

**DRAFT**

## Plans for End-of-Course Standard-Setting Activities for the Missouri Department of Elementary & Secondary Education (DESE) - November, 2009

Following are draft proposed plans for carrying out the standard-setting activities to be conducted for MO EOC assessments on November 2, 3, 4, and 5, 2009. These activities apply to seven of the state's End-of-Course (MO EOC) assessments as follows.

Integrated Mathematics II & III, English I, and American History (November 2 & 3)  
Algebra II, Government, and Geometry (November 4 & 5)

Note that standards will be set for the two separate Integrated Mathematics assessments during the same session – that is, a single panel will consider and recommend standards for two separate, though obviously related, assessments. Given the short (40-item) content of these tests, we believe it is feasible and reasonable to expect panelists to make judgments about two tests within the allotted time period.

The above seven MO EOC assessments are being administered operationally for the first time during the 2009-2010 school year. To establish performance standards for these assessments, we will involve six panels – each composed of approximately 15 panelists. The following are Riverside/Questar's proposals, based on our prior experiences in this area, our proposal for the MO EOC program, preliminary discussions with DESE, similar procedures conducted in 2008 with other EOC assessment, and review of the 2008 plans and results with the state's TAC. We welcome additional suggestions in refining and finalizing these plans.

The above pairs of three sessions will be facilitated by Michael Beck, Sheila Potter, and Leon Dreyfus, respectively, Questar's most-experienced facilitators. All have facilitated standards-setting sessions for multiple clients in the past – for both elementary-level and high-school assessments. Riverside will provide a data analyst and a psychometrician; an additional Riverside project professional will also be present for the entire session to assist as needed by DESE or panelists with logistics issues. In addition, representatives of Riverside's content-development group will be present at sessions to serve as resources should any content-related questions arise during the sessions.

Each panel will consist of approximately 14-18 members. One to three members of each panel may be participants in the earlier standards-setting session for other Missouri assessments; others will be new participants. New panelists should not have been members of any of the assortment of committees for MO EOC development activities. Approximately one-half of each panel will be active classroom teachers of the subject of the session; several other panel members will be other professional educators – e.g. administrators or curriculum coordinators. At least a couple of members of each panel should be from "the public" – parents, school board members, representatives of the business community or college professors. DESE is currently recruiting the members of each panel according to these general guidelines. Panel composition along the lines

outlined above will be comparable, to the extent possible, with similar panels used for the 2008 EOC sessions and other Missouri sessions.

It would be desirable for DESE to provide to panelists ahead of time for their review available non-secure descriptive materials concerning the MO EOC. Such materials are likely to be of particular value to panelists who have not previously dealt with MO EOC at its various developmental or tryout phases; this obviously pertains especially to non-educator panelists. If desired, Questar can prepare a short 1- or 2-page overview of the standard-setting process as an orientation to panelists. Descriptive materials should be sent to panelists about two weeks prior to the scheduled sessions.

Panels will use the same performance labels as have been used with previous Missouri assessments – *Advanced*, *Proficient*, *Basic*, and *Below Basic*. A major early activity of these standard-setting sessions will involve the panelists in fine-tuning the achievement-level descriptors (ALDs) for each of the assessments. As a starting point for this activity, we should provide to the panelists the ALDs for the Grades 3-8 assessments in the appropriate content area. These ALDs will serve two purposes – (1) to provide a suggested format for the ALDs, both in terms of structure and depth, and (2) to form a “downward extension” of the ALDs into the middle-school level. As was done in 2008, DESE will develop draft ALDs prior to the standard-setting session. Plans for this activity have been outlined by Riverside elsewhere and will not be repeated here. Standard-setting panels will “fine-tune” these draft ALDs during the first day of their sessions. Finally, panelists should have as a resource a copy of each MO EOC assessment blueprint and the appropriate Course Level Expectations. The Questar facilitators will use all of these materials to lead each panel into an extended discussion and elaboration of the desired ALDs. Panels will be encouraged to use these ALDs throughout the standard-setting session as a reference to their recommendations. In addition, they will make appropriate, though generally minor, revisions and refinements in these descriptors during the sessions and, if desired by DESE, at the end of the sessions.

All standards-setting sessions will involve three rounds of panel recommendations, consistent with the procedures used for establishing performance standards for previous Missouri assessments, including the 2008 EOC assessments. Between the first and second rounds of the panels’ work, panels will be given item-difficulty data for their consideration. Since the MO EOC is being administered operationally for the first time during the 2009-10 school year and essentially no operational tests will likely have been administered by November, these item data will of necessity be data derived from the earlier field test of these assessments. Panelists will be appropriately cautioned about the limitations of such data.

Prior to the last round of ratings for the sessions, panelists will see anticipated statewide “impact data” for the assessments as they are constituted. These are intended to serve as an “anchor” for the panelists’ recommendations. Obviously, again these data will be based on projected statewide score distributions generated from the field-test data. Again, panelists will be cautioned about relying overmuch on these data. Despite the limitations of field-test data for this activity, we believe that providing even these tentative data is desirable both to mirror procedures used for establishing standards for previous Missouri assessments and to provide panelists with an “external reality check” on their evolving recommendations. Note that similar procedures were used for the 2008 EOC sessions.

The specific methodology to be used for the standard setting activities will be a modified Angoff procedure. The state’s TAC recommended this procedure in 2008, and

Questar/Riverside concur with that recommendation. The strongest arguments for this method are that it was used in 2008, plus the fact that only field-test data will be available for use at the time of standard setting. This would make suspect the sequencing of items in the ordered-item booklet required by the “Bookmark” or “item mapping” procedure that has been used for previous Missouri standard settings. The Angoff procedure is a well-recognized and heavily researched method for establishing student performance standards for tests such as the EOC. Since it does not require the availability of item data, it was the method of choice for these sessions.

One of the mechanical aspects of the Angoff procedure that should be addressed by DESE and their TAC is the specific method that panelists use to judge each item. Panelists can be asked to judge EITHER “yes-no” for each item as they make their judgments OR they can judge the percent of borderline students who meet the standard who will answer each item correctly. Both of these methodological procedures have been used successfully in the past. The yes-no approach somewhat simplifies the panelists’ work – that is, they simply judge whether a student who barely meets the standard will answer the given item correctly. With the “percent correct” approach, panelists indicate the *percent* of students barely meeting the standard who will answer correctly. Again, either approach is methodologically acceptable. For the 2008 EOC standard-setting sessions, we used the latter (percent) approach; for that reason, we recommend the same procedure for the current sessions.

Note that all 7 of the MO EOC assessments for which standards will be set this fall are composed exclusively of 40 operational selected-response (multiple-choice) items. Thus, no constructed-response/multi-point items are included in this set of assessments. This will somewhat simplify the panelists’ task.

The EOC operational form intended for Spring, 2010 use will be the booklet on which panelists will make their judgments. We selected this form from the several available operational forms as it is the booklet that will be most widely used.

The process to be used requires panelists to read each item in the test booklet, consider the content standard assessed and the cognitive challenge posed by the item, and then estimate the percent of borderline Basic students who will answer the item correctly. Panelists then make like estimates for borderline Proficient and Advanced students, then move to the next item in the booklet to make three comparable decisions. This activity repeats for each item in the operational test. Item-difficulty data will be provided to panels prior to the second round of judgments; anticipated statewide impact data will be shared prior to the final round of judgments. All of these methodological events are consistent with typical modified Angoff activities and with the procedures used in 2008.

Panelists will clearly be told that their work is purely advisory to DESE, which will then consider these recommendations and select the final cut scores for each assessment.

### ***Outline of Specific Session Activities:***

The appended Session Agenda provides an outline of the proposed specific elements of the panel sessions along with a general guide as to the time devoted to each activity. However, these major session components are summarized below. Note that identical processes – in fact, the same slides and scripts – will be used in all six concurrent sessions to minimize any inter-session differences related to “facilitator” or “session” variance. All PowerPoint slides and related scripts will be prepared as soon as the overall session plans are approved.

Drafts of session scripts will be provided to and preapproved by DESE. Note that three of the sessions will be held on Monday and Tuesday; the other three sessions will be conducted (with totally or predominantly different panelists) on Wednesday and Thursday.

We arbitrarily conceive of the sessions as being made up of essentially nine distinct elements:

1. General Process Overview. Three separate activities will occur during the first 100-120 minutes of the 2-day session. First, DESE representatives will orient the panelists to the MO EOC program. This presentation includes a discussion of the current state of the program, the new elements introduced in the current academic year, and short- and long-term plans for the future of the program. This portion of the presentation ends with an overview of the specific assessments of focus for these two days of sessions.

The second initial activity is a brief general overview of “what is standard setting?” Michael Beck will lead this session for all panelists to ensure common understanding across all panelists of the fundamental elements of the process. Included at this time is a brief overview of the *general* process of establishing student performance standards, ground rules for how this activity occurs, and some general recommendations of key elements to focus on when attempting to recommend standards. Panelists will be told during this portion of the session that their work is advisory to DESE. This general overview of the standard-setting process will be expanded upon individually by each facilitator after the panels break into their content-specific groups.

The final activity will be a general introduction to Achievement-Level Descriptors (ALDs) and their importance to the overall process. Since the panels will be reviewing, editing, and expanding upon draft versions of the ALDs provided by the state, it is critical that they undertake this task seriously and that they understand the key role of ALDs in the standard-setting process. Sheila Potter of Questar will provide this general overview. Following this activity, panels will break into the three content-specific groups, where all remaining work for the sessions will take place.

2. Preview/“Take” the Assessments. Upon reconvening in the content-area subgroups, panelists will first introduce each other and sign DESE-provided security agreements. Following a continuation and extension of the overview of the process in which they will later be engaging, they will then review operational tests on which they will shortly recommend standards. For this activity, panelists will use the test-administration procedures, the actual test content, and all relevant scoring materials. It would also be desirable for them to have a copy of the test blueprint for reference purposes. The actual spring, 2010 operational tests will be used for this activity. Field-test items contained in these booklets will be removed from the booklets viewed by the panelists. Since these are “live” materials, the facilitators will stress the confidentiality of all of the items.

The primary purpose of the activity is to familiarize panelists with the actual, complete assessment content prior to beginning the standards-setting judgments. Following this review of the tests, each panel will spend a short time reacting to the assessment content – difficulty, sources of challenge, scoring issues, general and specific reactions. Experience has shown that providing this opportunity in a setting slightly distanced from the standards-setting activity expedites the later portions of the session. It also provides panelists – especially those not previously familiar with the assessments – with a mental “set” concerning the concrete definition of *proficient* as it conveyed in the assessments.

3. Discuss and “Fine-Tune” Achievement-Level Descriptors (ALDs). All three panels will begin this activity with a brief review of the ALDs for the elementary/middle-school levels of the Missouri assessment series. These previously established ALDs serve both as a model for and as a lower-level extension of the work of the MO EOC panels. The panel discussion will then review draft ALDs for the particular content area of concern for each of these sessions. These draft ALDs will have been developed and approved by DESE prior to standard setting. The ALD review activity is highly interactive, with panelists suggesting changes and other refinements to the draft ALDs. The essential task is to itemize specific performance-related behaviors indicating performance at the *Advanced*, *Proficient*, *Basic* and *Below Basic* levels in the content area. The activity is a “brainstorming” one in which each panel member’s ideas are recorded and considered without expecting consensus. Panel suggestions are written on chart paper around the room, and these pages along with the DESE-provided drafts are referred to by panelists throughout the actual judgment activities. The refinement of ALDs is an essential element of the entire standards-setting process. We expect to devote approximately 1 ½ hours to this activity. Our facilitators are highly experienced in involving all panelists – educator and non-educator – in this activity. The thoroughness of the ALD-refinement activities and the extent to which the panelists individually and as a group internalize these significantly impacts the soundness of the subsequent standards-setting activities. (Note that, as with the later standards-recommending process, panelists working with Integrated Mathematics content will have to carry out ALD-refinement activities with two separate though related tests.)

At the conclusion of the standards-setting sessions, DESE will collect the panelist recommendations for revisions in the draft ALDs for consideration in the wording of the final ALDs for the assessments.

4. Orientation to the Angoff procedure The final activity leading to the first round of committee judgments is an orientation to the specific task of carrying out a modified Angoff standards-setting process. Included in this overview is a reminder to focus on the *threshold* of performance in each category, a review of what each performance descriptor means, a reminder to focus on MO EOC students statewide, etc. While the MO EOC assessments will yield four categories of student performance, “cuts” are made at only three locations in the distribution. Panels will be guided to understand the process of making three determinations of the percent of borderline students who will answer each (multiple-choice) item correctly. Panelists are reminded that their judgments are anonymous and that it is important that they work independently and maintain security in the materials.

Panelists indicate their judgments on specially designed rating sheets developed for each test. These rating sheets list item percent-correct values for each item; panelists simply circle three such values to indicate the percents-correct for borderline students in each performance level.

The orientation to the Angoff technique ends with panelists conducting a “mini-session” of ratings using a practice test of approximately ten multiple-choice items. This exercise provides panels an opportunity to practice the Angoff judgment process and procedures for recording their recommendations prior to beginning work on the “real” test. The practice test enables facilitators to check on panelists’ understanding of the mechanics of the technique and corresponding recording of judgments.

Panelists typically take about 15 minutes to complete this practice activity. After this, facilitators will lead a review of the practice activity, seeking judgments from each panelist to ensure their understanding of the basic Angoff procedure.

Finally, a short “self-assessment” instrument will be administered to each panelist. This assessment gauges panelist understanding of the basic elements of the process in which they are about to engage. It permits the facilitators to reemphasize key elements of the process as well as giving the panelists an opportunity to check their understanding.

5. Round 1 of Judgments, with Feedback and Discussion. The final Day 1 activity involves the panelists independently making initial, “Round 1” judgments of item-by-item performances that together define Advanced, Proficient, Basic and Below Basic. At this stage, panel work is both anonymous (via judge numbers known only to Questar facilitators) and independent. Should panelists have questions during this stage, they are asked of the facilitator. The first round of judgments for this activity is likely to require approximately 75 minutes of the typical panelist. However, there is no time limit for this activity; individual panelists may require as much as 2 hours to complete their judgments. Typically, some panelists are still struggling at this time to understand the task, thereby requiring more time to make their judgments; others simply work more deliberately. Panelists will turn in their judgment forms and other materials as they complete their Round 1 judgments and then are free to leave for the day.

Panelists are told to read and make performance judgments about each successive item in the test booklet. When reading the item, panelists are to consider the underlying Course Level Expectation being assessed, the task(s) required of the student, its difficulty and importance. Then they are instructed to decide what percent of students taking this assessment who perform at a minimally *Basic* level in this content area should be able to answer Item 1 correctly. Panelists then decide what percent of minimally *Proficient* students should be able to answer Item 1 correctly. Finally, they make a comparable decision concerning minimally *Advanced* achievement in this content area, and indicate this percent. They then move on to each successive item in the booklet, making the same three sets of judgments about each item. A partial sample of an item-judgment form for the Angoff procedure (drawn from another state program with only three performance categories) is illustrated in Appendix C.)

**Day 2** will begin with an overview of the previous day’s activities and outcomes. We propose to provide an additional “skim” of the ALDs to initiate this day.

6. Feedback and Discussion of Round 1 of Judgments The first panelist activity for this day begins with an overview of the results of Round 1. Riverside data analysts will prepare graphic displays of the Round 1 results for panelist inspection and consideration. The first such graph displays all three cut scores (Basic, Proficient, and Advanced) in terms of raw score for each panel. This is a display of all panelists’ sets of judgments anonymously on a single graph. The second graph is a bar display of the same information, showing both cuts on the same graph for the group of panelists. Samples of these two displays (drawn from non-MO EOC instruments) are shown in Appendix B are provided for DESE and TAC review. Note that these displays are “anonymous;” that is, the Judge Numbers shown on these graphs are not the judge numbers assigned during the rating process. These – and/or related – graphic displays of the Round 1 recommendations are used as the foundation for a

discussion of the procedures used by the panelists in coming up with their interim recommendations, both individually and as a group. Should DESE have had successful experience with graphic feedback devices for earlier Missouri standard-settings, we would be willing to use these instead or additionally.

Using the Round 1 results, facilitators then lead an extended discussion of the Round 1 judgments. Most of the work at this time focuses on the interim judgments of panelists on individual test items. Given the length of the MO EOC assessments, it should be possible to discuss with the panel all or nearly all of the individual items. All panelists are actively involved in the discussion to permit facilitators to gauge whether or not the panelists are indicating the item percent-correct values that they intend, that the reasoning processes they follow in making their judgments are consistent with good practice, and that the mechanics of indicating recommendations are clear. Throughout these discussions, facilitators continually focus on the key elements of the standards-setting process: establishing the *threshold* of each cut, projecting the cuts for a *statewide* population of these students, focusing on the particular *course* and *performance level* of the target populations, and correct Angoff standards-setting thought and mechanical judgment-marking processes.

It is important to note that these discussions are “full-panel” discussions, not carried out in subgroups. This permits facilitators to ensure that all panelists understand the process and are carrying out the judgment process correctly. It also ensures (much like a jury activity) that all panelists have the opportunity to hear each of their peers’ comments before making their later recommendations. This phase of the process generally consumes about 90 minutes in the typical session; facilitators will permit discussion to continue until they perceive that all panelists are prepared to make their second round of judgments.

Following a thorough discussion of the initial cut recommendations, panelists will review statewide “item difficulty” data. These data will of necessity be field-test difficulty data collected during the field test of the MO EOC assessments. These data will be *p*-values, as conveying useful information concerning any latent-trait difficulty indices to a lay audience is challenging. Panelists will be told the source of these data (the statewide field test) and clearly told that the data are to be used as each panelist sees fit. We typically instruct panelists that these data are *relevant to* but *not critical to* the activity of establishing performance standards.

7. Round 2 of Judgments/Recommendations. After panelists clearly understand the meaning of the item-difficulty data, Round 2 of judgments takes place. Judges again work independently to make a second set of judgments for each item on their assessment. Judges are free to maintain their Round 1 judgments or to revise them for one or more items as deemed appropriate. Prior to beginning this round of judgments, panelists are again reminded of the key elements of the process, and are focused specifically once more on the ALDs for their assessment. Again there is no time limit, although this round typically takes significantly less time than does Round 1 because most panelists more clearly understand the judgment process. In addition, they are increasingly familiar by this point with the specific items upon which they are making the judgments. Further, many panelists begin – indeed, are encouraged - to make some or all of their Round 2 item-based decisions during the discussion of Round 1 results.

As panelists complete their Round 2 work and record their recommendations on the judgment forms, they submit these forms and leave the room for lunch. All materials used for the sessions remain in the room.

8. Round 2 Feedback and Discussion. What occurs at this stage is very similar to, although more focused than, the presentation and discussion following Round 1. Panelists again see graphic displays of their judgments. These displays are used to guide another discussion of specific items – typically at this point, items on which the variability of panelist judgments in Round 2 was greatest. (This information could be shared with panelists as DESE and TAC desire. Our typical recommendation with regard to sharing panelist judgments at an item level is to display such data only after Round 2 and only for selected items, although TAC experiences in this regard are welcome.) Discussion remains focused on the cognitive task(s) underlying the particular item/activity/score point, not on a statistical judgment concerning the interim cutscore recommendations.

Following this discussion, we will provide panelists with anticipated/estimated “statewide impact data,” that is, the percents of students statewide whose performance will likely be labeled Below Basic, Basic, Proficient, or Advanced were the panels’ Round 2 judgments to be adopted. The median and/or mean Round 2 judgment is used to convey this information; we strongly prefer not to provide these data individually by judge to reduce the likelihood of panelists attempting to make their final judgments on essentially a “norm-referenced” basis. Again, judges are told that the “impact data” are *relevant to, but not the essential aspect of*, establishing student performance standards. This cautionary information is especially important in the case of MO EOC as the data upon which this information is based are not derived from operational administrations of the assessments. We believe that it would be wise to inform judges at this stage of DESE’s intent to review the final recommended standards with an advisory panel near the end of the school year, when a representative and sizable proportion of the anticipated statewide population of students have been scored. Judges are free to consider and either use or dismiss these impact data in making their final judgments.

As soon as facilitators are comfortable that all panelists are prepared to make their final recommendations, Round 3 occurs. Again all judgments are made anonymously and independently.

9. Round 3 of Judgments and Meeting Evaluation. Panelists are given unlimited time to complete their Round 3/final recommendations, although most panelists will likely complete their judgments within 30-40 minutes. All panelists clearly understand that only the Round 3 judgments “count” as their recommendations, and that the three rounds are not combined in any way to form the proposed cuts.

Immediately following the final recommendations, panelists complete a short written evaluation of the process, requesting their opinions of the success of each major element of the meeting, the salience of various possible elements in making their judgments, their comfort with and confidence in their judgments on a round-by-round basis, the other comments they wish to make concerning the sessions. We assume that the DESE form used for this purpose in 2008 will be used again for these sessions; this will permit easier comparison of panelist judgments, should that be desired. Riverside and Questar will be responsible, as requested, for summarizing the results and including these in the final process report.

After completing the evaluation, individual panelists are free to leave the sessions (depending on DESE’s decision regarding a review of the ALDs – see the following paragraph). No feedback is typically provided to them as to their final recommendations, although this recommendation is subject to DESE decision. We anticipate that DESE will make the final report of the process – including round-by-round results, any adjustments to

these recommendations made by DESE, and interim standards – available either on paper or via the Internet to all panelists. If that is the case, we would so inform panelists, as they are typically very eager to see their final results.

An optional final task of the sessions could be to reconvene the panels for a short period for a final review of the ALDs. Since the panels have just finishing using and applying these ALDs to an operational test, it may be useful to give them a final opportunity to “fine-tune” or revise the statements prior to their departure. The only substantive argument against holding such a final review session is that it requires all panelists to remain in the area until all panelists complete their final judgments. This potentially conflicts with the desire to permit all panelists to have unlimited time to complete their final recommendations, and could make slow-working panelists feel pressured to complete their ratings on a faster schedule than they would prefer. TAC counsel on this issue is sought.

Note that all of the activities outlined in the nine arbitrary steps above are consistent with those used in standard-setting sessions conducted for the EOC assessments in 2008.

### ***Responsibilities:***

Riverside and Questar will advise or otherwise assist DESE as requested with the solicitation of potential panelists or providing information related to the sessions to panelists after their selection. DESE’s responsibilities will be to select and maintain contact with panelists and to approve the general procedures and, as details firm, the slides and scripts for the sessions. DESE is responsible for arranging for the location for the sessions, making travel and related logistical arrangements for panelists, and providing required logistical support during the sessions. Finally DESE will plan and lead - with Riverside support as requested – the process of developing the draft ALDs for each of the seven assessments.

Riverside will produce all materials required for the standard-setting sessions well in advance of November 1. Riverside will also prepare a handout for panelists of item difficulty values by assessment. These data will be provided to panelists prior to Round 2.

Appendix A to this plan provides a proposed detailed agenda for the sessions as described above. These agendas are subject to additional review and revision, but they convey the essential elements of the proposed sessions as described above.

Panel sessions will be conducted at a Missouri location to be determined on November 2, 3, 4, and 5; the draft report of results will be prepared jointly by Riverside and Questar staff and will be submitted to DESE by the end of November.

We will require further discussions with DESE about the process of reviewing the panels’ recommendations, for considering the desirability of making any indicated adjustments or “smoothings,” and for any desired late-spring review of the standards using operational data prior to “officially” establishing the standards. Subsequent to DESE’s official approval, Riverside (with Questar assistance) will prepare a final report of the procedures, outcomes, and relevant information to document all activities.

## Appendix A

# MISSOURI EOC DRAFT SESSION AGENDA

## Standards-Setting Sessions

### Day 1 – MORNING

**8:15 – 8:45**

#### **Welcome, Introductions, Logistics**

(DESE staff)

*(Large-Group session – all panels together)*

- Place of this activity in the overall MO EOC schedule
- Logistics – expenses/honoraria, schedule, problem-solving

#### **Overview of MO EOC Assessment System**

(DESE)

- Alignment of assessments with Course Level Expectations
- Overview of MO EOC assessments
- How the MO EOC assessments are scored

**8:45 – 9:30**

#### **Overview of the two days of sessions**

(Mike Beck)

- Agenda
- Delimit the panels' activities – "Ground rules"
- What does it mean to set "performance standards"?

**9:30 – 10:00**

#### **Intro. to Achievement-Level Descriptors (ALDs)**

(Sheila Potter)

- What are ALDs?
- Why are they the foundation of setting standards
- Merging state content standards and assessment results
- How the panels will "polish" the draft ALDs, starting with general descriptors, ALDs for earlier grades, and the draft ALDs developed in advance of these sessions by separate panels
- What do we to *do* in *this* session?

*(Panelists now break into 3 individual groups, separately facilitated; all subsequent panel work will take place in the separate sessions.)*

**10:15 – 11:15**

#### **Setting Performance Standards – General Process**

- Overview of the general process of setting standards
- Process of placing cut scores to segment a continuum of performance
  1. Drawing a discrete cutoff (threshold students)
  2. Errors of classification in any measurement process
  3. Why multiple rounds are required
  4. Keys to making good judgments
  5. What happens *next* – panels as advisory, not decision-makers

**11:15 – 12:00**

**“Experience” the Assessments**

- Review/“Take” the assessments on which standards will be set – answer the questions, take notes
- Discuss the test – content, concerns, difficulty, “construct” issues
- Relationship among the assessments, the content standards, and the performance descriptors

**Day 1 – AFTERNOON**

**1:00 – 2:20**

**Definitions and Description of Performance Standards**

- Generic Achievement Level Descriptors provided by the state and ALDS for elementary assessments - their import/use
- What does *mean* for a student to be described this way – What can these students *do*? What do they *know*?
- Review the DESE-created draft ALDs
- Use flip charts to generate panel suggested changes/revisions/additions/deletions for the draft ALDs in the specific content area; review these and ensure panel comfort with the ALD descriptors as evolved.

**2:20 – 2:50**

**Orientation to the Specific Standard-Setting Methodology**

- “Mechanics” of setting standards
- What is the “Angoff procedure” and how does it work
- Judges’ task
- Features of the procedure to be used
- How to indicate your judgments
- Practice Exercise & Discussion

**3:00 – 3:30**

**Preparation for Round 1 of Judgments**

- Reminders of key issues – threshold, ALDs, all MO EOC students
- Distribute materials and orient panelists to use
- What to do – mechanics of judgments for each item and all cuts
- “Self-Evaluation” of standard setting readiness & discussion
- Rules for judgments – anonymity, independence, security of materials

**3:30 – 5:00**

(or until completion)

**First Round of Judges’ Work**

Panelists work independently, recommending cuts for *their* particular MO EOC, turning in their rating sheets and leaving for day when completed.

## Day 2 – Morning

**8:30 – 9:15**

### **Review of Day 1 Activities and Discussions**

- ALDs, their importance to the process – any additional thoughts or recommended additions/deletions
- MO EOC assessment
- Standard-setting general methodology and outcome
- “Angoff” mechanics – what to do
- Round 1 trials and tribulations
- Questions/Observations of judges to the process in Round 1

**9:15 – 10:45**

### **Feedback & Discussion of Round 1 Judgments**

- Round 1 feedback by test – Graphic portrayal of panelists’ judgments (anonymous)
- Meaning of Round 1 judgments - distribution of cuts, median/mean cut
- Discussion of *WHY*'s for Round 1 (i.e., what led panelists to set their standards as they did? Problems, issues, confusions, rationales for preliminary standards)
- Discussion of selected items
- “Shaping” of panelists’ considerations, focusing on critical considerations (threshold performance, “should vs. will,” ALDs, item rating procedural confusions, construct issues)
- Purpose of Rounds 2 & 3 – reflection, reconsideration, and comfort, not consensus
- Present statewide student performance data (*item difficulty* values from the statewide field test)
- What the data *mean*; why they are only minimally useful in setting standards
- Reminder of key considerations

**11:00 – 11:15**

### **Preparation for Round 2 Judgments**

**11:15 – 12:15**

(or completion)

### **Round 2 of Judges’ Work**

Opportunity to reconsider and adjust Round 1 judgments

*(A buffet lunch should be available from about 12:00 – 1:15, as panelists will complete their work on different schedules.)*

## Day 2 – AFTERNOON

**1:30 – 2:45**

### **Review of Round 2 Judgments**

- Questions/Observations of judges on the process
- Feedback and discussions much like that for Round 1
- Projected statewide “impact data” – implications of the Round 2 recommendations
- Discussion of draft cuts across assessments, and the desirability of consistency across MO EOC’s
- Discussion of selected items

**2:45 – 3:00**

**Preparation for Final Judgments**

- Evaluation forms
- Questions, reminders

**3:10 – 4:00**

**Final Round of Judgments & Evaluation**

*(or until completion)*

(panelists are free as they finish work and turn in all materials and their evaluation forms)

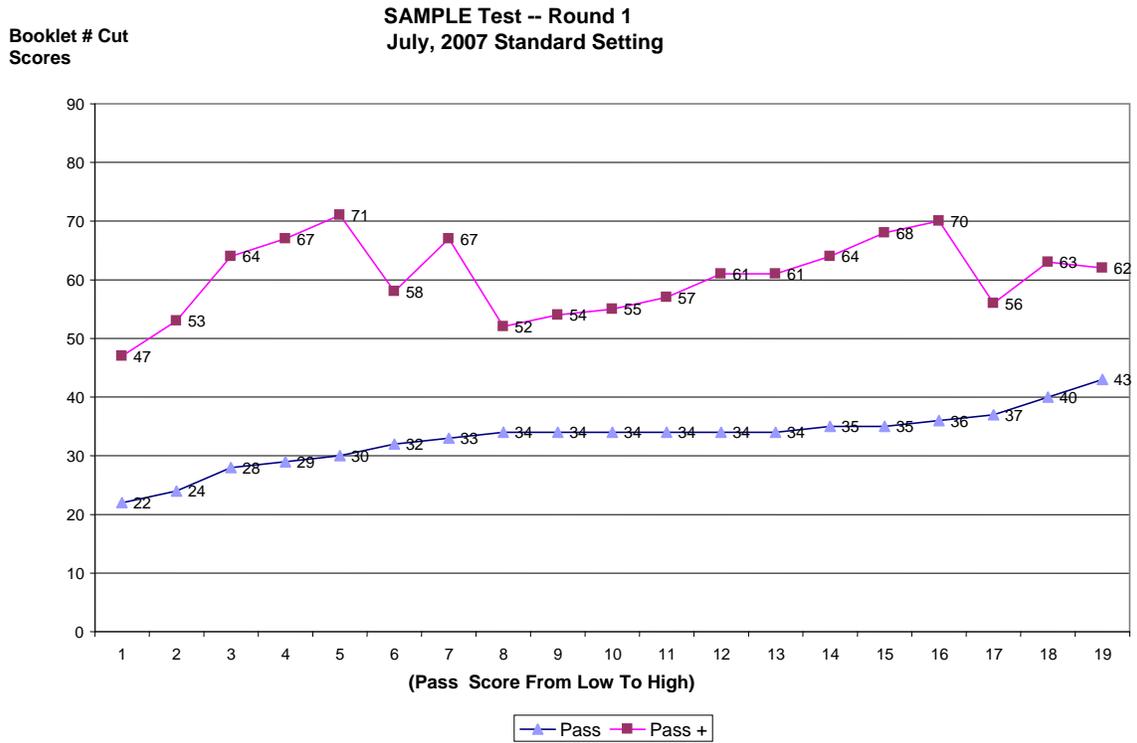
**4:15 – 4:30**

**Final review of ALDs & Session Wrapup *(as desired by DESE)***

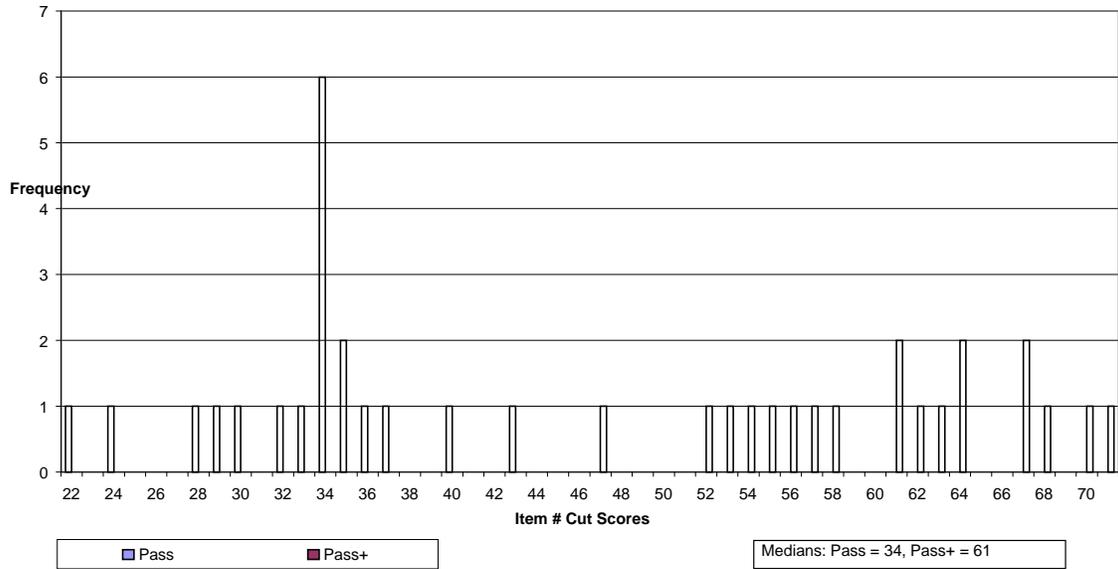
- Reconvene each panel for a short time to give a final review to the ALD lists. This permits any final refinements to the ALDs.
- Wrapup and thanks from DESE representative

## Appendix B

Sample displays provided to panelists following Rounds 1 and 2 of standards-setting sessions.



STATE XXX- Biology I Total Test -- Round 1  
August 1, 2007 Standard Setting



**Appendix C**

**XXXXX Standard Setting – Judges’ Rating Sheet**

**TEST: World Geography**

**Judge No.** \_\_\_\_\_

**Directions:** What percent of **borderline**/threshold students of each description should answer each of the test questions correctly?

Item No.	Description	Percent of Borderline Students Who Should Answer Correctly														
1	Proficient	20	25	30	35	40	50	60	65	70	75	80	85	90	95	100
	Advanced	20	25	30	35	40	50	60	65	70	75	80	85	90	95	100
2	Proficient	20	25	30	35	40	50	60	65	70	75	80	85	90	95	100
	Advanced	20	25	30	35	40	50	60	65	70	75	80	85	90	95	100
3	Proficient	20	25	30	35	40	50	60	65	70	75	80	85	90	95	100
	Advanced	20	25	30	35	40	50	60	65	70	75	80	85	90	95	100
4	Proficient	20	25	30	35	40	50	60	65	70	75	80	85	90	95	100
	Advanced	20	25	30	35	40	50	60	65	70	75	80	85	90	95	100
5	Proficient	20	25	30	35	40	50	60	65	70	75	80	85	90	95	100
	Advanced	20	25	30	35	40	50	60	65	70	75	80	85	90	95	100
6	Proficient	20	25	30	35	40	50	60	65	70	75	80	85	90	95	100
	Advanced	20	25	30	35	40	50	60	65	70	75	80	85	90	95	100
7	Proficient	20	25	30	35	40	50	60	65	70	75	80	85	90	95	100
	Advanced	20	25	30	35	40	50	60	65	70	75	80	85	90	95	100
8	Proficient	20	25	30	35	40	50	60	65	70	75	80	85	90	95	100
	Advanced	20	25	30	35	40	50	60	65	70	75	80	85	90	95	100
9	Proficient	20	25	30	35	40	50	60	65	70	75	80	85	90	95	100
	Advanced	20	25	30	35	40	50	60	65	70	75	80	85	90	95	100
10	Proficient	20	25	30	35	40	50	60	65	70	75	80	85	90	95	100
	Advanced	20	25	30	35	40	50	60	65	70	75	80	85	90	95	100
11	Proficient	20	25	30	35	40	50	60	65	70	75	80	85	90	95	100
	Advanced	20	25	30	35	40	50	60	65	70	75	80	85	90	95	100
12	Proficient	20	25	30	35	40	50	60	65	70	75	80	85	90	95	100
	Advanced	20	25	30	35	40	50	60	65	70	75	80	85	90	95	100
13	Proficient	20	25	30	35	40	50	60	65	70	75	80	85	90	95	100
	Advanced	20	25	30	35	40	50	60	65	70	75	80	85	90	95	100
14	Proficient	20	25	30	35	40	50	60	65	70	75	80	85	90	95	100
	Advanced	20	25	30	35	40	50	60	65	70	75	80	85	90	95	100
15	Proficient	20	25	30	35	40	50	60	65	70	75	80	85	90	95	100
	Advanced	20	25	30	35	40	50	60	65	70	75	80	85	90	95	100
16	Proficient	20	25	30	35	40	50	60	65	70	75	80	85	90	95	100
	Advanced	20	25	30	35	40	50	60	65	70	75	80	85	90	95	100

• • • •

•

•

•