	0.224	0.375	2800	7	1
18	0.365	0.376	2803	7	1
19	0.507	0.379	2805	7	1
20	0.652	0.383	2808	7	2
21	0.801	0.388	2811	7	2
22	0.954	0.395	2813	7	2
23	1.114	0.405	2816	7	2
24	1.282	0.416	2819	7	2
25	1.461	0.431	2822	8	3
26	1.655	0.45	2826	8	3
27	1.868	0.474	2830	8	3
28	2.108	0.506	2834	9	3
29	2.384	0.548	2839	10	3
30	2.715	0.605	2845	11	3
31	3.130	0.688	2852	12	3
32	3.692	0.821	2862	15	3
33	4.580	1.103	2878	20	3
34	5.945	1.891	2900	33	3

Table F-7. FI ELA Grade 11 Raw to Scale Score Conversion Table

TotalRawScore	Theta	SETheta	SS	SESS	PL
0	-5.094	1.836	3000	53	1
1	-3.862	1.02	3033	29	1
2	-3.128	0.737	3054	21	1
3	-2.679	0.615	3067	18	1
4	-2.344	0.546	3077	16	1
5	-2.071	0.501	3084	14	1
6	-1.837	0.469	3091	14	1
7	-1.628	0.446	3097	13	1
8	-1.437	0.428	3103	12	1
9	-1.260	0.414	3108	12	1
10	-1.092	0.404	3113	12	1
11	-0.933	0.395	3117	11	1
12	-0.780	0.388	3122	11	1
13	-0.631	0.383	3126	11	1
14	-0.485	0.38	3130	11	1
15	-0.342	0.377	3134	11	1
16	-0.200	0.376	3138	11	1
17	-0.059	0.376	3143	11	1
18	0.083	0.377	3147	11	1
19	0.226	0.38	3151	11	2
20	0.371	0.383	3155	11	2
21	0.520	0.388	3159	11	2
22	0.673	0.395	3164	11	2
23	0.833	0.404	3168	12	2
24	1.000	0.415	3173	12	2
25	1.178	0.428	3178	12	3
26	1.368	0.445	3184	13	3
27	1.576	0.467	3190	13	3
28	1.807	0.495	3196	14	3
29	2.070	0.532	3204	15	3
30	2.378	0.582	3213	17	3
31	2.757	0.655	3224	19	3
32	3.262	0.777	3238	22	3
33	4.063	1.054	3261	30	3
34	5.347	1.857	3298	54	3

Table F-8. FI Math Grade 3 Raw to Scale Score Conversion Table

TotalRawScore	Theta	SETheta	SS	SESS	PL
0	-4.827	1.845	2200	40	1
1	-3.573	1.036	2221	23	1
2	-2.806	0.758	2238	17	1
3	-2.326	0.639	2248	14	1
4	-1.962	0.572	2256	12	1
5	-1.660	0.529	2263	12	1
6	-1.397	0.499	2269	11	1
7	-1.159	0.478	2274	10	1
8	-0.939	0.462	2279	10	1
9	-0.730	0.452	2283	10	1
10	-0.530	0.444	2287	10	1
11	-0.334	0.440	2292	10	1
12	-0.141	0.439	2296	10	1
13	0.052	0.440	2300	10	1
14	0.247	0.444	2304	10	1
15	0.447	0.451	2309	10	1
16	0.656	0.462	2313	10	2
17	0.876	0.477	2318	10	2
18	1.113	0.498	2323	11	2
19	1.375	0.528	2329	12	2
20	1.675	0.571	2336	12	2
21	2.038	0.638	2343	14	2
22	2.516	0.756	2354	16	3
23	3.279	1.034	2371	23	3
24	4.531	1.844	2398	40	3

Table F-9. FI Math Grade 4 Raw to Scale Score Conversion Table

TotalRawScore	Theta	SETheta	SS	SESS	PL
0	-4.435	1.842	2300	39	1
1	-3.190	1.029	2333	22	1
2	-2.436	0.749	2349	16	1
3	-1.969	0.629	2359	13	1
4	-1.618	0.561	2366	12	1
5	-1.329	0.517	2372	11	1
6	-1.079	0.486	2378	10	1
7	-0.854	0.464	2382	10	1
8	-0.646	0.448	2387	10	1
9	-0.451	0.437	2391	9	1
10	-0.263	0.429	2395	9	1
11	-0.081	0.425	2399	9	1
12	0.099	0.423	2403	9	1
13	0.278	0.425	2407	9	1
14	0.460	0.429	2410	9	2
15	0.647	0.436	2414	9	2
16	0.842	0.447	2419	10	2
17	1.048	0.463	2423	10	2
18	1.272	0.485	2428	10	2
19	1.521	0.515	2433	11	3
20	1.808	0.559	2439	12	3
21	2.158	0.628	2447	13	3
22	2.623	0.747	2457	16	3
23	3.373	1.028	2473	22	3
24	4.617	1.841	2499	39	3

Table F-10. FI Math Grade 5 Raw to Scale Score Conversion Table

TotalRawScore	Theta	SETheta	SS	SESS	PL
0	-4.424	1.841	2400	40	1
1	-3.180	1.028	2431	22	1
2	-2.428	0.748	2447	16	1
3	-1.962	0.628	2457	14	1
4	-1.612	0.560	2465	12	1
5	-1.324	0.516	2471	11	1
6	-1.075	0.485	2476	10	1
7	-0.850	0.463	2481	10	1
8	-0.643	0.447	2486	10	1
9	-0.448	0.436	2490	9	1
10	-0.262	0.429	2494	9	1
11	-0.080	0.424	2498	9	1
12	0.099	0.423	2502	9	1
13	0.278	0.424	2505	9	1
14	0.460	0.428	2509	9	1
15	0.646	0.435	2513	9	1
16	0.840	0.446	2518	10	2
17	1.045	0.462	2522	10	2
18	1.268	0.483	2527	10	2
19	1.516	0.514	2532	11	2
20	1.802	0.558	2538	12	2
21	2.150	0.626	2546	13	3
22	2.613	0.746	2556	16	3
23	3.362	1.027	2572	22	3
24	4.604	1.840	2599	40	3

Table F-11. FI Math Grade 6 Raw to Scale Score Conversion Table

TotalRawScore	Theta	SETheta	SS	SESS	PL
0	-4.610	1.847	2500	40	1
1	-3.353	1.037	2527	22	1
2	-2.584	0.758	2544	16	1
3	-2.104	0.639	2554	14	1
4	-1.740	0.571	2562	12	1
5	-1.441	0.526	2569	11	1
6	-1.181	0.495	2574	11	1
7	-0.947	0.473	2579	10	1
8	-0.731	0.457	2584	10	1
9	-0.528	0.445	2588	10	1
10	-0.334	0.437	2593	9	1
11	-0.145	0.432	2597	9	1
12	0.040	0.430	2601	9	1
13	0.225	0.431	2605	9	1
14	0.413	0.435	2609	9	1
15	0.604	0.441	2613	10	2
16	0.803	0.452	2617	10	2
17	1.013	0.467	2622	10	2
18	1.241	0.488	2627	11	2
19	1.492	0.518	2632	11	3
20	1.782	0.561	2638	12	3
21	2.133	0.629	2646	14	3
22	2.600	0.748	2656	16	3
23	3.352	1.028	2672	22	3
24	4.595	1.841	2699	40	3

Table F-12. FI Math Grade 7 Raw to Scale Score Conversion Table

TotalRawScore	Theta	SETheta	SS	SESS	PL
0	-4.958	1.845	2600	41	1
1	-3.705	1.035	2618	23	1
2	-2.941	0.756	2635	17	1
3	-2.462	0.638	2645	14	1
4	-2.099	0.571	2653	13	1
5	-1.798	0.528	2660	12	1
6	-1.536	0.498	2666	11	1
7	-1.299	0.477	2671	10	1
8	-1.079	0.462	2676	10	1
9	-0.871	0.451	2680	10	1
10	-0.672	0.444	2685	10	1
11	-0.477	0.439	2689	10	1
12	-0.285	0.438	2693	10	1
13	-0.093	0.439	2697	10	1
14	0.101	0.443	2702	10	1
15	0.300	0.450	2706	10	2
16	0.507	0.460	2711	10	2
17	0.725	0.475	2715	10	2
18	0.961	0.496	2721	11	2
19	1.221	0.526	2726	12	2
20	1.519	0.569	2733	13	3
21	1.878	0.636	2741	14	3
22	2.353	0.754	2751	17	3
23	3.114	1.032	2768	23	3
24	4.363	1.843	2795	41	3

Table F-13. FI Math Grade 8 Raw to Scale Score Conversion Table

TotalRawScore	Theta	SETheta	SS	SESS	PL
0	-4.589	1.843	2700	38	1
1	-3.339	1.032	2733	21	1
2	-2.581	0.752	2749	16	1
3	-2.109	0.633	2758	13	1
4	-1.753	0.565	2766	12	1
5	-1.460	0.521	2772	11	1
6	-1.205	0.490	2777	10	1
7	-0.976	0.468	2782	10	1
8	-0.766	0.452	2786	9	1
9	-0.567	0.440	2790	9	1
10	-0.377	0.432	2794	9	1
11	-0.192	0.428	2798	9	1
12	-0.011	0.426	2802	9	1
13	0.171	0.427	2805	9	1
14	0.354	0.431	2809	9	1
15	0.543	0.438	2813	9	2
16	0.738	0.448	2817	9	2
17	0.946	0.464	2821	10	2
18	1.170	0.485	2826	10	2
19	1.419	0.515	2831	11	3
20	1.707	0.559	2837	12	3
21	2.056	0.627	2844	13	3
22	2.520	0.747	2854	15	3
23	3.270	1.028	2870	21	3
24	4.513	1.841	2895	38	3

Table F-14. FI Math Grade 11 Raw to Scale Score Conversion Table

TotalRawScore	Theta	SETheta	SS	SESS	PL
0	-4.592	1.845	3000	61	1
1	-3.340	1.034	3040	34	1
2	-2.578	0.754	3065	25	1
3	-2.104	0.634	3080	21	1
4	-1.747	0.566	3092	19	1
5	-1.453	0.521	3102	17	1
6	-1.198	0.490	3110	16	1
7	-0.969	0.468	3118	15	1
8	-0.758	0.452	3125	15	1
9	-0.559	0.440	3131	15	1
10	-0.369	0.432	3137	14	1
11	-0.185	0.428	3144	14	1
12	-0.003	0.426	3150	14	1
13	0.178	0.427	3156	14	2
14	0.362	0.431	3162	14	2
15	0.550	0.438	3168	14	2
16	0.746	0.448	3174	15	2
17	0.954	0.464	3181	15	2
18	1.178	0.485	3189	16	3
19	1.428	0.515	3197	17	3
20	1.715	0.559	3206	18	3
21	2.065	0.628	3218	21	3
22	2.530	0.747	3233	25	3
23	3.280	1.028	3258	34	3
24	4.523	1.841	3299	61	3

Table F-15. FI Science Grade 4 Raw to Scale Score Conversion Table

TotalRawScore	Theta	SETheta	SS	SESS	PL
0	-5.135	1.847	2300	32	1
1	-3.878	1.036	2323	18	1
2	-3.115	0.753	2336	13	1
3	-2.644	0.629	2344	11	1
4	-2.296	0.556	2350	10	1
5	-2.015	0.507	2355	9	1
6	-1.776	0.472	2360	8	1
7	-1.566	0.445	2363	8	1
8	-1.378	0.424	2367	7	1
9	-1.205	0.407	2370	7	1
10	-1.045	0.394	2372	7	1
11	-0.894	0.383	2375	7	1
12	-0.751	0.375	2378	7	1
13	-0.613	0.368	2380	6	1
14	-0.480	0.362	2382	6	1
15	-0.350	0.358	2385	6	1
16	-0.223	0.355	2387	6	1
17	-0.097	0.354	2389	6	1
18	0.028	0.353	2391	6	1
19	0.153	0.354	2393	6	1
20	0.279	0.356	2396	6	1
21	0.406	0.359	2398	6	1
22	0.536	0.363	2400	6	2
23	0.670	0.369	2402	6	2
24	0.809	0.376	2405	7	2
25	0.954	0.386	2407	7	2
26	1.107	0.398	2410	7	2
27	1.271	0.413	2413	7	3
28	1.449	0.432	2416	8	3
29	1.647	0.458	2420	8	3
30	1.872	0.492	2424	9	3
31	2.136	0.539	2428	9	3
32	2.463	0.610	2434	11	3
33	2.908	0.734	2442	13	3
34	3.639	1.019	2455	18	3
35	4.869	1.836	2476	32	3

Table F-16. FI Science Grade 7 Raw to Scale Score Conversion Table

TotalRawScore	Theta	SETheta	SS	SESS	PL
0	-5.341	1.838	2600	31	1
1	-4.106	1.022	2622	17	1
2	-3.369	0.738	2634	12	1
3	-2.919	0.615	2642	10	1
4	-2.586	0.544	2647	9	1
5	-2.317	0.496	2652	8	1
6	-2.088	0.462	2656	8	1
7	-1.886	0.437	2659	7	1
8	-1.705	0.417	2662	7	1
9	-1.538	0.401	2665	7	1
10	-1.382	0.388	2668	7	1
11	-1.236	0.378	2670	6	1
12	-1.097	0.369	2672	6	1
13	-0.963	0.362	2675	6	1
14	-0.834	0.356	2677	6	1
15	-0.709	0.352	2679	6	1
16	-0.586	0.348	2681	6	1
17	-0.466	0.345	2683	6	1
18	-0.347	0.343	2685	6	1
19	-0.230	0.342	2687	6	1
20	-0.113	0.342	2689	6	1
21	0.004	0.342	2691	6	1
22	0.121	0.343	2693	6	1
23	0.240	0.345	2695	6	1
24	0.360	0.348	2697	6	1
25	0.482	0.351	2699	6	1
26	0.606	0.355	2701	6	2
27	0.735	0.361	2703	6	2
28	0.867	0.368	2706	6	2
29	1.006	0.376	2708	6	2
30	1.151	0.387	2710	7	2
31	1.305	0.399	2713	7	2
32	1.471	0.415	2716	7	3
33	1.651	0.435	2719	7	3
34	1.851	0.460	2722	8	3
35	2.077	0.494	2726	8	3
36	2.344	0.541	2731	9	3
37	2.675	0.613	2736	10	3
38	3.122	0.736	2744	12	3
39	3.856	1.020	2756	17	3

TotalRawScore	Theta	SETheta	SS	SESS	PL
40	5.088	1.837	2777	31	3

Table F-17. FI Science Grade 11 Raw to Scale Score Conversion Table

TotalRawScore	Theta	SETheta	SS	SESS	PL
0	-5.387	1.835	3000	33	1
1	-4.158	1.017	3019	18	1
2	-3.430	0.731	3032	13	1
3	-2.990	0.607	3040	11	1
4	-2.667	0.534	3046	9	1
5	-2.408	0.486	3050	9	1
6	-2.190	0.450	3054	8	1
7	-1.999	0.424	3058	8	1
8	-1.829	0.403	3061	7	1
9	-1.674	0.386	3063	7	1
10	-1.530	0.372	3066	7	1
11	-1.396	0.360	3068	6	1
12	-1.270	0.351	3071	6	1
13	-1.150	0.343	3073	6	1
14	-1.035	0.336	3075	6	1
15	-0.924	0.330	3077	6	1
16	-0.816	0.326	3079	6	1
17	-0.712	0.322	3080	6	1
18	-0.609	0.319	3082	6	1
19	-0.509	0.316	3084	6	1
20	-0.409	0.314	3086	6	1
21	-0.311	0.313	3088	6	1
22	-0.213	0.312	3089	6	1
23	-0.116	0.312	3091	6	1
24	-0.018	0.313	3093	6	1
25	0.080	0.314	3095	6	1
26	0.179	0.316	3096	6	1
27	0.280	0.318	3098	6	1
28	0.382	0.321	3100	6	2
29	0.486	0.325	3102	6	2
30	0.593	0.329	3104	6	2
31	0.703	0.335	3106	6	2
32	0.817	0.341	3108	6	2
33	0.936	0.349	3110	6	2
34	1.062	0.359	3112	6	2
35	1.194	0.370	3114	7	2
36	1.336	0.384	3117	7	2
37	1.490	0.400	3120	7	3
38	1.658	0.421	3123	7	3
39	1.847	0.448	3126	8	3

TotalRawScore	Theta	SETheta	SS	SESS	PL
40	2.063	0.483	3130	9	3
41	2.319	0.532	3134	9	3
42	2.639	0.605	3140	11	3
43	3.076	0.729	3148	13	3
44	3.801	1.015	3161	18	3
45	5.026	1.834	3182	33	3

Table F-18. FI Social Studies Grade 5 Raw to Scale Score Conversion Table

TotalRawScore	Theta	SETheta	SS	SESS	PL
0	-4.690	1.836	2400	33	1
1	-3.459	1.019	2425	18	1
2	-2.727	0.735	2438	13	1
3	-2.281	0.612	2446	11	1
4	-1.953	0.540	2452	10	1
5	-1.687	0.493	2457	9	1
6	-1.461	0.460	2461	8	1
7	-1.261	0.435	2464	8	1
8	-1.081	0.416	2468	7	1
9	-0.914	0.401	2471	7	1
10	-0.759	0.389	2473	7	1
11	-0.611	0.380	2476	7	1
12	-0.469	0.373	2478	7	1
13	-0.331	0.368	2481	7	1
14	-0.197	0.365	2483	6	1
15	-0.065	0.363	2486	6	1
16	0.066	0.362	2488	6	1
17	0.198	0.363	2490	6	1
18	0.330	0.365	2493	6	1
19	0.464	0.368	2495	7	1
20	0.602	0.374	2497	7	1
21	0.744	0.380	2500	7	2
22	0.892	0.389	2503	7	2
23	1.048	0.401	2505	7	2
24	1.215	0.416	2508	7	2
25	1.395	0.435	2512	8	3
26	1.595	0.460	2515	8	3
27	1.821	0.493	2519	9	3
28	2.087	0.540	2524	10	3
29	2.417	0.612	2530	11	3
30	2.862	0.735	2538	13	3
31	3.595	1.019	2551	18	3
32	4.826	1.836	2573	33	3

Table F-19. FI Social Studies Grade 8 Raw to Scale Score Conversion Table

TotalRawScore	Theta	SETheta	SS	SESS
0	-4.875	1.836	2700	35
1	-3.646	1.018	2723	19
2	-2.916	0.733	2737	14
3	-2.473	0.610	2746	12
4	-2.146	0.538	2752	10
5	-1.882	0.491	2757	9
6	-1.659	0.457	2761	9
7	-1.462	0.432	2765	8
8	-1.284	0.412	2769	8
9	-1.120	0.397	2772	8
10	-0.967	0.386	2775	7
11	-0.822	0.376	2777	7
12	-0.683	0.369	2780	7
13	-0.549	0.363	2783	7
14	-0.419	0.360	2785	7
15	-0.290	0.357	2788	7
16	-0.163	0.356	2790	7
17	-0.037	0.356	2792	7
18	0.090	0.357	2795	7
19	0.218	0.360	2797	7
20	0.349	0.364	2800	7
21	0.483	0.369	2802	7
22	0.622	0.377	2805	7
23	0.768	0.386	2808	7
24	0.921	0.398	2811	8
25	1.085	0.413	2814	8
26	1.264	0.432	2817	8
27	1.461	0.458	2821	9
28	1.685	0.491	2825	9
29	1.950	0.539	2830	10
30	2.277	0.610	2837	12
31	2.721	0.734	2845	14
32	3.452	1.018	2859	19
33	4.682	1.836	2883	35

Table F-20. FI Social Studies Grade 11 Raw to Scale Score Conversion Table

TotalRawScore	Theta	SETheta	SS	SESS
0	-4.927	1.834	3000	33
1	-3.701	1.016	3024	18
2	-2.976	0.729	3037	13
3	-2.538	0.605	3045	11
4	-2.218	0.532	3051	10
5	-1.961	0.484	3055	9
6	-1.744	0.449	3059	8
7	-1.555	0.422	3063	8
8	-1.386	0.402	3066	7
9	-1.231	0.385	3069	7
10	-1.088	0.372	3071	7
11	-0.954	0.361	3074	7
12	-0.827	0.352	3076	6
13	-0.706	0.344	3078	6
14	-0.590	0.338	3080	6
15	-0.478	0.333	3082	6
16	-0.368	0.329	3084	6
17	-0.261	0.326	3086	6
18	-0.156	0.323	3088	6
19	-0.052	0.322	3090	6
20	0.051	0.321	3092	6
21	0.154	0.321	3094	6
22	0.257	0.322	3096	6
23	0.361	0.323	3097	6
24	0.466	0.325	3099	6
25	0.573	0.328	3101	6
26	0.682	0.332	3103	6
27	0.794	0.337	3105	6
28	0.910	0.343	3107	6
29	1.030	0.351	3110	6
30	1.156	0.360	3112	7
31	1.289	0.371	3114	7
32	1.431	0.384	3117	7
33	1.585	0.400	3120	7
34	1.753	0.421	3123	8
35	1.941	0.447	3126	8
36	2.157	0.482	3130	9
37	2.412	0.531	3135	10
38	2.732	0.604	3140	11
39	3.168	0.728	3148	13

TotalRawScore	Theta	SETheta	SS	SESS
40	3.891	1.015	3161	18
41	5.116	1.834	3184	33

Appendix G: Verification of Psychometric Work for Spring 2019 MI-Access Administration

UCLA | CRESST NATIONAL CENTER FOR RESEARCH ON EVALUATION, STANDARDS, AND STUDENT TESTING

PSYCHOMETRIC VERIFICATION OF MICHIGAN'S ALTERNATE ASSESSMENT PROGRAM (MI-ACCESS) FOR SPRING 2019 ADMINISTRATION

CRESST Psychometrics Team

National Center for Research on Evaluation,
Standards, and Student Testing (CRESST)
University of California, Los Angeles

OCTOBER 2019

Table of Contents

Executive Summary iii

1. Introduction 1

2. Methodology 2

2.1. FI ELA and Math 2

2.2. FI Science and Social Studies 4

2.3. P/SI ELA, Math, and Science 5

3. Verification Results and Conclusion 5

4. Discussion 6

References 7

Appendix A: Mean-Mean Equating Procedure for MI-Access Functional Independence ELA and Math Tests 8

Appendix B: 2019 MI-ACCESS Phase I Verification Activities 9

Appendix C: 2019 MI-ACCESS Phase II Verification Activities 10

Executive Summary

The National Center for Research on Evaluation, Standards, and Student Testing (CRESST) at the University of California, Los Angeles (UCLA) conducted an independent psychometric verification of two testing programs under contract with the Michigan Department of Education (MDE) from May to September of 2019. The two testing programs were the Michigan Student Test of Educational Progress (M-STEP) and Michigan's Alternate Assessment Program (MI-ACCESS). For each testing program, the verification work included two phases: (a) creation of raw-to-scale-score (RSS) tables and raw-to-performance-level (RPL) conversion tables and (b) item statistics computation for the item bank system (IBS).

This report documents the psychometric verification of the MI-ACCESS testing program which includes three testing populations: Functional Independence (FI), Supported Independence (SI), and Participation (P). Based upon the agreement with MDE, three software programs—WINSTEPS 3.92.1, DIFAS 5.0, and R 3.4.1—were used for RSS table creation, differential item functioning (DIF) analysis, and classical test theory (CTT) item statistics analysis, respectively.

The verification work included processing data (for FI science and social studies, P/SI English language arts [ELA], math, and science) and simultaneously conducting psychometric analyses by applying the same approach and methodology as MDE. It also involved evaluation of statistical methods and comparison of results between CRESST and MDE (e.g., item parameters, RSS tables, and item statistics) at each subtask stage and at the completion of the task.

During the process, very detailed and careful alignment of the analytic approaches was occasionally required to produce the exact same numerical results. Nonetheless, through active discussion and exchange of feedback regarding soundness of the procedures and results, CRESST and MDE found and corrected errors, and adequately resolved discrepancies.

As a result, CRESST and MDE produced the exact same numerical results, except for few small discrepancies (in the third or fourth decimal place) in the FI ELA and math final item parameters. These discrepancies were likely due to different rounding algorithms implemented in different software programs or software versions. The verification results showed that the psychometric analyses of the MI-ACCESS testing program were performed with high precision and were successfully verified.

Further detailed descriptions are provided in the body of the report, including methodology, verification results, and discussion. Following the discussion, tables in Appendix B and C outline the detailed account of the verification procedure and key findings for each of the content areas.

1. Introduction

This report documents the verification activities of Michigan's Alternate Assessment Program (MI-ACCESS) administered in 2019, which includes three testing populations: Functional Independence (FI), Supported Independence (SI), and Participation (P). The verification work consisted of two phases: (a) creation of raw-to-scale-score (RSS) conversion tables and raw-to-performance-level (RPL) conversion tables (b) item statistics computation for the item bank system (IBS). Based upon the agreement with the Michigan Department of Education (MDE), WINSTEPS 3.92.1 (Linacre, 2016), DIFAS 5.0 (Penfield, 2013), and R 3.4.1 (R Core team, 2017) were used for RSS table creation, differential item functioning (DIF) analysis, and classical test theory (CTT) item statistics analysis, respectively. Specifically, WINSTEPS 1-parameter logistic (1PL)/partial credit model (PCM) calibration was used for RSS table creation. The verification work involved the following tasks:

- FI English language arts (ELA) and math
  - Phase I work: (a) initial operational data calibration, (b) anchor evaluation, (c) mean-mean equating, and (d) RSS conversion table creation
  - Phase II work: (a) final data calibration, (b) classical statistics analysis, and (c) DIF analysis

- FI science and social studies
  - Phase I work: (a) anchor evaluation via fixed-parameter WINSTEPS calibration, (b) equating online operational items, (c) online field test item calibration, (d) paper-pencil item calibration, and (e) RSS conversion table creation
  - Phase II work: (a) final data calibration, (b) classical statistics analysis, and (c) DIF analysis

- P/SI ELA, math, and science
  - Phase I work: RPL conversion table creation
  - Phase II work: classical statistics analysis

For careful alignment of analytic approaches, supporting documents that described the analysis procedures, variables, and evaluation/flagging criteria were provided to CRESST by MDE. The following list is by no means exhaustive, but it details documents that are frequently referred to throughout this report:

1) MIAccess_CutScores_FIScalingConstants.xls
2) Spr2019 MIAccess_MinMax_ScoreRanges_PLtables.xls
3) Decision on FI Min and Max Scalescore Truncation.doc

4) MI-Access FI Variable List for Item Bank Stats Analysis.xls
5) MI-Access P_SI Variable List for Item Bank Stats Analysis.xls
6) Item Flagging Criterion List for MI-Access Data Review (7-24-19).doc

2. Methodology

2.1.FI ELA and Math

For ELA and math, CRESST received item-level data for online (OL) and paper-pencil (PP) forms. Both content areas were assessed at Grades 3 to 8 as well as Grade 11. For each of these grade and content combinations, the test consists of operational (OP) items and field-test (FT) items. We note that Accessing Print (AP) and Expressing Idea (EI) were combined to be reported as ELA, and that EI was only administered in PP forms.

- Phase I

  The process of obtaining OP item parameters involved the following subtasks: (a) initial operational data calibration, (b) anchor evaluation, and (c) mean-mean equating. For each grade and content combination, a free WINSTEPS run was first conducted to obtain initial item parameters for all OP items. The mean-mean equating approach (for more details, see Appendix A) was then applied to evaluate the suitability of equating items and to compute equating constants. Finally, the equating constants were added to the initial OP item parameters to obtain the equated item parameters.

  Given the initial RSS table generated from the free WINSTEPS run, the equating constants obtained in the previous step, scaling constants, and cut scores (provided in MIAccess_CutScores_FIScalingConstants.xls) were used to create the final RSS table. The equating constants were applied to the corresponding thetas. A linear transformation was then conducted to transform these thetas into scale scores. When necessary, the scale scores were truncated to the prespecified minimum and maximum scale scores (Decision on FI Min and Max Scalescore Truncation.doc).

- Phase II
  - Final online (OL) data calibration

    A one-step fixed-parameter WINSTEPS calibration approach was used for the OL data. All the OP item parameters were first fixed to the equated

parameters obtained from Phase I to calibrate the FT items. Stability of the fixed OP items was checked based on the displacement value. Items with absolute displacement values equal to or larger than 0.5 were flagged for exclusion from the anchor parameter list with the constraint that no more than 30% of the anchors should be dropped. The process of removing the flagged item with the largest displacement value was iteratively done until no item was flagged for exclusion.

- Final paper-pencil (PP) data calibration

  For the PP data, a two-step procedure was conducted. In Step 1, the equating item parameters obtained from Phase 1 were fixed to calibrate the paper OP items. The stability of the fixed items was checked using the same iterative procedure applied for the OL data. In Step 2, the paper OP item parameters were fixed to calibrate the paper FT items. The stability of the fixed OP items was checked in the same way as described in Step 1.

  With the two-step method, however, the EI constructed response (CR) OP item for each grade was not on the ELA anchor list, leading to different parameter estimates for the EI CR OP item between the modes. After discussion with MDE, the EI CR OP item was added to the anchor list in Step 1 and the final item parameters for the two EI CR FT items were borrowed from the online calibration.

- Classical statistics analysis

  Although FI ELA consisted of AP and EI items, classical statistics analysis and DIF analysis were separately conducted for each of AP OL, AP PP, and EI PP data sets. As for math, these analyses were conducted per mode (OL and PP).

  For each content/grade/mode combination, classical item statistics were calculated based on responses of the overall population and subpopulations. These subpopulations included groups classified by gender (*male* and *female*), ethnicity (*White, Black* or *African American*), economic disadvantage status (*ED* and *Non-ED*), and accommodation status (*Acc* and *Non-Acc*). It should be noted that classical item statistics were not computed for subgroups with $n < 30$.

  For each item, a list of classical item statistics (MI-Access FI Variable List for Item Bank Stats Analysis.xls) were calculated: the number of students that responded to the item, the proportion of each option among all responses (for multiple-choice items only), the proportions of scores 0-4 (for EI items), the mean score, the adjusted mean score, the score standard deviation, the raw and

adjusted correlations between the item score and the total score, and finally the correlation between each option and the total score.

When applicable, item flag variables for item difficulty (DIFFICFL), item total correlation (ITOTFL), mean-square infit (MSQINFL) and outfit (MSQOUTFL) were also obtained based on the corresponding item flagging criteria (Item Flagging Criterion List for MI-Access Data Review.doc).

- Differential Item Functioning (DIF) analysis

  DIF analyses were conducted using DIFAS 5.0 for the following comparisons: *male* vs. *female*, *White* vs. *Black* or *African American*, *ED* vs. *Non-ED*, and *Acc* vs. *Non-Acc*. DIF statistics were not computed for subgroups with $n < 30$. The sum scores of OP items were used as a stratifying variable and a stratum size was set to 1 in the first run and was incremented by 1 when DIFAS encountered a convergence issue.

  For each comparison, Mantel-Haenszel chi-square-related statistics (MCHI and MHD) were reported (for details, see MI-Access FI Variable List for Item Bank Stats Analysis.xls). Additionally, two flagging variables, DIF and FG, were computed according to the criteria (Item Flagging Criterion List for MI-Access Data Review.doc).

2.2.FI Science and Social Studies

For science and social studies, the OL and PP data were provided. Science was assessed at Grades 4, 7, and 11, and social studies was assessed at Grades 5, 8, and 11.

- Phase I

  (a) For each grade and content combination, a fixed parameter WINSTEPS run was conducted by fixing all equating item parameters to their corresponding item bank values. Then, stability of the equating items was evaluated based on the displacement value. Items with absolute displacement values larger than 0.5 were removed from the equating set.

  (b) A free WINSTEPS run was conducted using OL OP items, without fixing any item parameters. Steps 1–5 from the mean-mean equating approach (see Appendix A) were then applied to compute equating constants. Finally, the equating constants were added to the item parameters from the free run to obtain equated item parameters.

  (c) For OL FT item calibration, an anchored WINSTEPS run was done by fixing the OP item parameters to the adjusted parameters obtained in (b).

(d) An anchored WINSTEPS run was conducted to calibrate the PP data, by fixing OL item parameters to the parameters obtained in Steps (b) and (c).
(e) For each form shown in the testmap, a RSS table was generated via anchored WINSTEPS calibration based on the OP items, by fixing their parameters obtained from the previous steps. Random responses were generated for test forms with few records (e.g., braille and emergency forms) to allow WINSTEPS to generate the RSS tables.

- Phase II
  - Classical statistics analysis and DIF analysis

    Classical statistics analysis and DIF analysis were conducted following the procedures described for FI math. When applicable, the same classical item statistics and DIF statistics were computed for science and social studies as math.

2.3. P/SI ELA, Math, and Science

For P/SI, all three content areas were assessed at Grades 3 to 8, and Grade 11.

- Phase I
  - For P/SI, item response theory (IRT)-based scale scores and RSS tables were not computed. Instead, RPL conversion tables were created based on the raw score cuts (MIAccess_CutScores_FIScalingConstants.xls).

- Phase II
  - Classical statistics analysis

    Classical statistics analyses were conducted for the overall population and subgroups classified by gender (*male* and *female*), ethnicity (*White*, *Black* or *African American*), and economic disadvantage status (*ED* and *Non-ED).* The subgroups with $n < 30$ were excluded from the analysis. For each item, the corresponding classical item statistics (MI-Access P_SI Variable List for Item Bank Stats Analysis.xls) were computed. When applicable, item flag variables, DIFFICFL and ITOTFL were reported (Item Flagging Criterion List for MI-Access Data Review.doc) as well.

3. Verification Results and Conclusion

During the verification process, very detailed and careful alignment of the analytic approaches was occasionally required to produce the matching numerical results. CRESST and

MDE, however, adequately resolved discrepancies by detecting and fixing errors in software program files and statistical results, and maintaining the same level of precision (i.e., decimal places) through active discussion and exchange of feedback. In the case of slight discrepancies due to rounding errors, CRESST and MDE discussed an acceptable range and agreed that small differences in the third or fourth decimal place were acceptable.

The verification of the MI-ACCESS testing program indicated that Phase 1 psychometric analyses conducted by CRESST yielded the exact same numerical results as MDE, including the operational item parameters, the RSS tables, and the RPL tables. Similarly, the results from Phase 2 analyses exactly matched, except for a few small discrepancies in the FI ELA and math final item calibration results. Specifically, one FI math OL item showed a 0.0001 difference in the item infit measure and few FI ELA OL items showed a 0.001 difference in the item parameters. These discrepancies were likely due to different rounding algorithms implemented in different software programs (e.g., R, SAS, and Excel), and such differences (in the third or fourth decimal place) were well within the acceptable ranges.

In summary, the verification results indicate that the MI-ACCESS psychometric analyses were performed with high precision and were successfully verified.

4. Discussion

Among the statistical programs, DIFAS 5.0 encountered technical/convergence issues while conducting DIF analyses for the FI AP subpopulations. Specifically, depending on the stratum size, the program occasionally did not complete the analysis and/or did not provide DIF results for a few items. Another concern about using DIFAS is that for each single comparison, the number of strata had to be manually specified. Considering the recent manual for DIFAS was published in 2013, compatibility issues with new operating systems may arise in the near future. Thus, future efforts could consider using alternative software; the CRESST team has already begun exploring possibilities. For example, the psychometric team has started developing R code that provides the same DIF statistics and flagging variables as DIFAS. A pilot study showed that despite the difficulty in specifying the exact same settings as DIFAS, the current version of code still produced similar results to DIFAS. Upon request, this code can be made available and can be further updated to minimize discrepancies in numerical results. Other options include using up-to-date software such as flexMIRT (Cai, 2017).

References

Cai, L. (2017). flexMIRT® version 3.51: Flexible multilevel multidimensional item analysis and test scoring [Computer software]. Chapel Hill, NC: Vector Psychometric Group.

Linacre, J. M. (2016). Winsteps® (Version 3.92.1) [Computer Software]. Beaverton, Oregon: Winsteps.com.

Penfield, R. D. (2013). DIFAS 5.0. User's manual. Retrieved from https://soe.uncg.edu/wp-content/uploads/2015/12/DIFASManual_V5.pdf

R Core Team. (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.