

Missouri Assessment Program

Design for Standard Setting

Prepared for the Department of Elementary and Secondary Education (DESE)
by Ricardo Mercado and Joanna Tomkowicz, Data Recognition Corporation (DRC)

February 16, 2016

Version Note

This is version 3 of the *Design for Standard Setting*. This version reflects comments from recent conversations between DESE and DRC, including confirmation that the ALDs will not undergo substantive revision before the standard setting, details on the Technical Advisory Committee (TAC) recommendation on the configuration of the standard setting committee, and information on the use of writing tasks and the potential use of benchmarks at the standard setting. Previous versions of this document incorporated other suggestions from the TAC, details on the ways existing cut scores may be projected onto the new test scale as benchmarks, and information on the potential use of benchmarks at the standard setting.

Background

In summer 2016, a standard setting will be held to establish cut scores for the Missouri Assessment Program (MAP). The standard setting will focus on achievement standards for the MAP assessments of English language arts (ELA) and mathematics in Grades 3–8.

Achievement standards were established for MAP in 2005. At that standard setting, explicit linkages were made between the achievement standards and NAEP in compliance with newly-enacted state requirements (SB 1080). In the intervening years, the state joined the Smarter Balanced Assessment Consortium (SBAC), and the tests have been presented online starting with 2014-15 test administration. In that year, Missouri began using the SBAC common core standards (which are the same as Missouri Learning Standards), including the SBAC ALDs and cut scores.

DESE has indicated that it no longer intends to use SBAC items or test scales, and new tests and scales will be established in school year 2015-16 for the MAP program. In 2016, DESE seeks to establish achievement standards for the new MAP to reflect up-to-date expectations for student performance across the state of Missouri. During this standard setting, DESE will use the same set of content standards and SBAC's ALDs as a basis, seeking to develop cut scores on the MAP assessments that reflect these content-based expectations on the new tests.

The Bookmark Standard Setting Procedure (BSSP)¹ will be implemented to recommend cut scores for MAP. This method has a history of use on large-scale assessments across the nation. In Missouri, the Bookmark Procedure was used to establish the original achievement standards for English language arts and mathematics in 2005, and most recently in 2015 to validate achievement standards for science.

DESE has indicated that four achievement levels will be established for MAP English language arts and mathematics. These four achievement levels are designed to provide clear indicators of students' knowledge of the skills listed in the Show-Me Standards, and to indicate their ultimate progression toward college and career readiness. These achievement levels are *Below Basic*, *Basic*, *Proficient*, and *Advanced*.

Achievement Level Descriptors (ALDs)

ALDs summarize the knowledge, skills, and abilities expected of students in each achievement level. The existing ELA and Mathematics ALDs, used in 2014-15 MAP reporting, were developed by Smarter Balanced Assessment Consortium. ALDs are powerful tools for communicating the intended inferences associated with achievement standards, and for helping teachers understand the skills held by their students.

DESE has indicated that the content-based expectations for students in each achievement level, as described in the existing ALDs, still reflect the knowledge, skills, and abilities that are expected of Missouri students in each achievement level. As such, the existing ALDs will be used at the standard setting to convey these expectations to standard setting participants. DESE has indicated that it reserves the right to make minor adjustments to the ALDs in wording and style to the ALDs to promote clarity and consistency; however, these adjustments will not affect the overall level of content-based expectations for any given achievement level.

The ALDs will be presented to participants early in the standard setting process. This way, participants will have a detailed understanding of the types of knowledge and skills that DESE expects of students in each achievement level. This activity is described later in this document.

Configuration of the Standard Setting Committee

For each content area, six panels of approximately 10 participants each will be convened for a total of 12 panels and 120 participants. Within each content area, these six panels will each focus on a separate grade level. This configuration, recommended by the TAC, will allow Missouri educators to consider the achievement standards for each test carefully, while still allowing the entire committee to consider the articulation of the achievement standards across grades.

¹ Lewis, D. M., Mitzel, H. C., & Green, D. R. (1996). Standard setting: A Bookmark approach. Symposium presented at the Council of Chief State School Officers National Conference on Large-Scale Assessment: Phoenix, AZ.

Mitzel, H. C., Lewis, D. M., Patz, R. J., & Green, D. R. (2001). The bookmark procedure: Psychological perspectives. In G. J. Cizek (Ed.), *Setting performance standards: Concepts, methods, and perspectives* (pp. 249-281). Mahwah, NJ: Lawrence Erlbaum.

If space permits, the panels for each content area may all work in a single room to promote consistency in process and to facilitate across-grade discussions on the last day of the workshop. However, if space will not allow such an arrangement, two or three panels may work in a single room, overseen by a group leader from DRC. DRC group leaders will work in tandem with DESE and the rest of the workshop staff to promote consistency across the entire standard setting committee.

Each panel will be subdivided into two *tables* of approximately five participants each. Structured in this way, the two tables for each panel will replicate the standard setting process for each other, allowing DRC to estimate the amount of statistical variance in the cut score recommendations. The interaction between the two tables for each group, as well as the interaction between groups, is described later in this document.

DRC Education Program Management and DESE will collaborate to assign participants to tables so that each table is as representative and balanced as possible with regard to the relevant demographics (e.g., gender, geographic location). DRC Education Program Management and the DESE will also collaborate to designate one participant at each table as the table leader.

Table leaders. Table leaders facilitate discussion and keep the process on track within their tables. Their primary role is to monitor the group discourse, keep the group focused on the task at hand, and keep time for the group. As needed, table leaders must focus the discussion, find a diplomatic middle ground between participants, or request assistance from DRC and DESE. Table leaders need appropriate skills for group facilitation and should be very familiar with the content standards across multiple grades. Table leaders will also work as a subcommittee at the end of the workshop to study the recommended cut scores for across-grade articulation and, if needed, to recommend defensible adjustments to the cut scores to improve this articulation. Table leaders are members of the workshop committee, not workshop staff.

Workshop Staff

DRC will facilitate the workshop on behalf of DESE. Staff members from DRC will provide training, general facilitation, and data processing. Staff members from DRC will also serve as group leaders, facilitating the standard setting procedure in each room. Lastly, DRC will manage the logistics for the workshop.

Standard Setting Schedule

The standard setting comprises three days of activities. Table 1 provides a high-level daily agenda of the standard setting.

Table 1. Proposed Overview of the Daily Agenda for the Standard Setting

Day	Time	Activity
Day 1	AM	Opening session, workshop training, study standards, discuss ALDs
	PM	Take the test, study OIB
Day 2	AM	Complete study of OIB, supplemental training
	PM	Rounds 1 and 2
Day 3	AM	Round 3
	PM	Discuss recommendations across grades, refine ALDs, evaluate workshop

Benchmarks

At standard setting, *benchmarks* refer to any content- or policy-based information that is presented to participants that comes from an external source. Benchmarks could include achievement level descriptors from other assessments, cut scores derived from parallel tests, or impact data (consequences data) from previous years. The use of benchmarks at standard setting is well-established.² Indeed, many agencies have used benchmarks at standard setting to provide policy-based information to standard setting participants during the workshop. Participants then bring their content-based expertise to bear at the standard setting, joining it with the policy-based benchmarks. Thoughtful use of benchmarks can bring policy- and content-based information together in a meaningful way at standard setting.

The TAC has recommended that DESE consider (a) showing participants benchmarks based on the 2014-15 performance standards, as taken from Missouri’s participation in Smarter Balanced Assessment Consortium; or (b) showing participants benchmarks based on other, respected measures of student assessment. Several potential benchmarks and methods for presenting them to participants are listed here.

- Cut scores associated with 2014-15 impact data. In 2015, Missouri students took the SBAC assessments in ELA and mathematics. In 2016, the MAP assessment will measure the same content standards, but no linking items are available to directly translate the existing MAP cut scores onto the 2016 test scale: presenting the 2014-15 cut scores is not prudent, as the test scales will not be directly comparable. Instead, DRC can use the percentages of students in each achievement level (*impact data*) from 2015, and then identify benchmarked cut scores on the 2016 that yield approximately the same impact data. This approach is often used to create benchmarks in cases where test scales cannot be continued between two consecutive assessment years.

² 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., pp. 225-253). New York: Routledge.

At the standard setting, participants would be asked to evaluate the benchmarked cut scores using the content from the 2016 test. Participants would be told that if they recommended keeping the benchmarked cut scores, the 2016 MAP cut scores would have about the same level of stringency as the 2015 cut scores. Further information about the ways to communicate context to participants about these or any benchmarks can be found below.

- Performance data from other assessments. DESE may wish to show standard setting participants benchmarked information from other respected assessments. DRC recommends that DESE consider using these benchmarks, although DESE may have access to additional benchmark information that may also be presented.
 - the achievement standards for EOC assessments
 - the performance of Missouri students on NAEP
 - the proportion of Missouri students meeting the college and career readiness (CCR) benchmarks on ACT

Information about external assessments can give standard setting participants useful policy-based information about the context in which MAP cut scores are established. For example, participants may see how similarly (or differently) their recommended cut scores impact Missouri students in relationship to the NAEP, or how the percentage of students classified as *Proficient* in grade 8 may be connected to the percentage of students observed as college and career ready in high school.

If DESE chooses to share benchmarks with participants, DRC will indicate them for participants in terms of ordered item book page position (e.g., a “bookmark on page 19” of the ordered item book). Participants will be able to evaluate the knowledge and skills that students would need to demonstrate on the assessment in order to meet the cut score associated with this benchmarked cut score, and then may choose to recommend an adjustment, if needed, as based on the tested content reflected in the ordered item book.

As previously stated, the TAC has recommended that DESE consider using benchmarks at the standard setting. Specifically, the TAC has recommended that if DESE chooses not to reference the cut scores associated with the 2014-15 impact data at the standard setting, that benchmarks calculated from other assessments be used.

If DESE chooses to use benchmarks, DRC recommends that the Department work in partnership with DRC before the workshop to define the exact benchmarks that will be used, as well as the precise language that will be used to describe them. Different messages may be delivered to standard setting participants to describe the context around the benchmarks.

For example, participants may be told that the benchmarks are based on the existing MAP cut scores and are presented purely for their information, and they should use these benchmarked cut scores as a starting point for their deliberations. Alternatively, participants may be told that they are expected to retain the benchmarked cut scores unless there is a compelling content-based reason to change them. DESE may prefer the former language as a way to encourage participants to recommend achievement standards that are similar to the existing ones, albeit

not exactly the same. DESE may prefer the latter language if it is reasonably certain that the existing achievement standards should be kept without adjustment. DRC is happy to discuss these options with DESE prior to the standard setting.

Benchmarks and ALDs

As previously stated in this document, DESE has indicated that the SBAC ALDs, as used during the 2014-15 school year, still reflect the content-based expectations for Missouri students in each achievement level. Additionally, DESE has indicated it may show standard setting participants benchmarked cut scores based on the impact data from the 2014-15 test. In this context, the 2016 standard setting is a form of cut point validation: DESE has the same content-based expectations for students as last year, but seeks to determine the most appropriate cut scores on the 2015-16 assessment which reflect those expectations.

This approach assumes that DESE is reasonably satisfied with the results of the 2014-15 assessment, specifically in terms of the percentage of students classified in each achievement level. Indeed, these impact data may be used as benchmarks to help inform standard setting participants' decisions in 2016. However, if DESE is not satisfied with the impact data and expects the results of the 2016 assessment will be substantially different—for example, if the percentages of students classified in each achievement level for grades 6–8 mathematics will be more than 10% different than 2015—DESE may wish to adjust the ALDs for these grades to better reflect the content-based expectations for students. At the standard setting workshop, Missouri educators will then be able to validate the adjustments to the ALDs, plus will be able to recommend cut scores that link the cut scores and adjusted ALDs by using empirical data and the tested content.

Construction of the Ordered Item Books (OIBs)

The ordered item book (OIB) lies at the heart of the Bookmark Procedure. A separate OIB will be prepared for each grade and content area combination. An OIB is comprised of the test items which are ordered in terms of difficulty. Item difficulty, as defined by its scale location given a response probability (RP) value, is calculated based on data from Missouri students during the spring administration of MAP. Easier items appear earlier in the OIB, and harder items appear later. Items ascend in terms of difficulty throughout the OIB.

Multiple choice and constructed-response items are ordered together in the OIB. For constructed-response items, each non-zero score point is ordered separately. For example, for a two-point item, the constructed-response item appears twice in the OIB: once for the first score point, and once for the second score point. It is more difficult to earn two points on most items instead of a single point: accordingly, the first score point appears earlier in the OIB than the second score point.

Response probability (RP) criterion. To order the items, an RP criterion must be selected in advance of the workshop. For example, at the recent cut point validation for MAP science, an RP criterion of 0.55 was used: the location for each item is defined as the IRT scale value associated with a 0.55 chance of answering the item correctly after guessing is factored out.

(This criterion is sometimes abbreviated as RP55GA.) For a score point, it is defined as the value associated with a 0.55 chance of scoring *at least* that many points.

This response probability criterion was selected because the original standard setting for MAP science used RP55GA, and the cut point validation was designed to mirror the original standard setting process as closely as possible. Accordingly, DESE may or may not choose to use this same RP criterion in the future.

The choice of RP criterion is a policy decision, and different criteria have been used in various standard settings around the country. The most common criterion is 0.67, incorporating an adjustment for guessing (RP67GA).

The decision of which RP criterion to use can be informed by data. Approximately a week prior to the standard setting, DRC will construct OIBs for MAP using various RP criteria. For each of these OIBs, DRC will calculate the percentage of students that would be classified below a cut score associated with each page in the OIB. The resulting data, often represented as curves, show the potential impact data associated with each potential cut score recommendation using each OIB. DRC will then find the OIB page position associated with each of the benchmarks that DESE intends to present to standard setting participants. Ideally, the scale locations (difficulty estimates) of test items will match well with the observed distribution of student scores (e.g., not all the test items are only mastered by the top half of students), and the benchmarks will appear in positions throughout the OIB (e.g., the benchmarks are not all associated with the first few pages of the OIB). Choosing an RP criteria after viewing the results of this analysis can help DESE make an informed decision.

Construction of the OIB. Participants use the OIB to recommend cut scores. Accordingly, it is important that the items included in the OIB span the difficulty continuum—from easy to hard—and that items are found around the points on the test scale where cut scores are likely to appear (e.g., around the benchmarks).

For Grades 5 and 8 ELA, each OIBs will also contain a writing task. When constructing the OIB for each of these two grades, DRC will select a writing prompt that spans the difficulty continuum with score points that are easy, medium, and difficult for students to earn. The 10 points for each writing prompt will be ordered in the OIB alongside the other items from the ELA test. The rubrics for each prompt will be finalized as part of the rangefinding activity, scheduled for April 2016.

Each point will be labeled with the score point and rubric trait (e.g., Point 3 of Organization, Point 2 of Conventions). Accordingly, participants will be able to consider the relative difficulty of earning points on the writing task alongside the difficulty of the other items on the ELA test.

DRC will share the information about the RP criterion and augmentation with DESE just before the workshop, as data become available.

Bookmark Standard Setting Procedure

The first morning of the workshop, all participants will convene in a single room for the opening session. After this training session, participants will be divided into their pre-assigned rooms and tables.

Opening session. The opening session represents the participants' first opportunity to meet the staff from DESE and DRC. It is important that the participants feel appreciated and valued for their content expertise. DESE should set the tone for the workshop in the opening session by describing the development of the MAP assessments, their expectations for the type of cut scores that they anticipate from the standard setting process, and the procedure that will be used to review and approve the cut scores.

The tone DESE sets in the opening session may have a significant impact on the participants' judgments. It is important that DESE share its vision for student achievement, its goals for college and career readiness, and its expectations for the rigor of the standards. DESE staff should be available to answer policy-related questions during the standard setting.

DRC recommends that the state describe the benchmarks used during the standard setting. DRC will reinforce these messages in its training session which follows directly afterwards. By introducing these benchmarks during the opening session, standard setting participants will be better prepared to interpret the benchmarks in the manner intended by DESE.

Workshop training. Following the presentation by DESE, DRC will provide an overview of the purpose of the standard setting and will describe the implementation of the standard setting methodology. The participants will use the following materials.

The *achievement level descriptors (ALDs)* summarize the knowledge, skills, and abilities expected of students in each achievement level. Specifically, participants will be shown the threshold ALDs that summarize the knowledge, skills, and abilities expected of students who have just entered each achievement level.

The *ordered item book (OIB)* is comprised of the test items, which are ordered in terms of difficulty. The ordering is straightforward in that easier items are placed earlier in the book and harder items follow. For constructed-response items, each non-zero score point is ordered separately accordingly to the difficulty associated with earning it.

The *item map* summarizes the materials in the OIB. The item map indicates the order of difficulty, score scale location, item identification number, scoring key, and content standard that each item measures. On the item map, the participants answer two questions: (1) "What does this item measure? That is, what do you know about a student who can respond successfully to this item/score point?" and (2) "Why is this item more difficult than the preceding items?"

DRC will also describe the roles and responsibilities of group leaders, table leaders, and participants to the committee. DRC will thank table leaders for their service during the workshop and will provide additional guidance to their roles in the Bookmark Procedure.

Before the conclusion of the opening training session, participants will be reminded that it is important they keep confidential all of the test items, student data, and cut score recommendations they see during the workshop. Such confidentiality helps protect test equity, the validity of students' scores, and the state's investment in the testing program.

Following this presentation, participants will move to their rooms and pre-assigned tables. DRC will distribute the secure workshop materials to participants, using the guidelines described under the heading "Workshop Security."

Review the content standards and ALDs. Participants will study the content standards and ALDs associated with each assessment. Participants will be asked to consider the knowledge, skills, and abilities detailed in the content standards, and how they are reflected in the ALDs. Participants will study the threshold ALDs, the documents that describe the knowledge, skills, and abilities expected of students “just barely” in each achievement level. During the workshop, these students will be termed the “threshold students.” By engaging in this discussion, participants at each table will begin thinking about the knowledge, skills, and abilities held by these threshold students, and will later be able to consider the threshold students as they make their bookmark placements.

Discuss the Threshold Students. Participants will then reflect on the conversations about the content standards and the ALDs. Participants will then think explicitly about the *threshold students*. A threshold student is a theoretical student whose level of knowledge, skills, and abilities is “just barely” past the point-of-entry for a given achievement level. There are three threshold students for each assessment, one per cut score. By engaging in this discussion, participants from both tables in each group will gain a richer, more commonly-shared understanding of the skills of each threshold student.

Take the test. Participants will briefly review the first operational test, examining the test from the student’s perspective.

Study the writing prompt, rubric, and anchor papers. Participants in Grade 5 and 8 ELA will be given a special training session on the writing prompts contained in their OIBs. During this session, participants will examine the prompt itself, will discuss the scoring rubric, and will study the anchor papers which illustrate the ways students can earn each given score point. After this session, participants will understand how the writing prompt is scored, and they will know what types of writing performance is needed to earn each score point.

Review the benchmarks. Participants will be introduced to the benchmarks for their assessments. This includes an explanation of their purpose and how they should be considered in conjunction with the ALDs. Participants will be reminded that the benchmarks are provided for their consideration: participants should consider the benchmarks as they place their bookmarks later in the process.

Study the ordered item book (OIB). At each table, participants will study each of the items in the OIB in detail as a small group and take notes on their item maps. By studying each item on the test systematically in the OIB, participants will gain an understanding of the knowledge, skills, and abilities that are measured by the test. Participants will be instructed to focus on the elements of the content standards that are measured by each item on the test.

Bookmark training. After participants have completed examination of the OIB, they will be convened in a bookmark training session. During this presentation, DRC will instruct participants how to consider their bookmark placement and how bookmarks represent cut scores. Participants will be asked to keep the threshold students in mind as they place their bookmarks.

At the end of this training session, participants will be given a brief quiz, termed a *checkset*, to test their understanding of the bookmark placement process. This checkset asks participants to answer questions based on the process of bookmark placement to gauge their understanding of the process. Participants are also asked to signal whether they are ready to proceed with the process by checking a box on the checkset; or, if they are not ready, by writing down any questions they may have on bookmark placement. These questions, as well as any potential misunderstandings from the checkset, can then be address before the entire group before participants continue with the Bookmark Procedure. After a discussion of the answers to the checkset questions, participants will then be asked to return to their tables.

Round 1. Individually, participants will consider the knowledge, skills, and abilities measured by the items in the OIB, and will compare these against the ALDs. Using their understanding of the threshold students, participants will place their Round 1 bookmarks for all three cut scores without discussion. As they make their bookmark placements, participants will be asked to develop content-based rationales for each of their decisions, explicitly linking the content measured by the items before each bookmark with the threshold ALDs. Participants will indicate their bookmark placements on a special form for collection by DRC.

Review Round 1 recommendations. DRC will tabulate participants' Round 1 bookmark placements and calculate each table's cut score recommendations. The group's cut score recommendation is associated with the median bookmark placements.

DRC will also calculate *impact data* to present to the committee, as based on participants' median cut score recommendations. Impact data are the percentages of students who would be classified in each achievement level given a set of cut scores. DRC will describe to the group how the cut scores and impact data, were calculated.

DRC will also provide contextual information to participants regarding the impact data. For example, participants in grades 7–8 mathematics may be reminded that approximately 20% of students in grade 8 take the Algebra I assessment instead of the grade 8 mathematics assessment. Accordingly, these participants should consider that these students, who are likely among the more highly-skilled students in mathematics, will not be represented in the impact data for grade 8 mathematics. DRC recommends that DESE be present during this presentation to respond to any policy-related questions that might arise.

In their tables, participants will share their Round 1 bookmark placements with their colleagues. One at a time, in a discussion led by the table leader at each table, participants will share their bookmark placement, along with the content-based rationale behind their decision. Participants will be encouraged to refer to the OIB, item map, content standards, ALDs, and threshold student descriptions throughout this discussion.

Round 2. Following the discussion, participants will again individually consider their bookmark placements. Participants are free to either keep their bookmark placements from Round 1 or change their bookmark placements. All participants make their bookmark placements individually and without discussion.

Review Round 2 recommendation. DRC will then tabulate participants' Round 2 bookmark placements, calculate each table's cut score recommendations, and calculate impact data to present to the committee. DRC will then lead participants in a discussion of the Round 2 recommendations. Participants will be asked to share the rationales behind their bookmark recommendations. Whenever possible, participants will be asked to make explicit reference to workshop materials, such as the OIB and ALDs.

Round 3. Following the presentation, participants will take part in an across-table discussion about their Round 2 bookmark placements. This discussion allows all 10 participants in the group to discuss their bookmark placements together, facilitating group-wide discussions for each assessment. During this discussion, participants will summarize the conversations they had at their tables before Round 2 bookmark placement, will describe their reaction to the presentation of Round 2 recommendations, and will share their content-based rationales behind their bookmark placements. DRC will monitor the conversation to make sure participants focus primarily on content-based topics.

After this discussion, DRC will instruct participants to make their Round 3 bookmark placements. Participants are free to either keep their bookmark placements from Round 2 or change their bookmark placements. All participants make their bookmark placements individually and without discussion.

Review of recommendations across grades. DRC will then present each group with a display of the recommendations made for each grade in the group's assigned content area. These recommendations will be shown in terms of cut score and, if desired by DESE, in terms of impact data. By displaying the recommendations for all grades, participants will have a good sense of how their own recommendations correspond with the recommendations made for other grades.

Each group will be given time to consider the *articulation* of the recommendations across grades. In this sense, achievement standards are well articulated when the cut scores and impact data form a meaningful pattern across grades. Specifically, each group will be asked to consider the extent to which its own recommendations are articulated with the surrounding grades.

Working together, the group may suggest refinements to its recommendations to improve the articulation across grades. For example, a group may note that its median bookmark for a given cut score is on a certain page of the OIB, but that a certain number of bookmarks on either side of that median placement would also be acceptable and would still be congruent with the content standards and ALDs.

Participants will be encouraged to share their views about the recommended cut scores with their table leaders. After this discussion, table leaders will convene to share what they learned from their participants and to make recommendations on how to improve articulation, if needed.

Articulation discussion across grades for table leaders. In the articulation discussion across grades, and as recommended by the TAC, table leaders will make sure the cut scores given to DESE for consideration are reasonable and cohesive across grades. By including table leaders, the voices of all standard setting participants can be represented in the articulation discussion, but the number of discussion participants can still be limited to promote a more meaningful conversation overall.

The table leaders will be shown important content- and policy-based elements of the proposed achievement standards. Panelists will be encouraged to study and evaluate each element, with a special focus on whether the achievement standards are well-articulated across grades. These elements may include:

- The percentages of Missouri students who would be classified in each achievement level
- The proposed cut scores for the tests
- The ALDs for the tests
- External benchmarks of college and career readiness (e.g., NAEP)

Table leaders will then be presented with the cut score recommendations made by all the groups within their content area, along with the associated impact data. Participants will have an opportunity to comment on the level of articulation they observe within the system of cut scores.

During this discussion, the panel will be asked whether they would recommend that DESE adopt the achievement standards as recommended by the standard setting committee, or whether they would recommend changes to one or more cut scores. If the panel chooses to recommend adjusting one or more cut scores, the panel will be charged with providing its rationale for doing so. The panel's rationale may be content-based, or it may include a blend of content- and policy-based information. The panel will use consensus to drive its recommendations. However, if the panel cannot come to consensus, DRC will elicit majority and minority opinions from the panel.

All recommendations will be provided to DESE for review and evaluation, including the articulated cut scores which come from the articulation discussion, as well as the recommendations from Round 3.

Refinement of ALDs. The remaining participants will be asked to recommend possible refinements to the ALDs. DRC will advise participants that the overall level of rigor the ALDs should not be adjusted; however, the participants may have suggestions to make the ALDs more clear or concise for Missouri educators and other stakeholders.

Workshop evaluation. All participants will complete an evaluation of the standard setting. Results from the evaluation will be included in the standard setting technical report and can be used to document how fair and valid the participants perceived the standard setting process, and how satisfied participants were with their cut score recommendation. Subsequently, participants will be dismissed from the workshop with the thanks of DESE and DRC.

Workshop Security

Throughout the standard setting, security is of paramount importance. Secure test materials used during the workshop are numbered and assembled into packets, one per participant. Each participant signs out a specific packet and signs his or her name on each piece of secure material in the packet. DRC staff members monitor the meeting rooms to prevent the removal of secure materials. At the end of each day, each participant's materials are collected and inventoried against a master list. The secure materials are stored overnight in a secure room.

Documentation

Immediately following the standard setting, DRC will be prepared to share information about the participant-recommended cut scores, as well as the methodology used to obtain the recommendation, with DESE and its advisors.

Within five business days following the conclusion of the standard setting, DRC will provide DESE with a preliminary technical report of the standard setting. This preliminary report will contain the recommended cut scores for each round of activity, and tables reporting three types of standard error associated with the cut score: measurement error, cut score error, and a combination of the two. The preliminary report is designed to be used by DESE and its advisors in evaluating the participants' recommended achievement standards, as well as the standard setting process as a whole.

Within 60 business days following the standard setting, DRC will deliver the final technical report of the standard setting. This final report will contain all the information contained in the preliminary report, as well as narrative accounts of the events which took place at the standard setting, graphical representations of participants' judgments, and the results of the participant evaluation. The final report is designed to serve as lasting documentation of the standard setting process.