Kentucky Science Assessment Systems

Standard Setting Meetings

Grades 4 and 7

July 2018

Pearson

Version 1.0

Table of Contents

Executive Summary

This report summarizes the process and results of setting performance levels for the Kentucky Science Assessments for grades 4 and 7. The Kentucky Department of Education (KDE) and Pearson (Science assessment contractors) recommend the achievement levels shown in Table 2 of this report for adoption by KDE, State Board of Education, and the Commissioner of Education.

Kentucky Science Assessment Standard Setting Process and Results

Performance levels are used to classify and describe student performance on an assessment. In order to classify student performance into the different performance levels, the following components are generally required: 1) policy definitions, 2) Performance Level Descriptors (PLDs), and 3) cut scores. Policy definitions describe the performance levels in general terms that apply to all grades. PLDs illustrate the performance levels in terms that are specific to a grade. Cut scores represent the lowest boundary of each performance level on the scale.

The process of recommending performance standards for the Kentucky science assessments for grades 4 and 7 was in line with national best practice for standard setting. Results and details of the process are presented in the following sections.

Policy Definitions

Policy level descriptors for the Kentucky science assessments are shown in Table 1. The titles and descriptions of the achievement levels were defined to be part of a cohesive assessment system, and the achievement levels indicate a student's ability to demonstrate mastery on the Kentucky Academic Standards (KAS).

Performance Level Descriptors (PLDs)

There are two types of PLDs used as part of this standard setting: range and borderline. Borderline PLDs represent the knowledge and skills of a borderline student, one that is just barely past the point-of-entry for the performance level. As part of the standard setting process, participants developed borderline PLDs for the respective grade. These borderline PLDs were then used by the standard setting participants to recommend cut scores. The range PLDs represent the range of knowledge and skills a typical student in the performance level would likely demonstrate. Participants of the standard setting meeting developed range PLDs at the end of the standard setting using an anchored PLD development process. The recommended cut scores were used to divide Science assessment content (i.e., items) into the four performance levels and range PLDs were created using those item groupings. The range PLDs from the standard setting participants are being reviewed by KDE and will appear in the score reports release.

Table 1. Policy level descriptors for the Kentucky Science Assessment

Performance Level	Policy Level Descriptors
Distinguished	A student performing at the *Distinguished* level has a **comprehensive** understanding of science concepts and practices. The student consistently communicates ideas in a sophisticated and complex manner, using thorough supporting detail and explicit examples. The student reasons and solves problems by using appropriate strategies in an insightful way. Connections between science concepts/ideas, when appropriate, are justified and insightful.
Proficient	A student performing at the *Proficient* level has a **broad** understanding of science concepts and practices. The student usually communicates ideas accurately using clear and appropriate examples, supporting or justifying those ideas with relevant details and evidence. Problem-solving and critical thinking skills are used effectively. Connections between science concepts/ideas, when present are reasonable and appropriate.
Apprentice	A student performing at the *Apprentice* level has a **basic** understanding of science concepts and practices. The student communicates ideas in a basic manner, but explanations, solutions or justifications may be unclear or ineffective. The student demonstrates some problem-solving and critical thinking skills, but they are not consistently applied.
Novice	A student performing at the *Novice* level as a **minimal** understanding of science concepts and practices. The student communicates ideas ineffectively or inaccurately, providing little detail and little or no support. Attempts at problem solving or critical thinking are minimal or inappropriate.

Cut Scores

In order to create a common point of reference across the science assessments, cut scores and measures of student performance on the Kentucky science assessments are translated to a scale that ranges from 100 to 300 points and has a Proficient cut of 210. The common value of 210 for the Proficient cut score across assessments do not mean that they reflect the same difficulty, or that achievement levels can be compared in difficulty through the scale values of their cut scores across grades. Similarly, the percentage of students in a performance level is not directly comparable across grades. The population of students tested is different for each assessment. Performance levels from different tests are not comparable because the cut scores for these tests are criterion referenced—they are based on content-specific expectations of what students should know and be able to do.

The cut scores recommended for adoption are shown in Table 2. This table shows the scale score ranges corresponding to each performance level. The cut scores for the performance levels are the lowest cut score within each range. There is no cut score for Novice, since 100 is the lowest attainable scale score a student can earn.

Table 2. Cut Score Ranges for Kentucky Science Assessment Performance Levels

	Raw Score Ranges	
Performance Level	**Grade 4**	**Grade 7**
Distinguished	226 to 300	229 to 300
Proficient	210 to 225	210 to 228
Apprentice	191 to 209	192 to 209
Novice	100 to 190	100 to 191

Details pertaining to the general method for obtaining the recommended cut scores are provided below.

General Method

From July 17 to July 19, 2018, after the first year of operational administration, a standard setting committee meeting was conducted to provide cut score recommendations for the Kentucky science assessments for grades 4 and 7. Each committee was comprised of 15 individuals, including teachers and non-teacher educators. The participants were selected for the standard setting committee to provide content and grade-level expertise during the committee meeting and be representative of the state teaching population, including geographic region, gender, ethnicity, educational experience, community size, and community socioeconomic status.

The Extended Modified (Yes/No) Angoff standard setting method was used as the standard setting meeting (Davis & Moyer, 2015; Plake, Ferdous, Impara, & Buckendahl, 2005). This is a content- and item-based method which leads participants through a standardized process where they consider expectations of student performance, as defined by the borderline PLDs, and the individual items administered to students to recommend cut scores for each performance level. Since the items are presented in clusters on the assessment, the participants used the same process to provide judgments for the item clusters. The standardized process was used by the committees for each grade, which resulted in cut score recommendations.

The process started with participants experiencing the science test for the respective grade from the spring 2018 administration using paper test books from the spring administration. Based on their experience with the test items and a review of the borderline PLDs, participants reviewed each item on the test and answered the following question for each performance level:

“How many points would a borderline student at the [specific] performance level likely earn if he or she answered the question?”

The cut score recommendation for each individual participant was the expected raw score a borderline student at the respective performance level would likely earn, calculated as the sum of the individual item judgments. For the purposes of the standard setting, “likely” was defined as 2 out of 3 students at the borderline of the performance level. Each recommended cut score from the standard setting committee is the median of the recommendations from the individual participants in the committee.

It was determined by KDE that it would be beneficial to have a review the recommendations with the perspective of additional external data. A follow up meeting was held online for 100 minutes on Monday, July 30 to allow a subset of the standard setting participants to view the impact data resulting from their initial cut score recommendations along with the impact data from other assessments to consider whether and to what extent adjustments to the recommended cut scores may be warranted. The additional data presented during the meeting included:

- K-PREP Science for Grades 4 and 7 from 2012 through 2014
- National Assessment of Educational Progress (NAEP) Science for Grades 4 and 8 from 2015
- K-PREP Mathematics for grades 4 and 7 from 2012, 2013, and 2017

After a review of the data, the panelists discussed adjustments to the recommendations based on a desire to honor the work from the standard setting process, maintain high expectation for performance as represented by the standards, and to ensure the results were coherent and defensible. The cut scores recommended for approval are based on the recommendations from this follow-up meeting, building on the work of the standard setting meeting.

Following this meeting, KDE and the commissioner conducted a final reasonableness review of the cut score recommendations. The cut scores for each grade were reviewed to determine the reasonableness of the system of standards recommended, based on policy considerations. Based on this review, KDE adjusted the cut score recommendations for the Proficient and Distinguished performance levels for grade 7.

Results for Kentucky Science Assessments – Grades 4 and 7

Table 3 shows the percent of students who took the Kentucky science assessments for grades 4 and 7 during the Spring 2017-2018 administration that would be classified into each performance level based on the cut score. The percentage of students in an achievement level is not directly comparable across grades and subjects. The population of students tested is different for each assessment. Achievement levels from different tests are not comparable because the cut scores for these tests are criterion referenced—they are based on content-specific expectations of what students should know and be able to do.

Table 3. Percent of Students in Performance Levels

	Assessment	
Performance Level	**Grade 4**	**Grade 7**
Distinguished	3%	2%
Proficient	28%	24%
Apprentice	54%	53%
Novice	15%	21%

Overview of the Standard Setting Process

This chapter provides an overview of the standard setting process used for the Kentucky Science Assessments for grades 4 and 7, and includes the following sections:

- Goals of the standard setting meeting
- KY Science Assessment performance levels
- KY Science standard setting process

Goals of the Standard Setting Meeting

Once students are administered an assessment, various groups, including students, parents, educators, administrators and policy makers, want to know how the students performed on the assessment and how to interpret that performance. By establishing levels associated with different student performance on the assessment, a frame of reference is developed for interpreting student scores. Setting the level of performance on an assessment sufficient for student performance to be classified into each level is one of the most critical steps in developing an assessment program.

For a criterion standards-based assessment, such as the Kentucky Science Assessment program, performance on the assessment is compared to a set of predefined content standards. The standards communicated within the *Kentucky Academic Standards for Science* in grades K-12 define a set of performance expectations for what students should know and be able to do and are derived from the National Research Council's *Framework for K-12 Science Education*, also known as the *Next Generation Science Standards (NGSS)*. The cut scores established through the standard setting process represent the level of competence students are expected to demonstrate on the assessment to be classified into each performance level.

Kentucky Science Assessment Performance Levels

Federal statute requires that any statewide assessment used for accountability purposes includes at least three achievement or performance levels. The performance levels relate student achievement on the Kentucky Science Assessments directly to what students are expected to learn, based on the standards in the *Kentucky Science Assessments for Science*. Student achievement on all KY science assessments will be classified into four performance levels that delineate the knowledge, skills, and abilities for which students are able to demonstrate mastery.

The policy-level performance level descriptors (PLDs) provide general expectations for student performance on the KY science assessments to be classified into each performance level. These do not differentiate student performance between grade levels. The policy-level PLDs for the KY science assessments were developed prior to the standard setting meeting and approved by the Kentucky Department of Education (KDE) for use during the standard setting meeting.

The four performance levels with their respective policy description are shown in Table 4.

Table 4. Policy level descriptors for the Kentucky Science Assessment

Performance Level	Policy Level Descriptors
Distinguished	A student performing at the *Distinguished* level has a **comprehensive** understanding of science concepts and practices. The student consistently communicates ideas in a sophisticated and complex manner, using thorough supporting detail and explicit examples. The student reasons and solves problems by using appropriate strategies in an insightful way. Connections between science concepts/ideas, when appropriate, are justified and insightful.
Proficient	A student performing at the *Proficient* level has a **broad** understanding of science concepts and practices. The student usually communicates ideas accurately using clear and appropriate examples, supporting or justifying those ideas with relevant details and evidence. Problem-solving and critical thinking skills are used effectively. Connections between science concepts/ideas, when present are reasonable and appropriate.
Apprentice	A student performing at the *Apprentice* level has a **basic** understanding of science concepts and practices. The student communicates ideas in a basic manner, but explanations, solutions or justifications may be unclear or ineffective. The student demonstrates some problem-solving and critical thinking skills, but they are not consistently applied.
Novice	A student performing at the *Novice* level as a **minimal** understanding of science concepts and practices. The student communicates ideas ineffectively or inaccurately, providing little detail and little or no support. Attempts at problem solving or critical thinking are minimal or inappropriate.

Kentucky Science Assessment Standard Setting Process

The recommendations by the standard setting committees represent the level of competence students are expected to demonstrate to be classified into each of the performance levels. To establish the performance levels for each assessment, the Extended Modified (Yes/No) Angoff Method (Davis & Moyer, 2015; Plake, et al., 2005) was used to guide participants as they determined their performance level cut score recommendations. This standard setting procedure is a systematic method for combining various considerations into the process for recommending cut scores for the different performance levels, including content standards and educator judgments about what students should know, based on the *Kentucky Academic Standards for Science,* and be able to demonstrate at each performance level.

The following steps were used for the KY Science Assessment standard setting process.

- Pre-meeting development – In anticipation of the standard setting meetings, various tasks were completed, the development of materials for the participants, preparation of the Pearson Standard Setting website for participants and facilitators, presentation materials for the facilitators, and development of data analysis sources and procedures.
- Standard setting meetings – Committees of participants worked with Science assessment content and referenced borderline PLDs to make recommendations for cut scores that define the different performance levels for each assessment.
- Vertical articulation meeting – The recommended cut scores for each assessment were reviewed for reasonableness and alignment of performance level expectations across grades by the members of the standard setting committees.
- Development of grade-specific PLDs - The members of the standard setting committees used an anchored item approach to develop PLDs for each grade level.
- Reasonableness review - Meetings were held by KDE to review the reasonableness of the recommended cut scores based on additional external data.

The following chapters will describe the specific procedures and activities that occurred during each of these steps.

Preparations for the Standard Setting

This chapter provides an overview of the work that was completed prior to the standard setting meetings for the Kentucky Science Assessments for grades 4 and 7, and includes the following sections:

- Development of participant materials
- Development of presentation materials
- Facilitator training
- Preparation for data analysis during the meetings

Development of Participant Materials

The Kentucky science grades 4 and 7 standard setting required a large number of materials for use by the participants during the meetings. The Pearson standard setting team worked with KDE to develop the materials used during the meeting and to ensure that all materials provided to meeting participants communicated correct information. The following materials were developed for use by participants during the meeting:

- Meeting agenda
- Participant information survey*
- Meeting non-disclosure agreement
- Test form for each grade
- Experience the test activity response form for each grade
- Test form answer key*
- Open-ended item rubrics and exemplars*
- Practice item judgment set*
- Practice item judgment set answer key*
- Practice item judgment record form
- Practice item judgment survey*
- Item judgment round record form
- Item judgment round survey* – rounds 1, 2, and 3
- Ordered item set
- Process evaluations*

Since the standard setting meetings utilized the Pearson Standard Setting website as a tool for facilitating the meeting, the website for each committee needed to be developed. Several of the documents developed, indicated with an asterisk (*), were presented online through the website. After the initial development of the websites for the meetings, a complete quality control check was performed to verify that the information provided on the websites matched the information presented on the documents.

Using approved templates, documents were created for each specific committee meeting by the Pearson standard setting team. All documents developed for the website were reviewed and approved by KDE staff before being finalized for publication for the meetings. A sample set of materials for a committee are provided in Appendix A.

Development of Presentation Materials

PowerPoint presentations were developed to guide facilitators through the presentation of information and materials throughout the standard setting meetings. The Pearson standard setting team developed the initial PowerPoint presentations. Staff from KDE had the opportunity to review and provide suggested edits to the presentations, which were resolved by the Pearson standard setting team. The following PowerPoint presentations were created for the standard setting meetings.

- General Session Presentation and Standard Setting Overview
- Standard Setting Breakout Meeting – Day 1
- Standard Setting Breakout Meeting – Day 2
- Standard Setting Breakout Meeting – Day 3

The PowerPoint presentations for the breakout meetings, Day 1, Day 2, and Day 3, were customized to reflect the specific information for the grades for each committee. Additionally, a script was added to the notes section within each presentation to guide facilitators through the presentations.

Facilitator Training

Procedures employed in the standard setting meeting are specific to the goals and objectives of the project. So, even though the facilitators for the Kentucky science standard setting meeting had prior experience in facilitating standard setting meetings, training sessions were held to discuss the unique aspects of the Kentucky science standard setting and to walk through the process utilized for this meeting, demonstrate the use of the Pearson Standard Setting website, and discuss the PowerPoint presentations to be used during the standard setting meetings. Additionally, during the onsite standard setting meetings there were regular meetings with the facilitators to review the activities during the meeting, address any issues that may have occurred, and preview upcoming activities.

Preparation for Data Analysis During the Meetings

Creation and testing of analysis programs and the calculation of impact data lookup tables were conducted prior to the standard setting meeting. To facilitate the independent analysis for each judgment round during the meeting, each analyst independently completed the programming necessary to conduct all analysis using the SAS statistical software. A trial analysis was run with mock-data to ensure that each independent analysis generated the same results.

Impact data is the percent of students that fall within an achievement level based on the recommended cut scores at the given judgment round for a particular grade. The impact data is provided to participants during the standard setting meeting to present the expected results of their recommendations on student achievement level classifications. The analysis programs use impact data lookup tables to produce this output during the meetings, which need to be created prior to the standard setting meetings.

The impact data lookup tables were created using the data from students taking both operational forms of each grade assessment during the spring 2018 administration. During the spring 2018 administration of the Kentucky science assessments, each grade had two operational forms that were administered to students. Each operational form consisted of a

set of common items which were on both forms and a set of items which were unique to each form.

The impact data lookup tables were created to represent the expected impact from both operational forms using panelists' recommendations based on only one form. For each form, a unique raw score to ability value conversion table was created using student responses from the spring 2018 administration. These conversion tables were used to assign each student administered the test an ability value, so all student scores were on the same scale. Using the raw score to ability value table for the selected form, the impact data lookup tables were created to represent the percentage of students that had ability values equal to or greater than the ability value associated with each possible raw score value for the test.

Standard Setting Meeting

This chapter provides details about the cut score setting process used for the Kentucky Science Grades 4 and 7 standard setting meetings. The sections of this chapter include:

- Purpose of standard setting meetings
- Committee participant composition
- Standard setting meeting facilitators and staff
- Standard setting pre-meeting activities
- Standard setting meeting proceedings
- Recommended performance level cut scores

Purpose of the Standard Setting Meetings

Standard setting is based, to a large degree, on the judgment of educators. Committees of educators make expert recommendations about the level of achievement expected for each performance level based on their experience with different groups of students and knowledge of the assessed content. A specific process, or standard setting method, is used to capture the educator judgments and to translate these into cut scores for the performance levels. The purpose of the standard setting meetings for the Kentucky Science grades 4 and 7 assessments was to gather expert recommendations from groups of educators from across Kentucky for the cut scores that define the different performance levels on each assessment.

Student performance on each of the Kentucky Science assessments is classified into one of four performance levels. Each committee was asked to recommend three cut scores that would define the boundaries between the performance levels. These recommended cut scores represent the performance on each assessment that a student would need to meet or exceed to be classified into the specific performance level.

Committee Participant Composition

KDE started the process of selecting participants for the standard setting meeting by requesting volunteers for the meeting from across Kentucky. All participants for the standard setting committees were selected by KDE from the individuals which volunteered and then invited to participate in the standard setting meetings. The process of selecting committee participants included selecting a sample of participants that would be as representative of the state as possible, including demographic variables (gender, race, etc.), geographic representation, and background (educational experience, education, etc.). When selecting participants, KDE placed an emphasis on those educators who had relevant content knowledge as well as experience with a variety of student groups.

There was a total of 30 participants at the standard setting meetings, who were divided between two breakout committees. Each committee focused on providing cut score recommendations for one grade, either grade 4 or grade 7. The participants were assigned to the committee prior to the meetings based on their teaching experience. The tables in Appendix B summarize the characteristics and experience of the participants in each committee. These tables provide demographic information about the committee participants as well as information about the participant's current positions in education, their experience

working with various types of student populations, and the types of districts they represent. Participant's responses to the gender and ethnicity questions was voluntary.

The participants in each committee were assigned to table groups. The table groups were selected prior to the meeting to ensure that, to the greatest extent possible, the participants at each table were representative of the committee. The participants were placed into table groups to facilitate discussions during the standard setting meetings and ensure that each participant had the opportunity to fully engage in the process.

Standard Setting Meeting Facilitators and Staff

Staff members from Pearson and KDE collaborated to conduct the Kentucky Science standard setting meeting. These staff members worked in facilitative and observational roles and did not contribute to the cut score recommendations during the meeting.

Meeting Facilitators

The lead facilitator of the standard setting meeting was Eric L. Moyer, Ph.D., from Pearson. For each of the two breakout committees at the standard setting meeting there was an assigned process facilitator. The process facilitator was a member of the Pearson psychometric staff with experience in facilitating standard setting meetings and was responsible for leading the participants through the standard setting process. The process facilitator for the grade 4 science meeting was Mark Robeck, Ph.D. The process facilitator for the grade 7 science meeting was Eric L. Moyer, Ph.D.

Meeting Data Analysts

For the standard setting meeting, two data analysts performed all of the analysis for both committees. The data analysts were Brian Wrobel and Andrew Owens, members of the research assistant staff at Pearson. During the meeting, the analysts collected participant judgment data from the Pearson Standard Setting Website, performed independent analysis to verify analysis results, and prepared participants feedback.

KDE Staff

KDE staff members attended the standard setting meeting to observe the process, answer assessment and curriculum questions, and address policy questions. KDE staff also monitored the cut score recommendations for each performance level throughout the standard setting meetings.

Materials

The following section describes the materials used by the committee members during the standard setting breakout sessions. Separate materials were developed for the anchored PLD meeting.

Pearson Standard Setting Website

The Pearson Standard Setting website was used as the online platform for housing the materials for the standard setting meeting and collecting participant judgments throughout the standard setting process. The website is based on Moodle, an open source e-learning platform that provides access to the necessary information and tools for completing the standard setting meeting. The website allowed participants access to online documents that provided background information about the KY science assessments in preparation for the standard setting meeting. The preparation materials on the website included:

- Standard setting orientation video
- Kentucky academic standards
- Kentucky science policy-level performance level descriptors
- Next Generation Science Standards - Appendix F and Appendix G
- Kentucky standard setting non-disclosure agreement

The website also allowed the participants access to materials and tools necessary for completing the standard setting activities. The standard setting materials and tools on the website included:

- Test item map and answer key
- Borderline PLD worksheet
- Practice judgment activity items
- Practice judgment readiness quiz
- Practice judgment survey
- Judgment items for rounds 1, 2, and 3
- Judgment readiness quiz for rounds 1, 2 and 3
- Judgment survey for rounds 1, 2, and 3
- Judgment feedback folders for rounds 1 and 2
- PLD development ordered item map
- PLD development ordered items
- PLD development worksheet
- Process evaluations 1 and 2

Each standard setting meeting had a unique site within the Pearson Standard Setting Moodle site. The site was formatted by individual activities, to facilitate the order of the standard setting process. During the meeting, the facilitator allowed access to the individual activity sections that the participants needed at that point in the meeting. At the end of each day, the website was closed to prevent participants from accessing the website outside of the meeting.

Panelist Folder

In addition to the online resources provided through the website, participants were provided with a meeting folder to organize a variety of hard copy materials they would need to work with throughout the meeting. These materials included:

- Agenda
- Experience the test response form
- Item judgment record forms

The folders were prepared in advance. Participants were required to check out and check in their folders at the start and end of each day of their meetings. Participants were provided additional materials throughout the meeting, which they were instructed to insert into their folder.

Computers

Each participant was provided a laptop computer in his or her meeting room to access the online resources through the website. The laptops were Dell latitudes with 15.6" screens, standard keyboards with full-size number pad, and an external mouse. Participants were not provided with external keyboards, numeric keypads, or external monitors. Participants were seated in table groups in pod configuration with 3 to 4 participants each, to provide each participant with enough space to work with the computer and folder materials. The power supplies were centrally located in the middle of each table. The participants used Google Chrome to access the standard setting website, which was programmed with a white list of websites to restrict participants use of the computers to work associated with the cut score setting meeting.

Standard Setting Procedure

The Extended Modified (Yes/No) Angoff Method (Davis & Moyer, 2015; Plake, Ferdous, Impara, & Budkendahl, 2005) was used during the standard setting meeting to assist participants in recommending performance level cut scores for each assessment. The design of the KY science assessment, with independent items being clustered into item sets with an associated stimuli, led to a modification of this method. For round 1, participants were asked to review each independent item from the operational administration and answer the following question:

"How many points would a borderline student at the [specific] performance level likely earn if he or she answered the question?"

The judgment participants provided for each item was in terms of whole number point options. For round 2, in addition to the item level judgment, the participants were asked to review each cluster of items associated with the same stimuli and answer the following question:

"How many points would a borderline student at the [specific] performance level likely earn if he or she answered all the questions associated with the cluster?"

For round 3, the participants only provided their cluster level judgment for each cluster on the assessment. For the standard setting meeting, "likely" was defined statistically as the student having at least a 2/3 chance of earning the number of points. The participants completed the task for each performance level.

Participants completed three rounds of item judgments. Between the judgment rounds they were provided feedback information including data relative to participant agreement, student performance on the items, and student impact data.

Standard Setting Meeting Proceedings

The standard setting meetings were conducted across three days, July 17 - 19, 2017, in Lexington, Kentucky. Appendix C includes the complete agenda for the standard setting meetings. Table 12 presents a high-level agenda for the standard setting committee meetings.

Table 5. Standard Setting Agenda Topics

General Session	• Welcome and Introductions • Overview of the KY Science Assessment System • Overview of Cut Score Setting Process
Breakout Sessions	• Introductions • Experience the Assessment • Borderline Performance Level Descriptors • Standard Setting Training • Round 1: Judgment and Feedback • Round 2: Judgment and Feedback • Round 3: Judgment and Feedback • Vertical Articulation • Anchored Performance Level Descriptor Development • Evaluation and Closing Remarks

The following will describe the steps used to guide the participants through the entire standard setting process.

Standard Setting Meeting Pre-Work

The standard setting meeting participants were allowed access to a set of activities prior to attending the onsite meetings. The purpose of the pre-work was to expedite the training of the participants by providing the participants an opportunity to familiarize themselves with information that would be used throughout meeting. The pre-work included:

- Pearson Standard Setting Website – The pre-work was provided via documentation or links embedded within the secure Pearson Standard Setting Website developed for the standard setting meeting. The participants were provided their unique login and temporary password through an email sent to the email address they provided during registration. The participants were asked to login to the website to complete the pre-work activities, which also gave the participants experience in accessing the website and navigating through the pre-work sections and activities.
- Participant information survey – Participants were provided a survey to document their demographic information as well as current teaching position, experience, and school information. Participants were able to access this survey before and during the meetings.
- Standard setting orientation video – Participants were provided access to a short video which introduced them to the purpose and concepts associated with the KY science standard setting meetings.

- Borderline PLD Development – Participants were asked to develop draft borderline performance level descriptors (PLD) for a specific set of performance expectations. To assist with this task, participants were provided with an instructional video which introduced them to the concepts of performance level descriptors and borderline expectations and the steps in completing the activity. Borderline PLD worksheets, found on the website for each participant, identified the specific performance expectations they were expected to develop associated borderline expectations for.
- Security and Non-disclosure – Participants were provided access to the security and non-disclosure agreement for the standard setting meeting so they would be familiar with its content before signing the agreement at the meeting.

Participants were provided with their standard setting website login information via email. This login provided them access to the specific section of the website associated with the standard setting meeting for which they were registered. Participant access was restricted to only the respective site for the standard setting meeting they were attending.

General Session

The purpose of the general session was to welcome the participants, provide background information about the Kentucky science assessment system, and introduce the standard setting process. A single general session including all standard setting participants was conducted at the beginning of the standard setting meeting. An overview of the cut score setting process was provided by Dr. Eric Moyer, the lead research scientist from Pearson facilitating the standard setting process. The presentation slides for the general session are included in Appendix D.

Breakout Session

After the general session, participants moved into grade-specific breakout sessions for the remainder of the standard setting meeting. Each committee was responsible for providing recommendations for cut scores for each of the performance levels for the grade-specific science test. The following activities were used to guide the participants through the standard setting process. The presentation slides used during the breakout sessions are included in Appendix D.

Experience the Test. Participants experienced one of the two operational test forms that the students were administered during the spring 2018 administration. The participants experienced the test just as students did. Since the students were administered the assessment on paper test forms, the participants were presented paper test books from the spring 2018 administration. This activity provided participants the opportunity to experience the items in a similar method as students and to develop initial perspectives on the knowledge and skills needed to provide correct responses to the item.

Participants recorded their responses to the items on a separate item response form, provided in the participant folder. After the participants completed the ‘Experience the Test’ activity, the participants were given the opportunity to review the correct responses for each item. An item map was provided on the website for participants to review their item responses. The item map provided correct responses for multiple choice items and scoring rubrics and student exemplars for open-response items.

Borderline Performance Level Descriptors. An essential component to the standard setting process is the development of borderline performance level descriptors. As part of the pre-work the participants developed draft borderline PLDs for select performance expectations. Prior to the breakout meetings, the meeting facilitators, collected the participant draft borderline PLDs from the pre-work into interactive worksheets the participants would use during the breakout sessions.

To help inform the borderline PLD development activity, the facilitators reviewed the performance levels and the policy-level performance level descriptors for Kentucky science assessments with the panelists. This provided participants with a common understanding of the knowledge, skills, and abilities typical students should demonstrate within each performance level. During this group activity, participants were asked to discuss the differences between the expectations at the different performance levels.

The participants were then introduced to the concept of the difference between a *typical* student and *borderline* student within a performance level. The borderline student was described as the minimally qualified student to be classified within a particular performance level, possessing just enough knowledge, skills, and abilities to achieve the specific performance level classification. The table groups reviewed the draft borderline PLDs from the participant pre-work, included on borderline PLD worksheets, accessed through the website. The participants revised the draft borderline PLDs assigned to their table group to define the expectations for students at the borderline of each performance level. The borderline PLDs from each group were then collected into a master document and reviewed by the whole group. Edits to the master document were made to create a common set of borderline PLDs that would be used by the participants throughout the meeting. The final list of borderline PLDs were printed and provided to each participant to place in his or her folder as a reference for subsequent activities.

Item Judgment Process Training. The process facilitator for each committee provided the participants with training on the Extended Modified (Yes/No) Angoff standard setting procedure (Davis & Moyer, 2015; Plake, et al., 2005) and how to use the website to record their individual item judgments. They were instructed to review each item from the assessment, which was presented in the test book and on the website as a PDF, review the borderline PLDs, the answer key, and, if needed, the rubric and student exemplars for the item. Based on their review of the item and the related materials, the participants answered the following question for each of the three performance levels:

“How many points would a borderline student at the [specific] performance level likely earn if he or she answered the question?”

The response to judgment question for each item was recorded in the judgment survey in the website. Figure 1 presents an example item judgment survey in the website. Participants completed the item judgments for each performance level for an item before moving on to the next item.

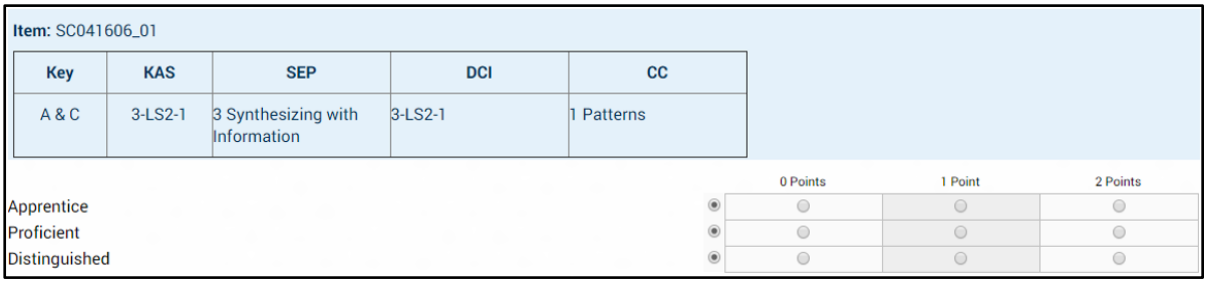


Figure 1: Example Item Judgment Survey from Website

The participants also kept a record of their item judgments on their paper Judgment Round Record Sheet. This document was provided to them as part of the materials in their folder. It included the unique item number, Kentucky Academic Standard (KAS), and maximum possible points for the item. The participants were shown how to use the unique item number to ensure that they were referencing the correct item on all documents within the judgment survey and within the test online.

To provide the participants practice in making item judgments, they completed a practice judgment task. The participants made judgment for all performance levels on a set of practice items, including both dichotomously and polytomously scored items. They were expected to complete their judgments independently and without discussion from other participants. After all the participants completed the practice judgment activity, a group discussion was used to review the judgment process, review the participant responses, demonstrate how their item judgments were used to determine a test level recommendation, and answer any questions that they had about the judgment process.

Item Judgment Rounds. After receiving training on the standard setting process, the participants worked through three rounds of judgments. Before starting each of the three judgment rounds, the facilitator completed a review of the item judgment process, including explicit instructions on which materials would be needed to complete the judgment task. Participants were required to complete a readiness survey in the website indicating that they understood the task and process used to complete the item judgments. The participants had to answer "yes" to all readiness survey questions before continuing with the judgment round. If they responded "no" to any question, they were asked to notify a facilitator for additional assistance. Figure 2 presents an example of the readiness quiz participants needed to complete before starting the item judgment task.

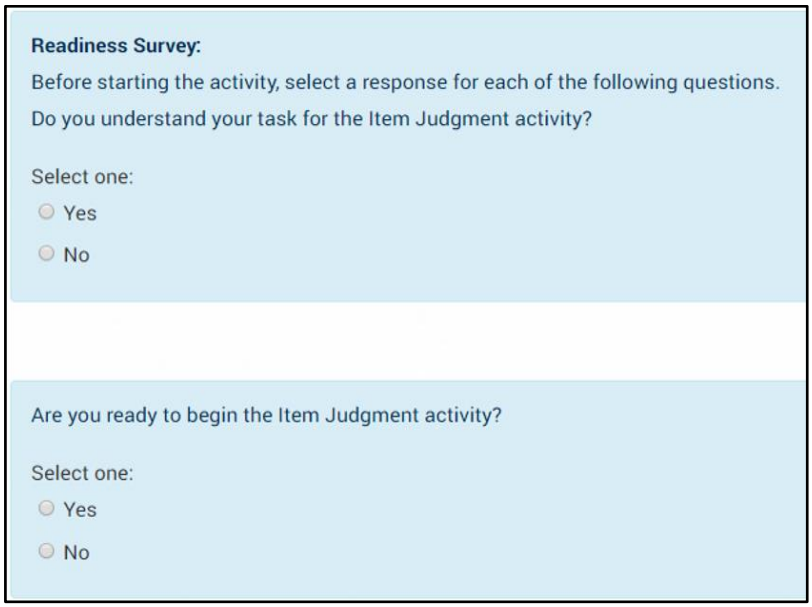


Figure 2: Example Readiness Quiz Before Item Judgment Task

Once the participants had completed the readiness survey, they were provided access to the item judgment survey for the respective round. Participants independently completed the item judgment task. The item judgment survey required that participants provide judgments for each item prior to submitting the judgment survey.

Once all the participants had completed their item judgments, data analysts from Pearson collected the data from the website and performed the analysis to determine an aggregate recommendation for the committee. The participants were provided feedback after each

judgment round which could be used to inform subsequent judgments. Table 6 displays the type of feedback that was provided to participants after each round of judgments.

Table 6. Feedback Data Provided to Participants After Each Judgment Round

Feedback	Round		
	1	2	3
Individual item-level judgment record	Yes	Yes	Yes
Individual test-level recommendations	Yes	Yes	No
Table test-level recommendations	Yes	Yes	No
Committee test-level recommendations	Yes	Yes	Yes
Item-level participant agreement	Yes	Yes	No
Test-level participant agreement	Yes	Yes	No
Item score mean and score distribution	Yes	Yes	No
Impact data	No	Yes	Yes

Appendix E provides examples of each of the feedback data provided to participants, along with a brief description of the feedback presented.

After feedback from round 1 judgments were provided to participants, they participated in table-level discussion of the rationale for each of their round 1 item judgments, facilitated by the table leaders. After feedback from the round 2 judgments were provided to participants, both table-level and committee discussions were facilitated where participants could discuss feedback data and rationale for individual round 2 judgments. Since the round 3 judgments were the participants' final judgments, the feedback data was provided to facilitate the participants' evaluation of the final recommendation by the committee and to discuss any additional changes they felt were necessary. This step was completed to assist the discussion during the vertical articulation meeting.

Process Evaluations. After the round 3 judgments and feedback, participants were asked to complete a process evaluation survey in the website. The purpose of these surveys was to collect information about each participants' experience in recommending cut scores for the performance levels associated with the assessments. The survey asked participants to provide feedback on the following:

- The level of success of the various components of the meeting
- The usefulness of the activities conducted during the meeting
- The adequacy of the various components of the meeting
- The adequacy of opportunities to ask questions, etc., at the meeting
- How confident participants were that the recommended cut scores accurately reflected student performance at each performance level
- Whether committee members thought that their judgments and opinions were treated with respect by facilitators and fellow participants

All participants were also allowed to provide any additional information concerning their evaluation of the process of the standard setting meeting through an open response question.

Recommended Cut Scores from Standard Setting Committees

During the standard setting meeting it was expected that there would be variation between participants' cut score recommendations for each performance level. To determine a single cut score recommendation for a performance level from a committee, each participant's cut score recommendation for that performance level was averaged across participants. Specifically, the median cut score from a set of participants' cut score recommendations was used to determine the recommended cut score for a performance level for the committee. The recommendations resulting from the round 3 judgments were considered as the committee's recommendation for the standard setting meeting. Table 7 displays the recommended cut scores for each performance level based on the round 3 recommendations for each grade. Figure 3 displays the impact data based on the recommended cuts scores from round 3 from each committee.

Table 7. Cut Score Recommendations from Standard Setting Committees

Grade	Maximum Score	Apprentice		Proficient		Distinguished	
		Raw Score	% Correct	Raw Score	% Correct	Raw Score	% Correct
4	48	15	31.2%	27	56.3%	36	75.0%
7	48	13	27.1%	26	54.2%	38	79.2%

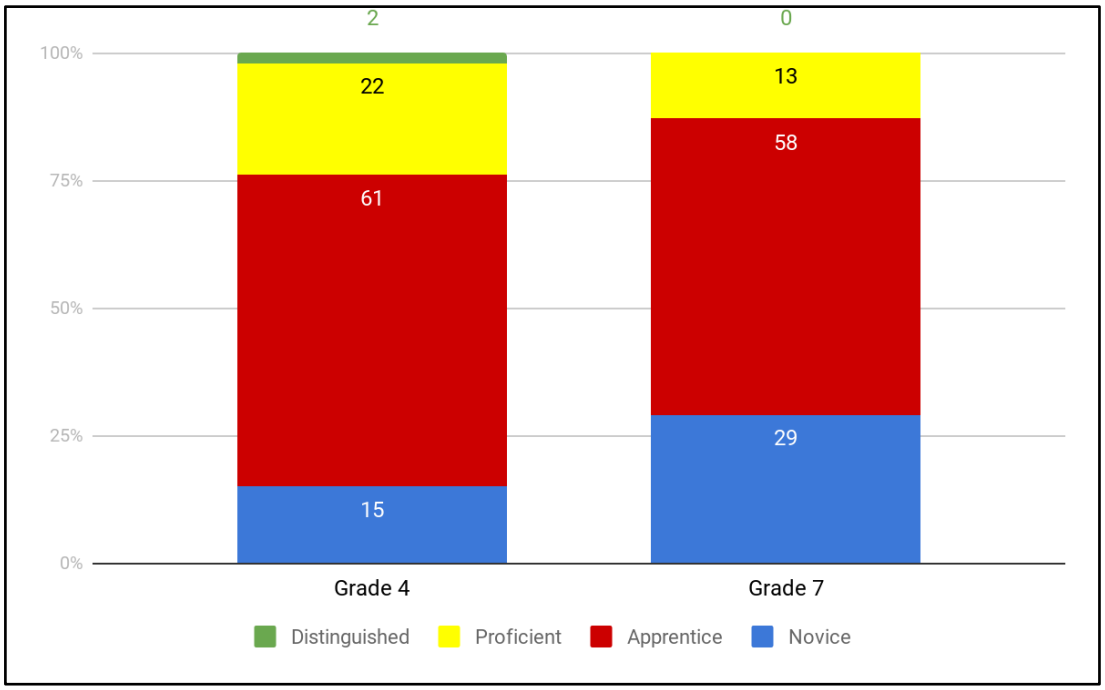


Figure 3: Impact Data from Round 3 Recommendations

The recommended cut scores for each achievement level from the three judgment rounds for each standard setting committee, represented as raw scores, are presented in Appendix F. The summary statistics for the recommended cut scores for each performance level from the three judgment rounds for each standard setting committee are shown in Appendix G. The participant agreement data for each performance level for judgment rounds 1 and 2 for each standard setting meetings are shown in Appendix H. The estimated impact data after

judgment round 3 for each performance level for each standard setting committee are shown in Appendix I.

Vertical Articulation

The purpose of the vertical articulation meeting was to review the cut score recommendations from the two standard setting committees and evaluate the reasonableness of the recommendation. Where the recommendations from the standard setting committees were made with a specific focus on the respective content for this committee, the focus of the vertical articulation committee was to view the cut score recommendations across grades to evaluate whether the recommendation resulted in a cohesive assessment system. The standard setting participants were guided through a specific process where they would review the recommendations from the standard setting committee and, if necessary, recommend and review changes to the recommendation. The result was a set of recommended cut scores from the vertical articulation committee.

The participants from both breakout sessions came together on the morning of Thursday, July 19, 2018 to participate in the vertical articulation meeting. The facilitator for the vertical articulation meeting was Eric L. Moyer, Ph.D., the lead facilitator of the standard setting meeting. The participants remained in the same table groups that they were assigned in the breakout sessions.

Meeting Process

The process for the vertical articulation meeting involved two steps:

- Review and discussion of the cross-grade impact data
- Discuss adjustments to recommended cuts scores

At the beginning of the vertical articulation meeting, the participants were instructed to the purpose of the vertical articulation which was the opportunity to review the recommended cut scores from the standard setting meetings across the grades, ensuring that they represented a cohesive assessment system. In the previous standard setting breakout sessions, they were focused primarily on the content related to the grades within their committees, where in this meeting they would review the recommendation from all the standard setting committees from a policy perspective.

The participants were presented with the cross-grade impact data chart reflecting the results from the round 3 judgments of all standard setting committees. The impact they were presented is shown above in Figure 3. The groups had the opportunity to discuss how the results looked across grades, based on their initial expectations.

Based on their initial expectations of student impact, the participants were provided the opportunity to investigate changes to the recommended cut scores from round 3 using an interactive spreadsheet, which was accessed through the Pearson website. Figure 4 presents the interactive spreadsheet for the vertical articulation meeting.

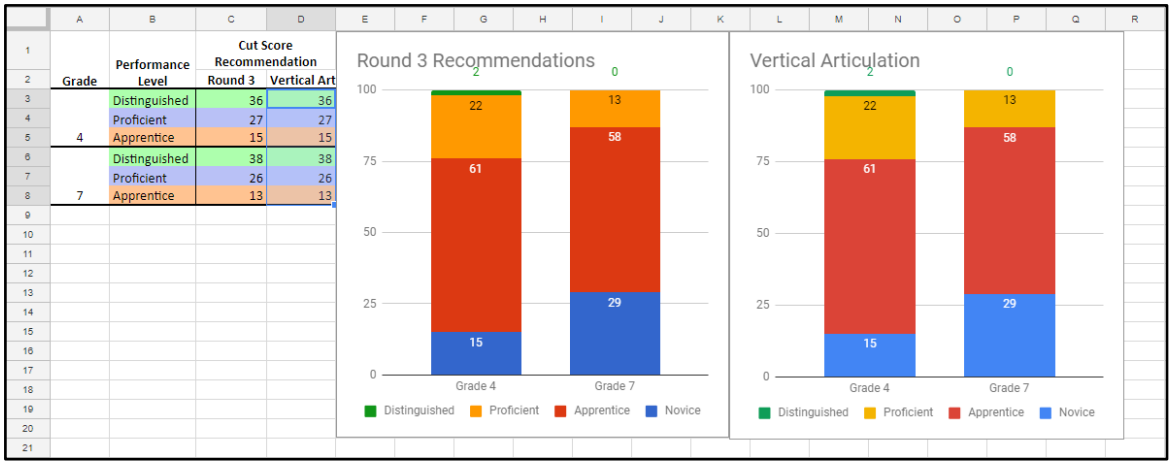


Figure 4: Interactive Spreadsheet for Vertical Articulation Meeting

The interactive spreadsheet allowed participants to investigate possible changes to the cut scores by adjusting the current cut scores and simultaneously viewing the change to the impact data. The participants had the opportunity to recommend changes to cut scores for the performance levels for the grade associated with their breakout group. When a change in cut score was recommended by a group, it was entered into the interactive spreadsheet by the meeting facilitator for the entire committee to view the change in cut scores and pattern of impact data across grades and performance levels. One recommended change at a time was viewed, discussed, and then either accepted or rejected by the committee. This process was repeated until all recommended changes were discussed and the vertical articulation committee agreed with the entire set of cut score recommendations across the grades.

The participants discussed the differences in impact data across the grades and their impressions of the recommendations. Due to the expected difference in difficulty of the assessments and the implementation of the new Kentucky Academic Standards for science, the participants did not discuss any need for adjustments to the cut score recommendations.

Performance Level Descriptor Development

Performance level descriptors (PLDs) are statements that articulate the knowledge, skills, and abilities that students classified into a particular performance level should be able to demonstrate. All Kentucky science assessments have four performance levels, as defined above in Table 4. The performance levels range from Novice, representing the lowest level of student performance, to Distinguished, representing the highest level of student performance.

The PLDs are associated with the performance levels in the following way:

- *Performance levels* indicate a student's level of competency of the standards defined in the Kentucky Academic Standards for science through classification of their performance on an assessment for the specific grade as *Novice, Apprentice, Proficient,* or *Distinguished.*
- *Performance level descriptors* indicated the knowledge, skills, and abilities expected of students to demonstrate competency at each grade level to be classified in each performance level.
- *Cut scores* partition the test scale and represent the minimum test score that a student must earn on an assessment for each grade level to be classified into a given performance level.

A well-defined set of PLDS is useful in delineating the knowledge, skills, and abilities which are associated with each performance level.

The development of the PLDs for the Kentucky science assessments was done after the vertical articulation meeting, once the participant cut score recommendations for each performance level were finalized. The participants of the standard setting meeting worked in their breakout groups to develop PLDs for the associated grade for which they recommended cut scores. The following describes the process which the participants were led through to develop the PLDs for the respective grade.

Meeting Process

Since the PLDs development was conducted after the standard setting meeting, an anchored development process was used to facilitate the development of the PLDs. The result of the standard setting meeting was cut score recommendations for each performance level for the assessment. The anchored development process uses the ability level associated with the cut score recommendations to align the items for the assessment with each performance level. The knowledge, skills, and abilities needed to respond to each set of items are then used to define the expectations communicated by the PLDs for each performance level, Apprentice, Proficient, and Distinguished. No PLDs were developed for the Novice performance level.

The set of ordered items was essential to the implementation of the anchored PLD development process. The ordered item set presented the test items from the spring 2018 administration of the Kentucky science assessment for the respective grade from the easiest item to the most difficult item. The order of item difficulty was based on the Rasch item parameters for each item, determined using the student responses to the items from the spring 2018 administration. Each multiple choice item was represented one time in the item set. Items which are polytomously scored, with maximum score greater than one, including short answer and extended response items, were represented in the item set one time for each non-zero score point. Polytomous machine-scored items (i.e., multi-select items) were not represented in the item set. An item map was developed to communicate to participants the order of the items in the item set, along with item information, such as item keys and standards.

The items associated with each performance level were determined using the cut score recommendations for each performance level from the vertical articulation meeting. The ability level associated with each cut score recommendation was determined using the raw score to ability level conversion table. For each item in the ordered item set, the ability value associated with a 67% probability of providing a correct responses was determined, also known as an RP67 value. The item with an RP67 value closest to but greater than or equal to the cut score ability level was the first item in the set of items associated with the performance level. The item set for a performance level included all the items from the first item associated with the performance level through the item before the first item for the next performance level. Table 8 displays the item range associated with each performance level for each grade.

Table 8. Item Ranges Associated with each Performance Level

Grade	Novice		Apprentice		Proficient		Distinguished	
	Min	Max	Min	Max	Min	Max	Min	Max
4	1	4	5	26	27	53	54	77
7	1	1	2	30	21	60	61	77

The facilitator introduced the ordered item set to the panelists and discussed how it was constructed and the relationship between the items in the set and the performance levels. The participants were provided with the item sequences that separated the item sets associated with each performance level. The participants worked in their table groups to review the items associated with each performance level that their group was assigned. From their review of the items, the participants defined a set of knowledge, skills, and abilities that represented a reasonable set of expectations for students classified into the performance level.

After the table groups created their draft PLDs, the expectations were collected into a master document. The master document was shared with the whole group for final review. The participants were able to make recommendations for editing the PLDs for each performance level. The facilitator led the discussion, making edits to the master PLD document, as recommended by participants. This final review was to ensure consistency in student expectations across the performance levels within the grade.

The final PLDs from the grade 4 and grade 7 committee were reviewed by KDE and revisions were made to ensure comparability in the PLDs across grades. The final PLDs from this review are in Appendix J.

Post-Standard Setting

This chapter provides details about the process used after the standard setting to approve the performance level standards for the Kentucky Science Grades 4 and 7 assessments. The sections of this chapter include:

- Reasonableness review
- Follow-up review meeting
- Final approved standards

Reasonableness Review

Following the standard setting meeting, an executive summary was provided to KDE to facilitate a review of the recommendations from the standard setting meeting. The executive summary included a brief overview of the methodology and process used to obtain the cut score recommendations, the participants' cut score recommendations for each performance level, and the impact data associated with the recommended cut scores. This summary was provided to KDE on Friday, July 20, 2018.

Using the executive summary, KDE reviewed the reasonableness of the cut score recommendations for the Kentucky science assessments from the standard setting meeting. The purpose of this review is to evaluate the reasonableness and alignment of the recommendations with other data, expectations for alignment across grades, and usefulness in the communication of results within the context of the state accountability system. Members from KDE along with technical advisors for the science assessment program participated in the reasonableness review discussion.

The recommendation from this review of the proposed cut scores from the standard setting meeting was to convene a follow-up review meeting to review the recommended standards while also considering additional evidence. This follow up committee would consist of members from the original standard setting meeting, which would be able to recommend informed revisions to the cut score recommendations.

Follow-up Review Meeting

On Monday, July 30, 2018, a standard setting follow up review meeting was convened to consider the reasonableness of the performance level cut score recommendations from the initial standard setting committee using performance data from other academic measures, and to recommend adjustments, if there existed a compelling reason. All participants from the original standard setting meeting were invited to participate in a 1 hour and 30-minute online meeting to review the recommended performance level cut scores for each assessment in conjunction with the external data. There were 19 participants that indicated that they would join the meeting and received login information for accessing the discussion. The meeting was facilitated by Dr. Eric L. Moyer, as the lead facilitator for the standard setting meeting, with members of KDE also attending.

During the meeting, the participants were instructed that the purpose of the meeting was to review the performance level cut score recommendations for each assessment with additional information provided in the form of performance data from other assessments that

was not presented during the standard setting meeting. The discussion started with a review of the performance level cut score recommendations from the standard setting meeting with the associated impact data. The participants in the meeting discussed the initial rationale for recommending these performance level cut scores. The participants were then introduced to the use of external data to further evaluate the cut score recommendations against. The data that was presented to the participants during the meeting included:

- K-PREP Science for Grades 4 and 7
- K-PREP Math for Grades 4 and 7
- National Assessment for Educational Progress (NAEP) for Grades 4 and 8

Figure 5 displays the historical data that was presented for the K-PREP Science tests for grades 4 and 7. The data was presented as impact data from the most recent administrations of the K-PREP Science assessments for grades 4 and 7, from 2012 through 2014. The participants were able to discuss the association between the previous K-PREP Science assessments and the new Kentucky Science Assessments. The increase in rigor and expectations defined by the new Kentucky Academic Standards for Science from the previous standards was discussed by the participants.

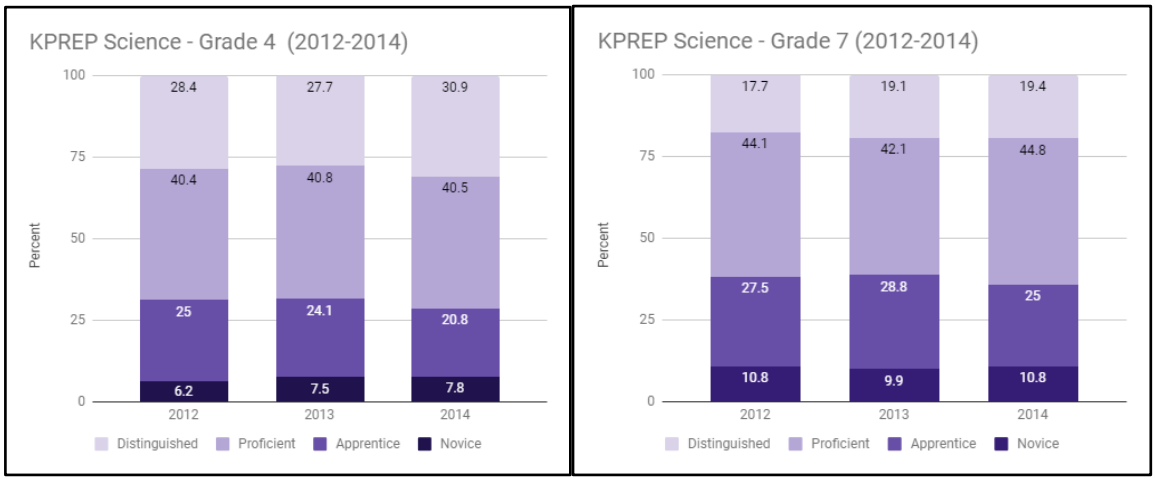


Figure 5: K-PREP Science impact data

Figure 6 displays the data that was presented for the K-PREP Math tests for grades 4 and 7. The impact data from the first two administration years was displayed along with the impact data resulting from the standard setting cut score recommendations from the most recent administration. The purpose of presenting this data was to show how the impact data changed as the test was administered over several years. The participants discussed the degree to which there is an association between math and science performance. They also discussed the difference between the impact data from the first administration year for math and the impact data from the science cut score recommendations.

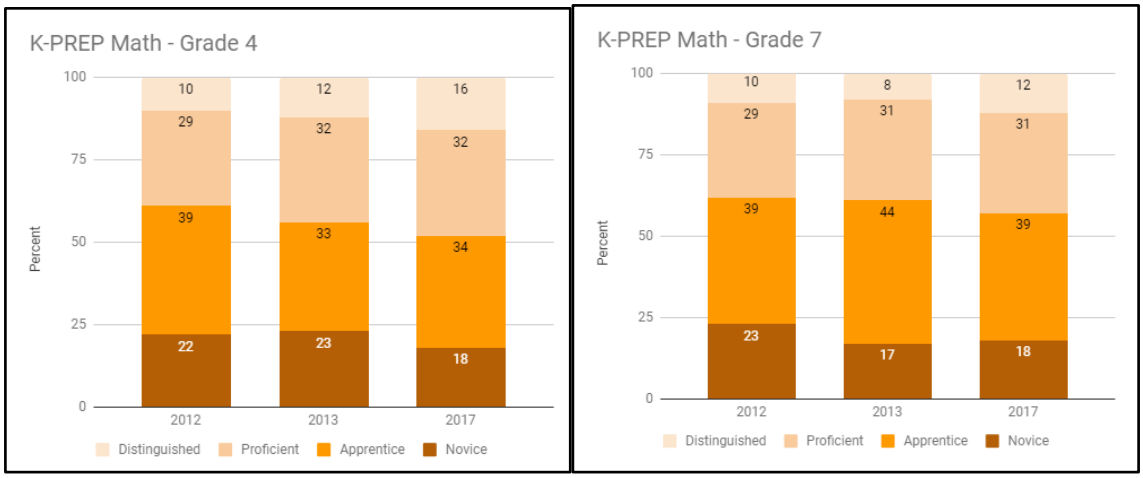


Figure 6: K-PREP Math impact data

Figure 7 displays the data that was presented for the National Assessment of Educational Progress (NAEP) for science grades 4 and 8. The most recent data that was available was from the 2015 administration. This assessment is only administered to students in grades 4 and 8. The results for both the nation and for the state of Kentucky were presented for comparison. The participants discussed the differences between the expectations for the NAEP science assessments and the Kentucky Science Assessments.

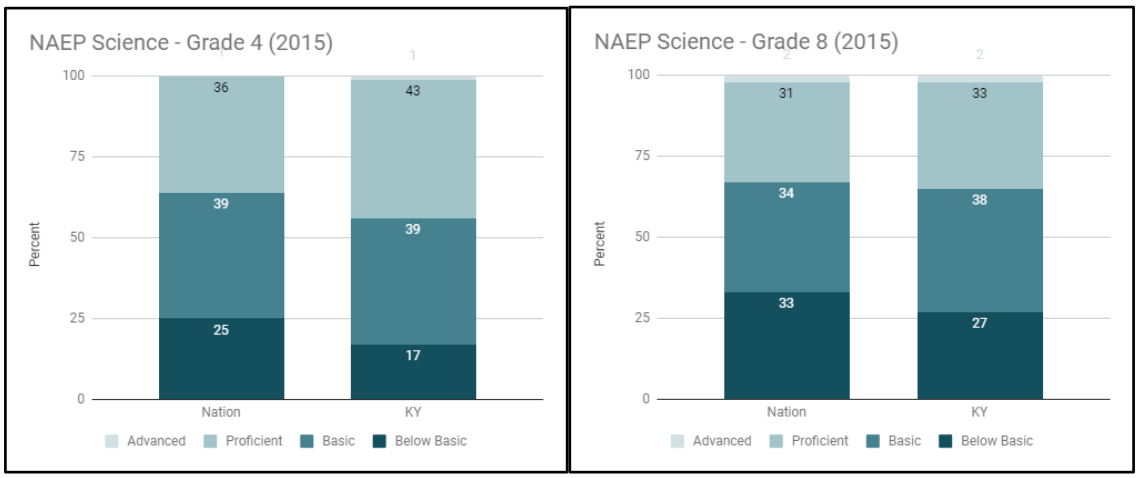


Figure 7: NAEP Science impact data for grades 4 and 8 from 2015

Following the presentation and discussion of the external assessment data, the participants were then able to discuss whether any adjustments were warranted to the original cut score recommendations. An interactive spreadsheet was used to assist in the discussion where participants were able to see the results of changes to the performance level cut scores as part of the discussion. During the discussion, participants were able to recommend changes to particular cut scores along with the rationale for the recommended adjustment. Once an adjustment was recommended by a participant, the other participants had the opportunity to discuss the rationale for the adjustment. The adjustment was kept if a majority of the participants agreed with the recommendation, otherwise the cut score was returned to the original value. This process was repeated for each recommended cut score.

The participants in the follow-up meeting were sent an online process evaluation survey to collect their information about each participants' experience in the follow-up meeting, including reviewing the additional data and recommending adjustments to the original cut scores. The survey asked participants to provide feedback on the following:

- The level of success of the various components of the meeting
- The adequacy of the various components of the meeting
- The adequacy of opportunities to ask questions, etc., at the meeting
- How confident participants were that the recommended cut scores accurately reflected student performance at each performance level

All participants were also allowed to provide any additional information concerning their evaluation of the process an open response question.

Table 9 displays the recommended performance level cut scores resulting from the follow-up meeting. For grade 4, the Distinguished cut score was decreased by 1 and the Proficient cut score was decreased by 1. For grade 7, the Distinguished cut score was decreased by 3, the Proficient cut score was decreased by 2, and the Apprentice cut score was decreased by 1. For each of these adjustments, the cut score recommendation was within the range of cut score recommendations from Round 3 of the standard setting meeting, between the minimum and maximum recommended cut scores. Figure 8 displays the impact data associated with the cut score recommendations from the follow-up meeting.

Table 9. Cut Score Recommendations from follow-up meeting

Grade	Maximum Score	Apprentice		Proficient		Distinguished	
		Raw Score	% Correct	Raw Score	% Correct	Raw Score	% Correct
4	48	15	31.2%	26	54.2%	35	72.9%
7	48	12	25.0%	24	50.0%	35	72.9%

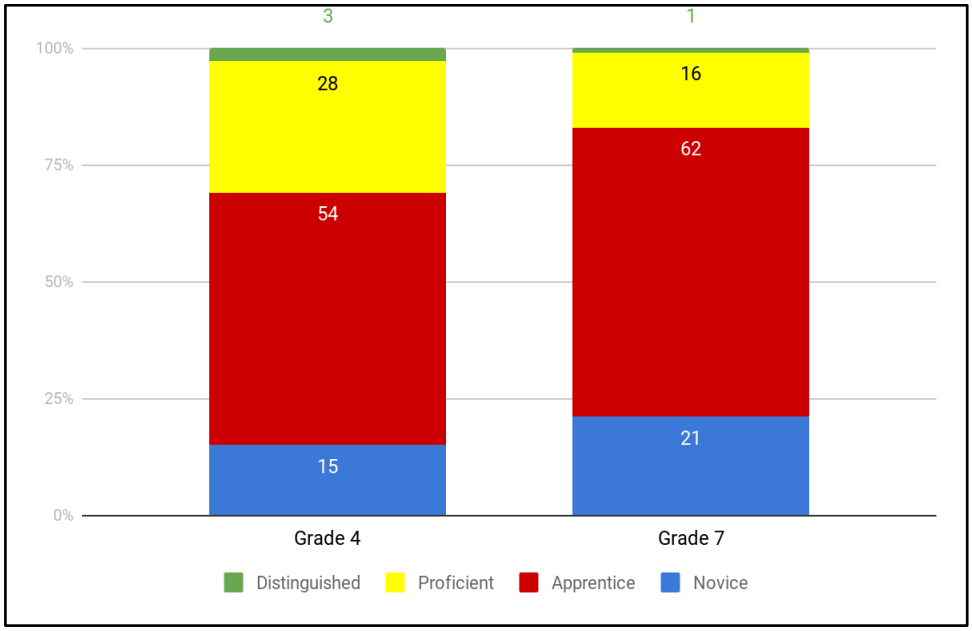


Figure 8: Impact data from cut score recommendations from follow-up meeting

Final Performance Level Standards

A final reasonableness review meeting was conducted with KDE and the Commissioner for the performance level cut score recommendations from the follow-up meeting. As with the previous reasonableness review, the cut score recommendations for each grade were reviewed for reasonableness and alignment of the recommendations with other data, expectations for alignment across grades, and usefulness in the communication of results within the context of the full accountability system. Based on this review, KDE explored additional adjustments to the cut score recommendations for the grade 7 Proficient and Distinguished performance levels. Table 10 displays the final recommended performance level cut scores resulting from the reasonableness review.

Table 10. Final Cut Score Recommendations

Grade	Maximum Score	Apprentice		Proficient		Distinguished	
		Raw Score	% Correct	Raw Score	% Correct	Raw Score	% Correct
4	48	15	31.2%	26	54.2%	35	72.9%
7	48	12	25.0%	22	45.8%	34	70.8%

The performance level cut score recommendations were initially defined in terms of raw scores on the test. Student results are not reported as raw scores, since the overall difficulty of tests may change from administration to administration, so results may not be comparable across administrations. To address this, student results on the Kentucky science assessments are reported using scale scores, which are comparable across administrations. The reporting scale for each assessment is an independent horizontal scale with a lowest obtainable scale score of 100 and a highest obtainable scale score of 300. The cut score on the reporting scale for the Proficient performance level is set at 210. The cut scores on the reporting scale for the Apprentice and Distinguished performance levels is determined using the same scaling factor from the previous K-PREP science reporting scale. The Table 11 displays the scale score ranges for each performance level for each grade. Figure 9 displays the impact data associated with the final cut score recommendations as scale scores.

Table 11. Final Cut Score Recommendations on Reporting Scale

Performance Level	Scale Score Ranges	
	Grade 4	Grade 7
Distinguished	226 to 300	229 to 300
Proficient	210 to 225	210 to 228
Apprentice	191 to 209	192 to 209
Novice	100 to 190	100 to 191

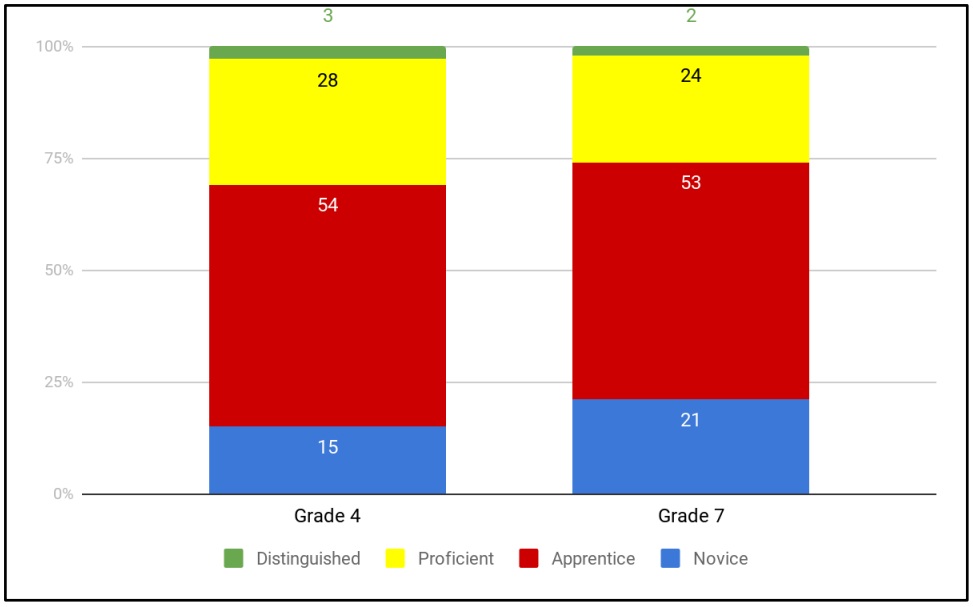


Figure 9: Impact data from final cut score recommendations

An executive summary was provided to KDE with a brief overview of the methodology and process used to obtain the final cut score recommendations, the final cut score recommendations for each performance level on the reporting scale, and the impact data associated with the final recommended cut scores. This summary was provided to KDE on August 1, 2018. The executive summary provided is at the beginning of the technical report.

Evidence of Procedural Validity of the Standard Setting Process

This chapter details various evidence for the validity of the process used during the standard setting meetings. The sections in this chapter include the following:

- Committee representation
- Committee training
- Participants' perceived validity of the meeting

Committee Representation

As part of the standard setting evaluation, participants completed a demographic survey that collected information about their background relevant to educational experience. The results of the self-reported demographic characteristics of the participants are documented in Appendix B.

As part of the survey, participants were asked to report their highest level of education (Table B.6), their current position (Table B.1), their number of years in education (Table B.2), and the number of years teaching a course related to their standard setting meeting (Table B.4). In each of the committees, the participants that had master's or doctoral degrees composed a large majority of the committee. At least 80 percent of the participants of each committee were teachers in grades K–12. The teachers in the committees had a range of teaching experience, with at least 60 percent of the teachers having greater than 10 years of experience in education.

The experience of the teachers in each committee was relevant to the recommendations they were making, with a majority of participants in each committee indicating they had experience teaching the subject in the grades relevant to their committee, as presented in Table B.3. The experience of the teachers in the committees included experience teaching different populations of students, as displayed in Table B.4. A large majority of participants of each committee had experience teaching general education, mainstream special education, and English language learners.

The participants in the committee were representative of the different regions of Kentucky, as presented in Table B.6. A large majority of participants were currently working in school districts, as presented in Table B.10. The participants that worked within school districts represented the various types of districts across the state, including size, type, and socioeconomic status. The set of participants for this standard setting was well selected for representing the teachers across the state in this process, which was noticed consistently by the facilitators of the meeting.

Committee Training

During the standard setting meeting, it was essential that participants understood how to make judgments as part of the Extended Modified (Yes/No) Angoff standard setting methodology. The training on the standard setting methodology was provided during the general session and in the breakout standard setting committees for each grade-level test.

The training on the implementation of the standard setting process was standardized across committees through the PowerPoint training slides, script, and materials used.

Participants went through a practice item judgment round as an opportunity to implement the standard setting methodology without consequence, including making judgments within the Pearson website. During the practice item judgment round, the participants reviewed a reduced set of items and provided item judgments for the three achievement levels, *Apprentice, Proficient,* and *Distinguished*. After the practice round, a whole-group discussion was facilitated by the process facilitator to identify and respond to any questions or issues participants encountered while implementing the standard setting process. Before each judgement round, participants responded to a readiness survey that asked whether participants were prepared for making their judgments. Participants were not able to continue to the item judgment survey unless they answer yes to both questions on the readiness quiz and were encouraged to ask the facilitator questions if they responded “no” to either question.

At various points within the standard setting meeting, participants completed a process evaluation survey to record their impressions of the effectiveness of the materials and methods employed through the process. Figure 9 above displays the results of the evaluation survey across grade-level committees for several questions related to the training on the standard setting process. The results of these process evaluations for each individual committee are presented in Appendix K.

As part of the evaluation survey, the participants were specifically asked about the effectiveness of the training they received on the standard setting process. One question asked participants to rate the success of the overview of the standard setting process during the standard setting training. Generally, the initial introduction to the standard setting process was perceived as successful with over 90 percent of all participants responding that it was either *Successful* or *Very Successful*. The perception of the training on the standard setting process in the breakout groups was also good, where more than 90 percent of participants responded that it was either *Adequate* or *Very Adequate*. More than 90 percent of all participants indicated that the practice judgment activity for the standard setting process was either *Successful* or *Very Successful*. These responses indicate that, overall, most participants believed that the training provided prepared them to implement the standard setting procedure, providing cut score recommendations for each assessment for which they were responsible.

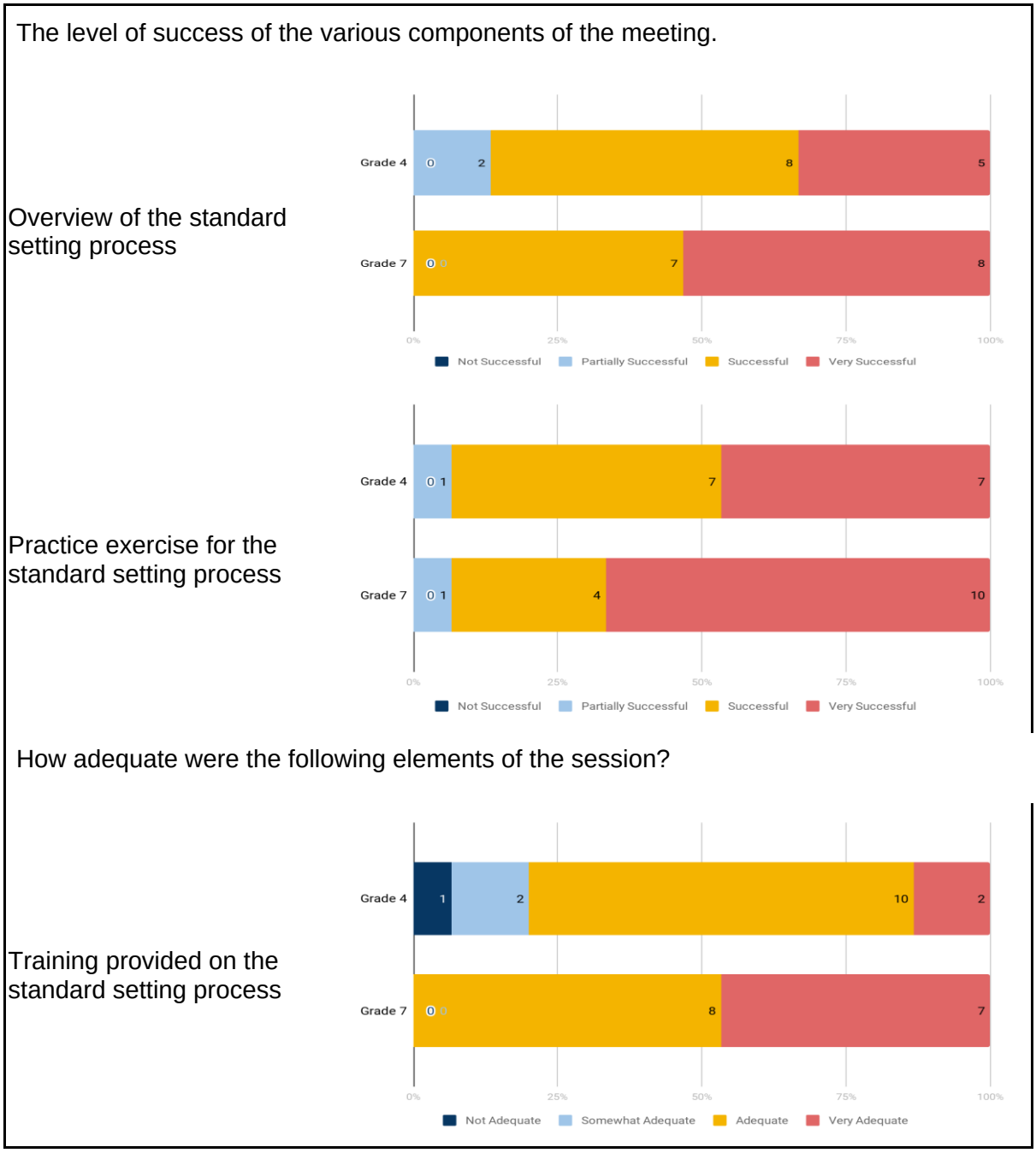


Figure 10: Evaluation results on standard setting process training activities

Perceived Validity of the Workshop

Participants communicated their perceived validity of the workshop and the recommended cut scores as part of the workshop evaluation. Evaluations are important evidence for establishing the validity of recommended cut scores for the performance levels.

Generally, the participants were satisfied with their recommendations and with the workshop as a whole. The participants were provided the opportunity to indicate their confidence in the cut scores recommended by the standard setting committees. Figure 11 displays the results of the evaluation survey across grade-level committees for their confidence in the recommended cut scores. This evaluation was completed prior to the vertical articulation

meeting. During the vertical articulation meeting, the grade 4 participants were provided the opportunity to recommend adjustments to the cut scores for the performance levels. Although there were grade 4 participants which indicated that they were either *Not Confident* or *Somewhat Confident* in the cut score recommendations, no grade 4 participant recommended adjustments to the cut score recommendations during the vertical articulation process.

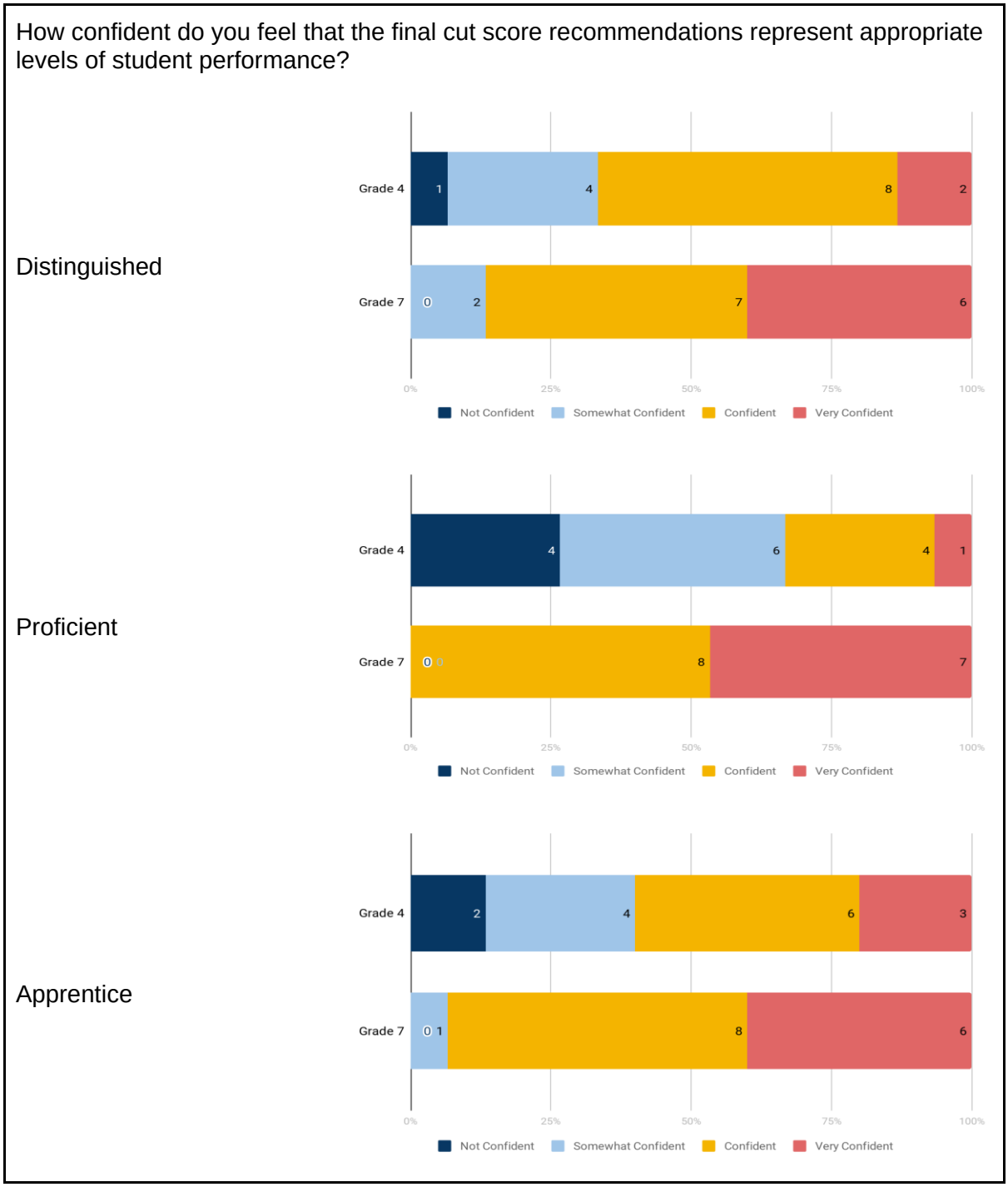


Figure 11: Evaluation results on reasonableness of the cut scores for each performance level

Some of the participants in the standard setting meeting also participated in the follow-up meeting. As part of the follow-up meeting, the participants were provided an opportunity to indicate their confidence in the adjusted cut score recommendations for each performance level. Figure 12 displays the results of the evaluation survey for their confidence in the adjusted cut score recommendations from the follow up meeting.

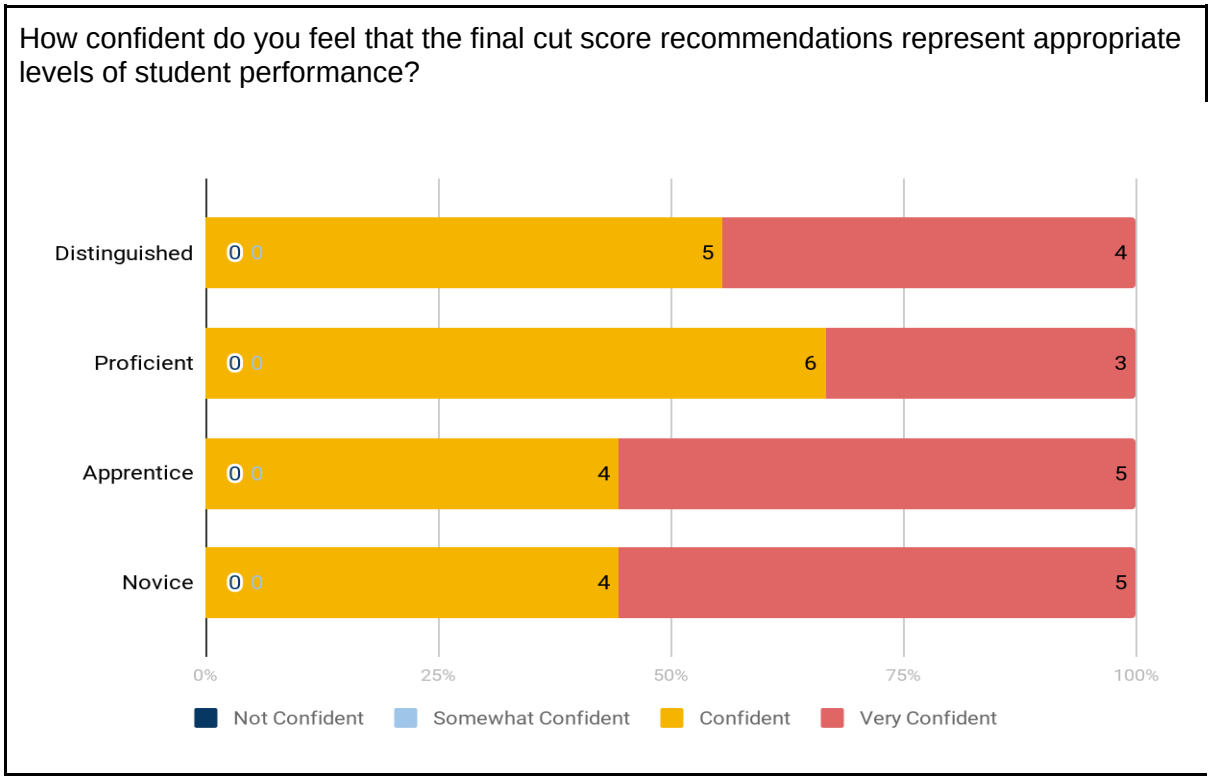


Figure 12: Evaluation results on confidence in adjusted cut score recommendations

The participants were also provided the opportunity to provide additional feedback that indicated overall agreement with the process and the cut score recommendations. Some comments for the standard setting meeting included:

> "Our facilitator for the 7th grade group was wonderful and kept the flow of the group going. He very much understands our process and assessment in KY." Grade 7 participant

> "Great experience as an educator to hear and collaborate with other educators regarding the standard setting process. I feel more prepared and knowledgeable regarding the standards." Grade 7 participant

> "I appreciate being part of the standard setting process, I feel as though as a participant I will be able to share that TEACHERS had an impact on the process rather than outside sources." Grade 4 participant

> "What a very important process that I enjoyed being a part of." Grade 4 participant

Some comments for the follow-up meeting included:

> "Thanks for the opportunity to participate. The additional data was helpful in some cases, but I still feel like we are comparing two very different types of tests, and that the other data was of limited usefulness in setting cut scores for this test."

“Thank you for allowing me to be a part of such a great experience!”

Overall, this feedback from the standard setting participants provides evidence for the validity of the cut score recommendations for each of the performance levels from the standard setting process.

References

Cizek, G. J., Bunch, M. B., & Koons, H. (2004). Setting performance standards: Contemporary methods. *Educational Measurement: Issues and Practice*, 23(4), 31-50.

Davis, L. L. & Moyer, E. L. (2015). PARCC performance level setting technical report. Available from Partnership for Assessment of Readiness for College and Careers (PARCC), Washington, D.C.

Hambleton, R. K., & Plake, B. S. (1995). Extended Angoff procedure to set standards on complex performance assessments. *Applied Measurement in Education*, 8, 41-56.

Plake, B.S., & Cizek, G. J. (2012). Variations on a Theme. In G.J. Cizek (Ed.), Setting performance standards (2nd ed.). New York: Routledge.

Plake, B. S., Ferdous, A. A., Impara, J. C., & Buckendahl, C. W. (2005). *Setting Multiple Performance Standards Using the Yes/No Method: An Alternative Item Mapping Method.* Meeting of the National Council on Measurement in Education. Montreal, Canada.

Appendix A – Participant Meeting Materials

The materials developed for the Kentucky science grade 7 standard setting committee are provided as an example of the materials developed and provided to the participants. Since the materials provided to participant contained secure information, any place where secure information would be provided, that information would be removed. Additionally, the following materials will not be provided within the appendix.

- Test form – This was presented to participants using actual student test books from the spring 2018 administration.
- Open ended item rubrics – These documents presented the scoring rubrics and scoring notes for each open-ended item presented to participants.
- Student exemplars – These documents presented student produced responses for each open-ended item presented to participants.
- Practice item judgment set – This was presented to participants through the Pearson Standard Setting Website as a pdf document.

Kentucky Science Assessment
Standard Setting Meeting
Grades 4 and 7

Agenda

Day 1 – Tuesday, July 17

Introductions and Meeting Orientation

Standard Setting Overview

Experience the Assessment

Kentucky Standards Performance Levels

Lunch

Borderline Performance Level Descriptors

Standard Setting Training

Practice Judgment Activity

Day 2 – Wednesday, July 18

Round 1 Judgments

Round 1 Judgment Feedback and Discussion

Round 2 Judgments

Lunch

Round 2 Judgment Feedback and Discussion

Round 3 Judgments

Round 3 Judgment Feedback and Discussion

Day 3 – Thursday, July 19

Vertical Articulation

Performance Level Descriptor (PLD) Development

Next Steps and Evaluations

Kentucky State-Required Assessments
Nondisclosure Agreement Form

Kentucky state-required assessments requires that all materials used during the standard setting process remain secure. To protect the security of the test items, only authorized persons are permitted to work with or view the materials. All test items and/or components of items, draft or final, and all supporting assessment materials or notes, student responses, and feedback from the standard setting process are to be regarded as secure documents. Thus, they may not be reproduced, discussed, or in any way released or distributed to unauthorized personnel during or after the standard setting process. As a member of the standard setting committee, you may not use any information gleaned from the standard setting process to gain/provide an unfair advantage to schools/districts.

The undersigned is an employee, contractor, consultant, advisory committee member or person otherwise authorized to view material associated with the standard setting process, and hereby agrees to be bound to the terms of this agreement restricting the disclosure of said materials.

_____

Name (printed)

_____

Signature

_____

Date

Participant Information Survey

Page 1

Kentucky Science Assessment
Standard Setting Meeting

Participant Information Survey

Grade 7

Professional Experience

* What is your current position?

- Teacher (K-12 Education)
- Teacher (Higher Education)
- Administrator (School)
- Administrator (District)
- Other Position:

* How many years of professional experience in education do you have?

- None
- 1 to 5 years
- 6 to 10 years
- 11 to 15 years
- 16 to 20 years
- More than 20 years

* How many years of professional experience do you have teaching science grade 7?

- None
- 1 to 5 years
- 6 to 10 years
- 11 to 15 years
- 16 to 20 years
- More than 20 years

* For which of the following populations do you have educational experience with? (Check all that apply.)

- [ ] **Students receiving mainstream special education services**
- [ ] **Students receiving self-contained special education services**
- [ ] **Students who are English language learners**
- [ ] **Students who are receiving general education instruction**
- [ ] **Students who are receiving vocational technical instruction**

* What is the highest degree you have completed?

- ○ **High School Diploma**
- ○ **Associates degree (A.A., A.S.)**
- ○ **Bachelors degree (B.A., B.S.)**
- ○ **Masters degree (M.A., M.S.)**
- ○ **Doctoral degree (Ph.D., Ed.D.)**

Page 2

Demographic Information

* The map displays different regions of Kentucky.

In which region of Kentucky do you reside?

- ○ **West**
- ○ **North West**
- ○ **South West**
- ○ **North Central**
- ○ **South Central**
- ○ **North East**
- ○ **South East**

What is your gender?

○ Male ○ Female ◉ No answer

What is your ethnicity?

○ Hispanic or Latino ○ Not Hispanic or Latino ◉ No answer

What is your race?

○ American Indian or Alaskan Native
○ Asian
○ Black or African American
○ Native Hawaiian or Pacific Islander
○ White
◉ No answer

* Do you currently work in a school district?

○ Yes
○ No

School District Information

* Which word best describes the size of the school district where you work?

○ Small
○ Medium
○ Large

* Which word best describes the type of school district where you work?

○ Rural
○ Metropolitan/Urban
○ Suburban

* Which word best describes the socioeconomic status of the school district where you work?

○ Low
○ Moderate
○ High

Kentucky Science Standard Setting Meeting
July 2018

Experience the Assessment Record Sheet

Grade 7 Science

Sequence	Item ID	KAS	Key	Max Point	Response	Notes
1	SC071600_01	06-ESS1-1		1		
2	SC071600_02	06-ESS1-1	Rubric	4		
3	SC071600_03	06-ESS1-1		1		
4	SC071600_04	06-ESS1-1		1		
5	SC071600_05	5-ESS1-2		1		
6	SC071600_09	5-ESS1-2		1		
7	SC071600_07	5-ESS1-1		1		
8	SC071600_08	5-ESS1-1		1		
9	SC071620_01	07-PS4-2	Rubric	4		
10	SC071620_02	MS-ETS1-1		1		

Note: Only the first page of this document is presented as an example.

Grade 7 Kentucky Science Test Map

Seq	UIN	Key	Points	KAS	SEP	DCI	CC
1	SC071610_01		1	06-LS2-2	2 Processing and Evaluating Information	LS2.A: Interdependent Relationships in Ecosystems	2 Cause and Effect
2	SC071610_02		1	06-LS2-2	2 Processing and Evaluating Information	LS2.A: Interdependent Relationships in Ecosystems	1 Patterns
3	SC071610_03		1	06-LS2-2	3 Synthesizing with Information	LS2.A: Interdependent Relationships in Ecosystems	1 Patterns
4	SC071610_04		2	06-LS2-2	3 Synthesizing with Information	LS2.A: Interdependent Relationships in Ecosystems	1 Patterns
5	SC071610_05		1	5-ESS2-1	3 Synthesizing with Information	ESS2.A: Earth Materials and Systems	3 Systems
6	SC071610_06		1	5-ESS2-1	3 Synthesizing with Information	ESS2.A: Earth Materials and Systems	3 Systems

Note: Only these rows of the test map are presented as an example.

Kentucky Science Standard Setting Meeting
July 2018

Judgment Round Record Sheet

Grade 7 Science

"How many points would a borderline student of the performance level likely earn if they answered the question?"

Seq	Item ID	KAS by Topic	Answer Key	Max Score	Judgment Round 1			Judgment Round 2			Judgment Round 3		
					A	P	D	A	P	D	A	P	D
1	SC071600_01	06-ESS1-1		1									
2	SC071600_02	06-ESS1-1	Rubric	4									
3	SC071600_03	06-ESS1-1		1									
4	SC071600_04	06-ESS1-1		1									
5	SC071600_05	5-ESS1-2		1									
6	SC071600_09	5-ESS1-2		1									
7	SC071600_07	5-ESS1-1		1									
8	SC071600_08	5-ESS1-1		1									
			CLUSTER SC071600 SCORE:	*11*			Panelist's CLUSTER Judgment:						

Kentucky Science Assessment
Standard Setting Meeting
Grade 7

Practice Item Judgment Survey

You are now ready to begin!

For each item in the Practice Judgment item set, do the following for each performance level:

- Review the item in the online system.
- Review the information provided about the item in the item map and answer key. For open response items, review the information in the rubric and exemplars.
- Review the borderline descriptors for the performance level.
- Answer the following questions:

"How many points would a borderline student of the performance level likely earn if he or she answered the question?"

- Record your response to the question for the performance level for the specific item on the judgment record sheet and in the online survey.

Continue reviewing the items until you have provided judgments for each performance level for all of the items.

You will now start the Item Judgment Process for the items in the practice item set.

For each of the items, answer the following question:

"How many points would a borderline student at each performance level likely earn if they answered the question?"

Item: SC071610_01

Key	KAS	SEP	DCI	CC
	06-LS2-2	2 Processing and Evaluating Information	LS2.A: Interdependent Relationships in Ecosystems	2 Cause and Effect

	0 Points	1 Point
Apprentice		
Proficient		
Distinguished		

Item: SC071610_02

Key	KAS	SEP	DCI	CC
	06-LS2-2	2 Processing and Evaluating Information	LS2.A: Interdependent Relationships in Ecosystems	1 Patterns

	0 Points	1 Point
Apprentice		
Proficient		
Distinguished		

Note: Only a couple of items are displayed as an example.

Appendix B – Committee Participant Composition

Table B.1: Participant Position

	Science	
	Grade 4	Grade 7
Teacher (K–12)	12	12
Teacher (Higher Ed.)	1	0
Administrator (School)	1	0
Administrator (District)	0	2
Other	1	1

Table B.2: Years of Teaching Experience

	Science	
	Grade 4	Grade 7
None	0	0
1 to 5 years	0	2
6 to 10 years	4	6
11 to 15 years	4	4
16 to 20 years	1	1
More than 20 years	6	2

Table B.3: Years of Teaching Experience Subject Within Grades

	Science	
	Grade 4	Grade 7
None	0	6
1 to 5 years	1	2
6 to 10 years	7	5
11 to 15 years	3	1
16 to 20 years	3	1
More than 20 years	1	0

Table B.4: Experience Teaching Student Populations

	Science	
	Grade 4	Grade 7
Mainstream special education	13	13
Self-contained special education	4	6
English language learners (ELL)	11	10
General education	15	15
Vocational technical education	0	5

Table B.5: Highest Education Degree

	Science	
	Grade 4	Grade 7
High School Diploma	0	0
Associates degree	0	0
Bachelor's degree	0	1
Master's degree	13	14
Doctoral degree	2	0

Table B.6: Demographic: Regions of Kentucky

	Science	
	Grade 4	Grade 7
West	2	1
North West	3	6
South West	1	1
North Central	5	2
South Central	3	3
North East	0	2
South East	1	0

Table B.7: Demographic: Gender

	Science	
	Grade 4	Grade 7
Male	1	5
Female	14	10
No answer	0	0

Table B.8: Demographic: Ethnicity

	Science	
	Grade 4	Grade 7
Hispanic or Latino	0	0
Not Hispanic or Latino	15	12
No answer	0	0

Table B.9: Demographic: Race

	Science	
	Grade 4	Grade 7
American Indian or Alaskan Native	0	0
Asian	0	0
Black or African American	1	3
Native Hawaiian or Pacific Islander	0	0
White	14	10
No answer	0	0

Table B.10: Currently Work in a School District

	Science	
	Grade 4	Grade 7
Yes	13	15
No	2	0

Table B.11: Size of School District

	Science	
	Grade 4	Grade 7
Small	3	6
Medium	3	3
Large	7	6

Table B.12: Type of School District

	Science	
	Grade 4	Grade 7
Rural	5	10
Metropolitan/Urban	3	3
Suburban	5	2

Table B.13: Socioeconomic Status of School District

	Science	
	Grade 4	Grade 7
Low	6	13
Moderate	6	2
High	1	0

Appendix C – Standard Setting Meeting Agenda

Kentucky Science Assessment
Standard Setting Meeting
Grade 7
Agenda

Day 1

	General Session
8:00 - 8:30 am	Welcome and Orientation
8:30 - 9:00 am	Standard Setting Overview
9:00 – 9:10 am	*Break*
	Breakout Session
9:10 - 9:30 am	Introductions, material orientation, meeting security
9:30 - 10:30 am	Experience the Assessment
10:30 – 11:00 am	Scoring the Assessment
11:00 – 11:30 am	Review and Discuss Standards and Policy Level Descriptors
11:30 – 12:15 pm	*Lunch*
12:15 – 12:45 pm	Borderline Performance Level Descriptors Training
12:45 – 1:30 pm	Borderline PLD Table Discussion
1:30 – 1:40 pm	*Break*

1:40 – 3:00 pm Borderline PLD Group Discussion

3:00 – 3:30 pm Standard Setting Training

3:30 – 5:00 pm Practice Judgment Activity and Discussion

Day 2

8:00 – 9:15 am Round 1 Judgments (Item level judgments)
Round 1 Readiness Form
Panelists work independently to make Round 1 judgments

9:15 – 9:45 am *Break*

9:45 – 10:15 am Round 1 Judgment Feedback
Item Level - Item means and distributions
Test Level – Cut score recommendations; Panelist agreement

10:15 – 10:45 am Table Discussion - Round 1 Feedback
Panelists discuss feedback data at their tables

10:45 – 11:45 am Round 2 Judgments (Item and cluster level judgments)
Round 2 Readiness form
Panelists work independently to make Round 2 judgments

11:45 – 12:30 pm *Lunch*

12:30 – 12:45 pm Round 2 Judgment Feedback
Item Level - Item means and distributions
Test Level - Threshold score recommendations;
Panelist agreement

12:45 – 1:15 pm Table Discussion - Round 2 Feedback

1:15 – 1:45 pm Whole Group Discussion - Round 2 Feedback

1:45 – 2:00 pm *Break*

2:00 – 2:45 pm Round 3 Judgments (Cluster judgments)
- Round 3 Readiness form
- Panelists work independently to make Round 3 judgments

2:45 – 3:00 pm *Break*

3:00 – 3:30 pm Round 3 Judgment Feedback and Discussion
- Test level – Cut score recommendations
- Impact data

3:30 – 4:00 pm Whole Group Discussion

Day 3

Vertical Articulation

8:30 – 8:45 am Introduction and Purpose

8:45 – 9:30 am Cross-grade Impact Data Review

9:30 – 10:00 am Recommend Changes to Round 3 Recommendations

10:00 – 10:15 am *Break*

10:15 – 10:45 am PLD Development Process
- Introduction to the anchor process
- Orientation to the ordered item book (OIB)

10:45 – 11:00 am Table Discussion for Proficient Achievement Level PLDs

11:00 – 11:45 am Whole Group Discussion for Proficient Achievement Level PLDs

11:45 – 12:30 pm *Lunch*

12:30 – 1:15 pm Table Discussion for Apprentice Achievement Level PLDs

1:15 – 2:00 pm	Whole Group Discussion for Apprentice Achievement Level PLDs
2:00 – 2:15 pm	*Break*
2:15 – 3:00 pm	Table Discussion for Distinguished Achievement Level PLDs
3:00 – 3:45 pm	Whole Group Discussion for Apprentice Achievement Level PLDs
3:45 – 4:00 pm	Next Steps, Process Evaluation, and Close Out

Appendix D – Presentations

The presentations are embedded in the document. Select the presentation slide to download the presentation.

General Session

Grade 7 Science Breakout Session – Day 1

Grade 7 Science Breakout Session – Day 2

Grade 7 Science Breakout Session – Day 3

Appendix E – Examples of Feedback Data

Feedback data was provided to participants after each judgment round. The following are examples of feedback data provided to participants.

Individual Item-Level Judgments

This provided participants with the actual item-level judgments that were recorded in the Judgment Survey for the participant during Rounds 1 and 2. This was provided so that the participant could check that the system recorded the judgments correctly.

UIN	A	P	D
SC071600_01	0	1	1
SC071600_02	2	3	4
SC071600_03	1	1	1
SC071600_04	1	1	1
SC071600_05	0	1	1
SC071600_09	0	0	1

Individual Cluster-Level Judgments

This provided participants with the actual cluster-level judgments that were recorded in the Judgment Survey for the participant During Rounds 2 and 3. This was provided so that the participant could check that the system recorded the judgments correctly.

UIN	A	P	D
SC071600	2	5	10
SC071620	3	7	10
SC071602	2	7	11
SC071618	4	9	12

Individual Test-Level Recommendation

This provided the participant with the recommendations for test-level cut scores based on their item judgments for the Apprentice (A), Proficient (P), and Distinguished (D) performance levels.

A Raw Score	P Raw Score	D Raw Score
16	34	48

Table-Level Test-Level Recommendation and Overall Test-Level Recommendation

This provided the participant with the aggregate test-level recommendation, based on the individual participants at the table, including the number of participants, the mean recommendation, the median recommendation, the minimum and maximum recommendation, and the first and third quartiles for each performance level.

	N	Mean	Median	Min	Max	Q1	Q3
A Raw Score	3	14.67	15	13.00	16.00	13.00	16.00
P Raw Score	3	30.67	29	29.00	34.00	29.00	34.00
D Raw Score	3	46.00	46	44.00	48.00	44.00	48.00

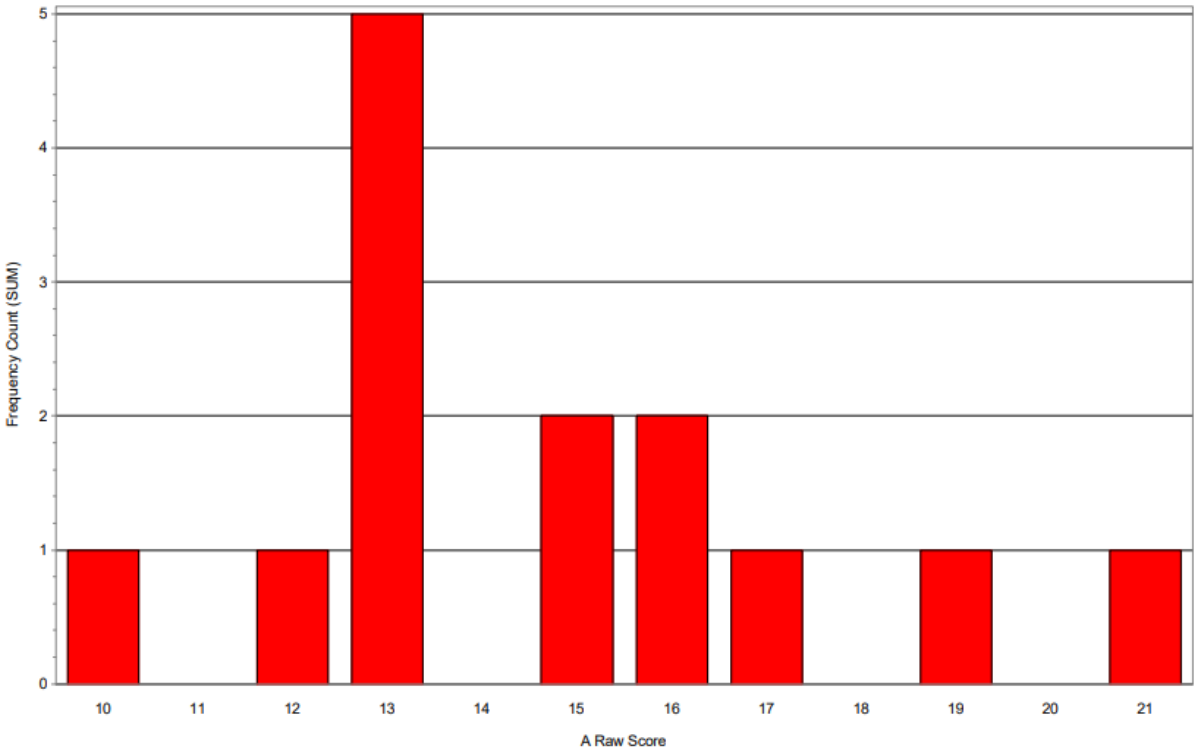
Item-Level Judgment Agreement

This provided the participants with item-level judgment distributions for the committee for each item for each performance level judgment. Additionally, for each performance level, the items with the greatest level of judgment disagreement were identified.

UIN	Max Points	0	1	2	3	4
SC071620_02	1	50%	50%	0%	0%	0%
SC071602_01	2	50%	50%	0%	0%	0%
SC071602_05	4	21%	57%	21%	0%	0%
SC071618_04	1	57%	43%	0%	0%	0%
SC071618_02	1	57%	43%	0%	0%	0%

Test-Level Participant Recommendation Agreement

This feedback was presented to participants by the facilitator. It presented bar graphs displaying the distribution of participant recommendations for the cut score, by raw score, for each performance level: Apprentice (A), Proficient (P), and Distinguished (D). Graphs displaying consecutive performance levels on the scale graph were also presented.



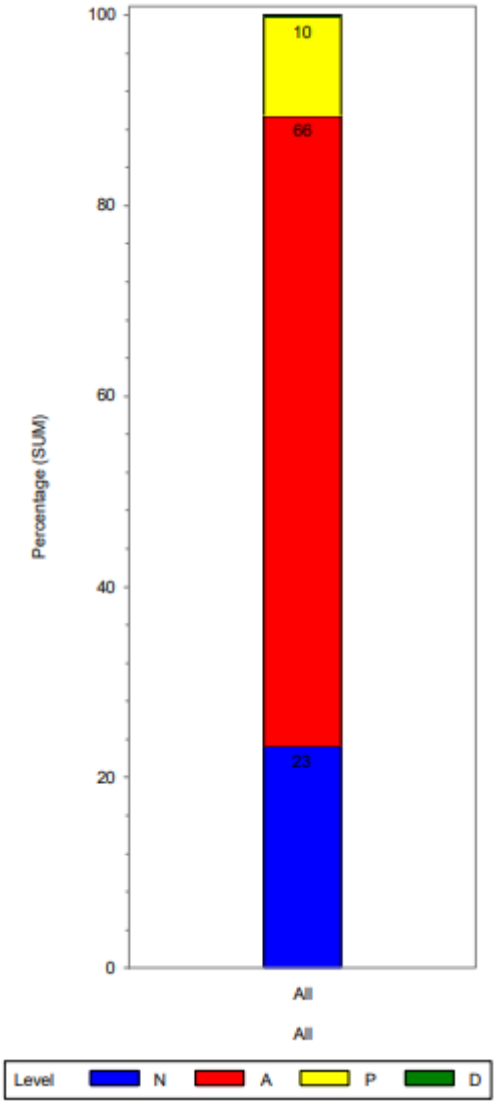
Item Score Mean and Score Distribution

This provided, for each item, the mean score and the distribution of scores received by students during the Spring 2018 administration.

Sequence	Item	Maximum Points	Score Mean	Score Distribution				
				0 pts	1 pt	2 pts	3 pts	4 pts
1	SC071600_01	1	0.40					
2	SC071600_02	4	1.07	39%	30%	18%	10%	2%
3	SC071600_03	1	0.40					
4	SC071600_04	1	0.62					
5	SC071600_05	1	0.16					

Impact Data

This provided the percentage of students expected to be classified into each performance level, Novice (N), Apprentice (A), Proficient (P), and Distinguished (D), based on the committee test-level cut score recommendations for that round. These results were based on the sample of student data from the Spring 2018 administration.



Appendix F – Committee Recommended Cut Scores by Round

Table F.1: Science Grade 4

Performance Level	Maximum Score	Rounds			Vertical Articulation	Follow-Up	Final
		1	2	3			
Apprentice	48	14	11	15	15	15	15
Proficient		26	25	27	27	26	26
Distinguished		37	36	36	36	35	35

Table F.2: Science Grade 7

Performance Level	Maximum Score	Rounds			Vertical Articulation	Follow-Up	Final
		1	2	3			
Apprentice	48	14	12	13	13	12	12
Proficient		30	27	26	26	24	22
Distinguished		43	39	38	38	35	34

Appendix G – Recommended Cut Score Summary Statistics

Table G.1: Science Grade 4 Cut Score Summary Statistics

Round	Statistic	Performance Level		
		Apprentice	Proficient	Distinguished
1	Mean	14.93	28.33	38.20
	Minimum	10	20	32
	Q1	12	25	36
	Median	14	26	37
	Q2	18	32	41
	Maximum	22	39	47
2	Mean	11.13	24.07	35.73
	Minimum	6	18	29
	Q1	9	22	33
	Median	11	25	36
	Q2	13	27	38
	Maximum	18	30	41
3	Mean	14.93	26.87	36.47
	Minimum	12	21	32
	Q1	13	26	36
	Median	15	27	36
	Q2	16	28	37
	Maximum	20	32	41

Table G.2: Science Grade 7 Cut Score Summary Statistics

Round	Statistic	Performance Level		
		Apprentice	Proficient	Distinguished
1	Mean	14.71	29.57	41.86
	Minimum	10	20	33
	Q1	13	29	41
	Median	14	30	43
	Q2	16	30	44
	Maximum	21	40	48
2	Mean	12.40	27.80	39.27
	Minimum	2	19	30
	Q1	10	25	38
	Median	12	27	39
	Q2	16	30	43
	Maximum	19	36	44
3	Mean	12.20	26.13	38.07
	Minimum	6	18	29
	Q1	12	25	37
	Median	13	26	38
	Q2	13	28	40
	Maximum	15	30	44

Appendix H – Test-Level Participant Judgment Agreement

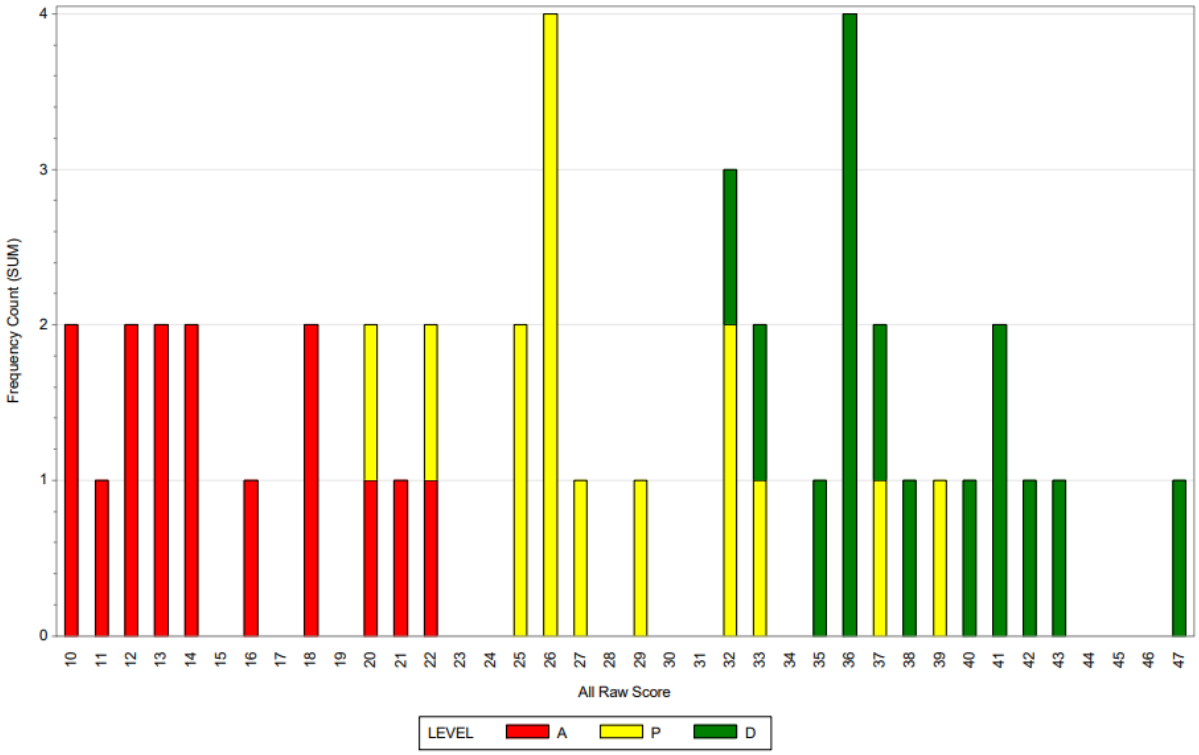


Figure H.1: Grade 4 Science Round 1 Panelist Agreement

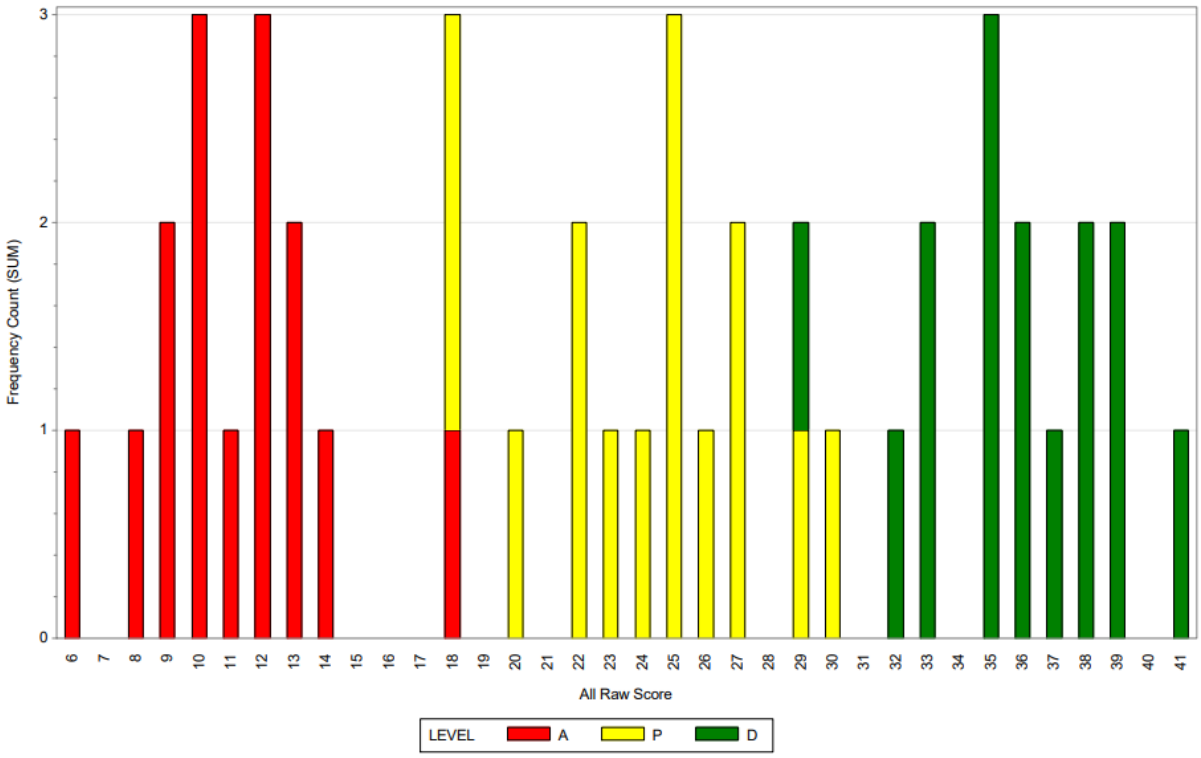


Figure H.2: Grade 4 Science Round 2 Panelist Agreement

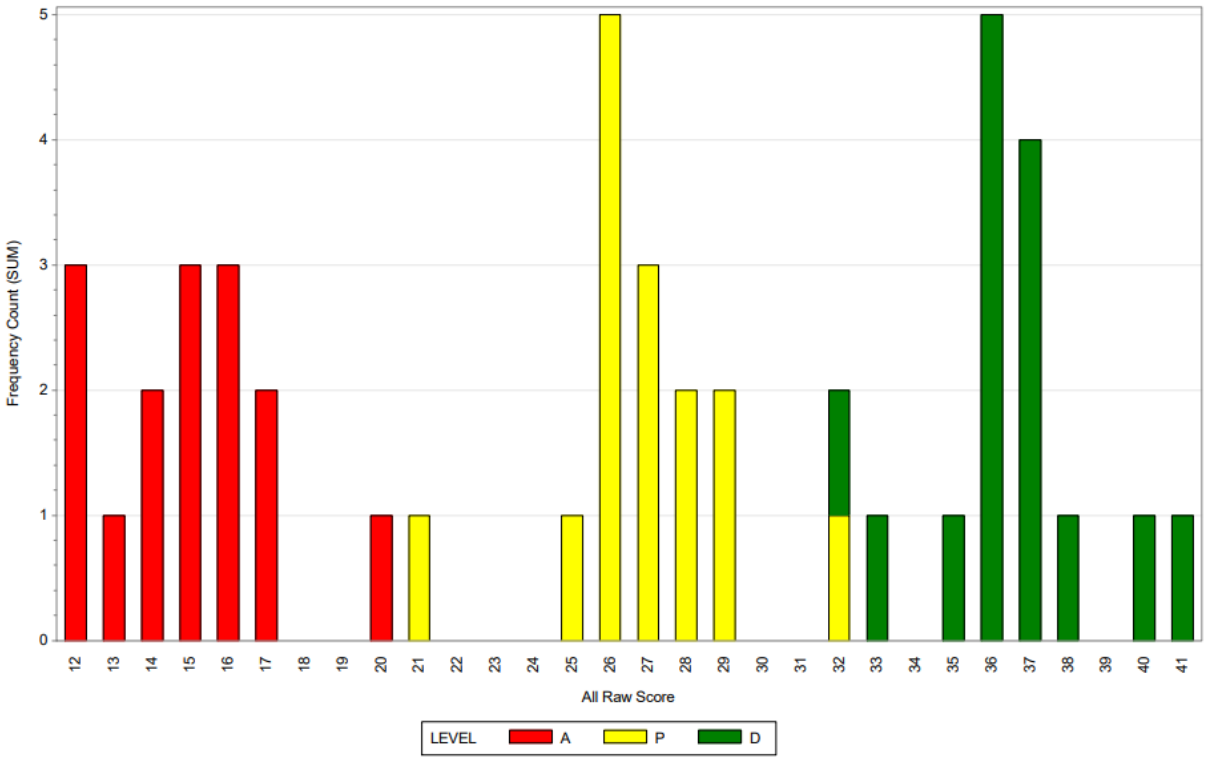


Figure H.3: Grade 4 Science Round 3 Participant Agreement

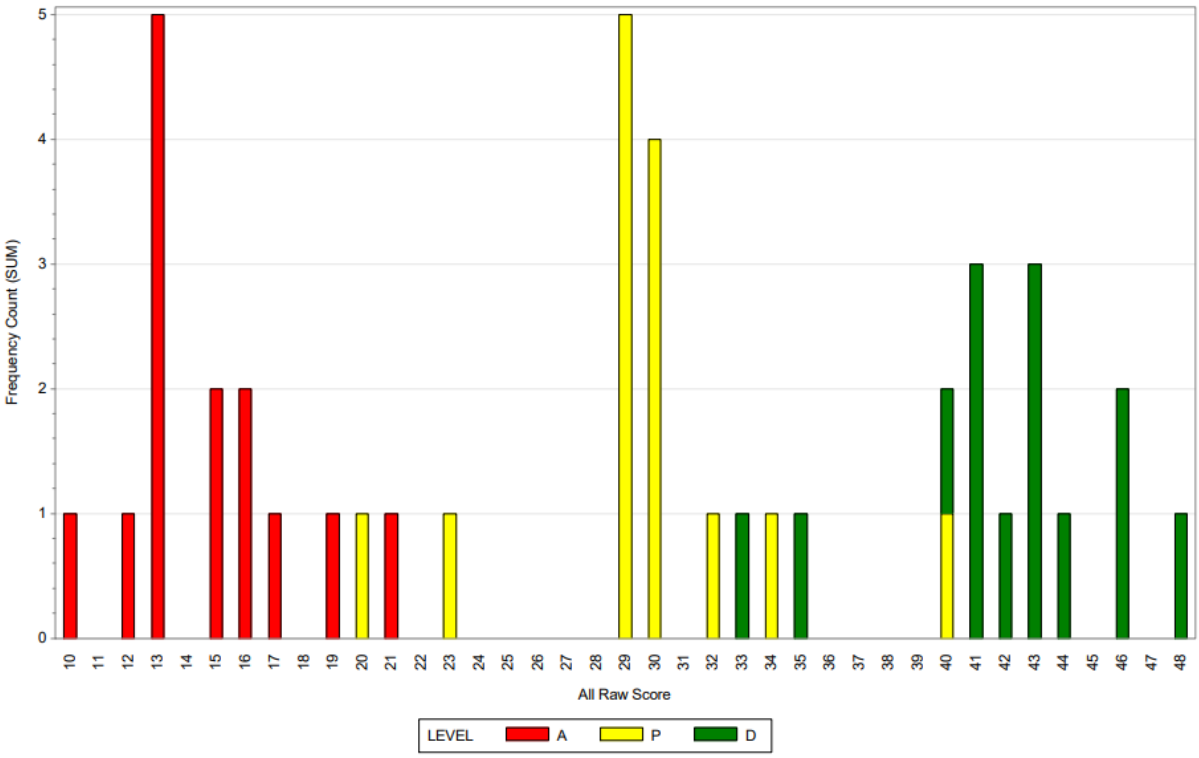


Figure H.4: Grade 7 Science Round 1 Panelist Agreement

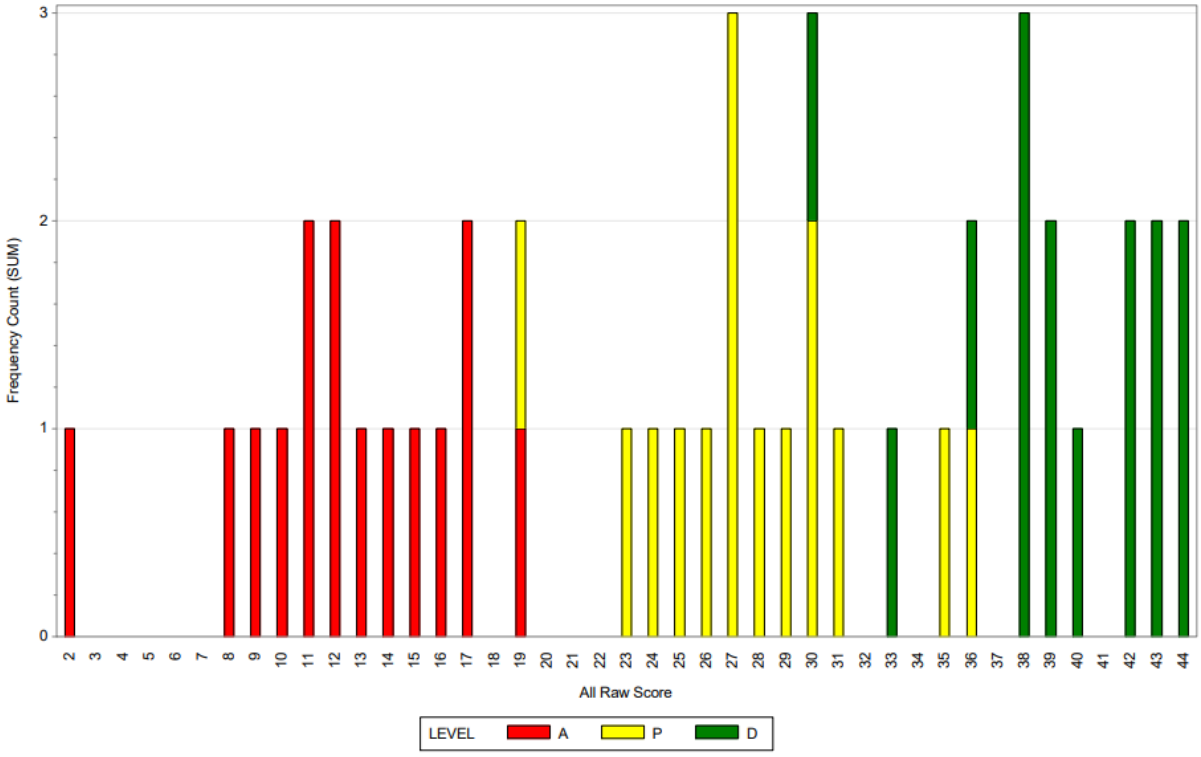


Figure H.5: Grade 7 Science Round 2 Panelist Agreement

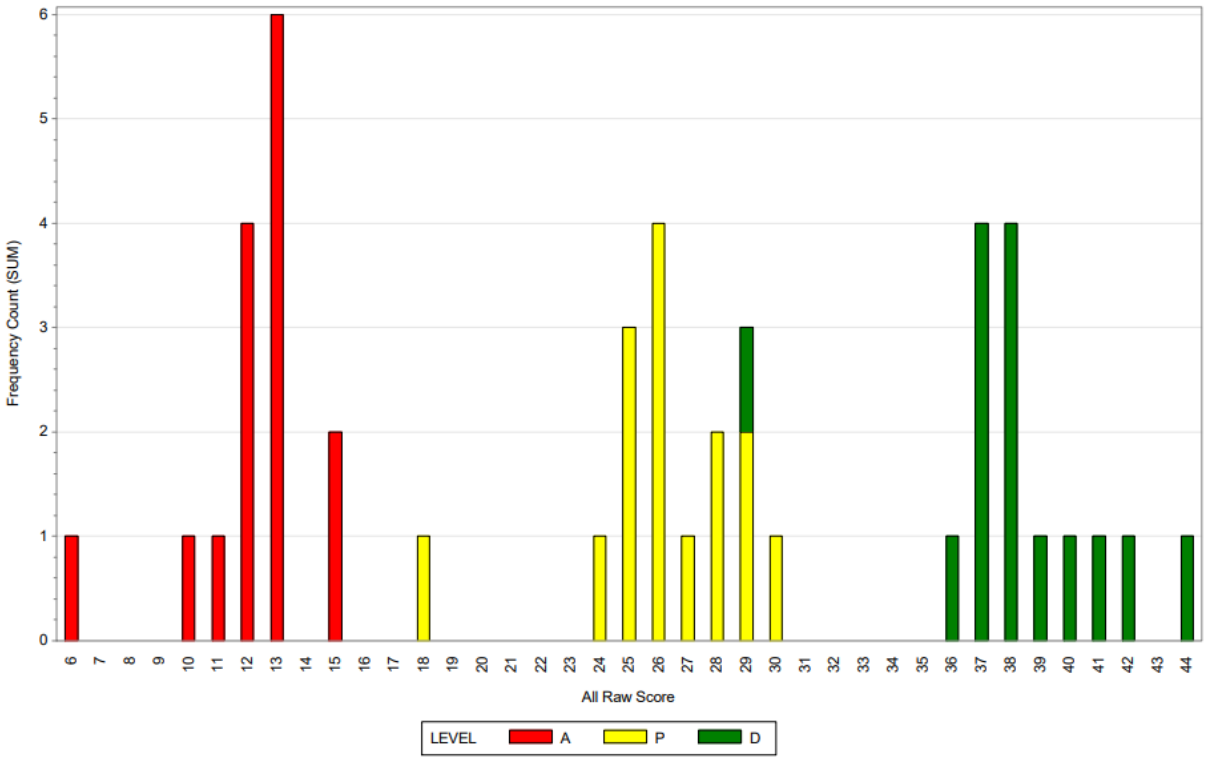


Figure H.6: Grade 7 Science Round 3 Panelist Agreement

Appendix I – Impact Data

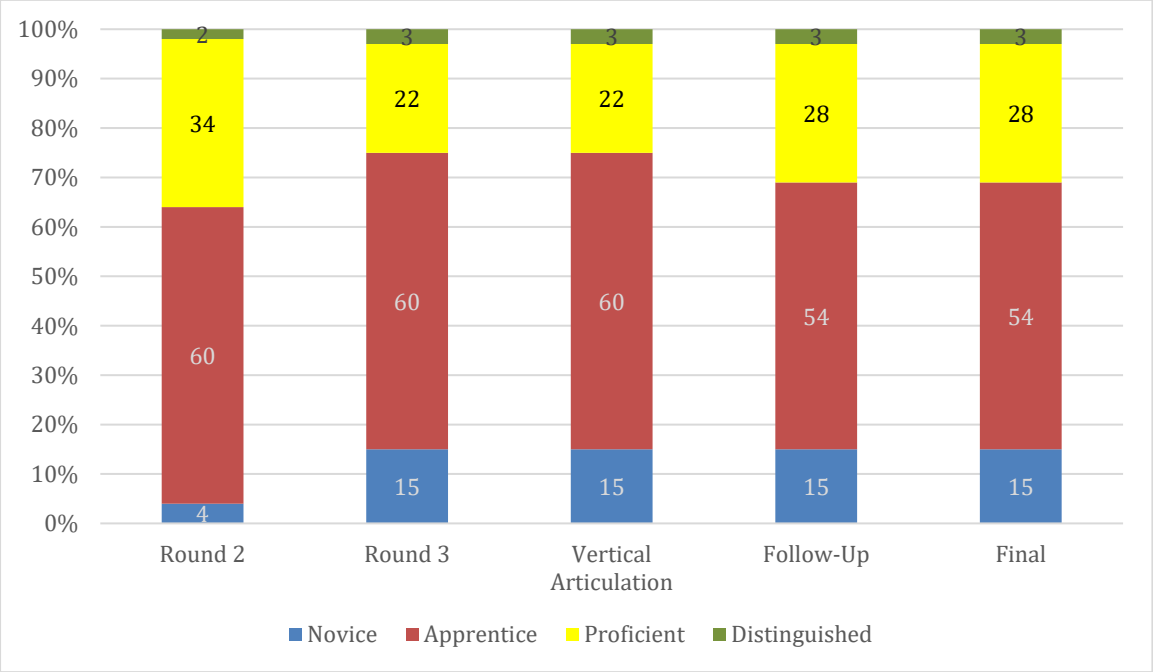


Figure I.1: Grade 4 Impact Data by Round

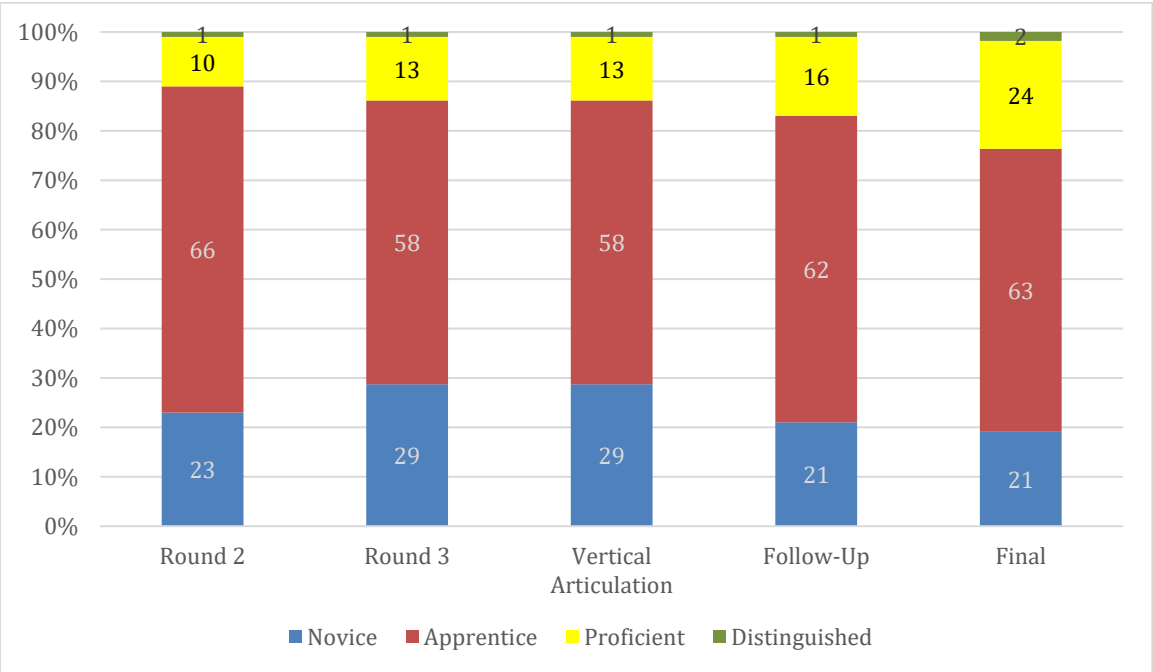


Figure I.2: Grade 7 Impact Data by Round

Appendix J – Performance Level Descriptors

Grade 4 Science Performance Level Descriptors

Kentucky Science Assessments

Performance Level Descriptors

Science - Grade 4

Distinguished Performance Level

A student performing at the **Distinguished** performance level for grade 4 science has a comprehensive understanding of science and engineering concepts and practices incorporated in the Kentucky Academic Standards for Science up to grade 4. The student consistently communicates ideas related to foundational concepts in a sophisticated and complex manner, using thorough supporting detail and explicit examples. The student reasons and solves problems by using appropriate strategies by in an insightful way. Connections between concepts/ideas from different areas of science, when appropriate, are justified and insightful.

The student at the **Distinguished** performance will demonstrate knowledge, skills, and abilities related to the Kentucky Academic Standards for grade 4 science such as:

- Can analyze and interpret data to organize information and make sophisticated and complex predictions.
- Can generate and/or analyze a claim, citing relevant evidence, and utilize cause and effect relationships to support a sophisticated design solution.
- Can accurately and insightfully analyze and interpret data and observations about patterns they observe.

- Can construct and analyze an argument, providing insightful and relevant evidence to support a claim.
- Can define and analyze simple design problems, including identifying criteria for success and considering appropriate constraints.
- Can interpret data to identify patterns used as evidence to construct an insightful explanation.
- Can formulate insightful questions to predict insightful outcomes.
- Can accurately develop or use a model that insightfully explains components and relationships within a system.

Proficient Performance Level

A student performing at the Proficient level for grade 4 science has a broad understanding of science and engineering concepts and practice incorporated in the Kentucky Academic Standards for Science up to grade 4. The student usually communicates ideas accurately using clear and appropriate examples of foundational concepts, supporting or justifying those ideas with relevant details and evidence. Problem-solving and critical thinking skills are used effectively and connections between concepts/ideas from different areas of science, when present, are reasonable and appropriate.

The student at the **Proficient** performance will demonstrate knowledge, skills, and abilities related to the Kentucky Academic Standards for grade 4 science such as:

- Can analyze and interpret data to organize information and make predictions.
- Can make and/or analyze a claim, citing relevant evidence and utilizing cause and effect relationships, to support a design solution.
- Can accurately analyze and interpret data and observations about patterns they observe.
- Can construct an argument with reasonable evidence to support a claim.
- Can analyze a simple design problem, including identifying criteria for success and considering appropriate constraints.
- Can interpret data to identify patterns used as evidence to construct a reasonable explanation.
- Can formulate questions to predict reasonable outcomes.
- Can accurately develop or use a model that explains most components and relationships within a system.

Apprentice Performance Level

A student performing at the Apprentice level for grade 4 science has a basic understanding of science and engineering concepts and practices incorporated in the Kentucky Academic Standards for Science up to grade 4. The student communicates ideas about foundational concepts in a basic manner, but explanations, solutions or justifications may be unclear or ineffective. The student demonstrates some problem-solving and critical thinking skills using concepts/ideas from different areas of science, but they are not consistently applied.

The student at the **Apprentice** performance will demonstrate knowledge, skills, and abilities related to the Kentucky Academic Standards for grade 4 science such as:

- Can analyze or interpret data in a limited way to organize information or make predictions.
- Can make and/or analyze a claim, citing little or incomplete evidence.
- Can analyze or interpret data and observations about patterns they observe.
- Can construct an argument with limited evidence to support a claim.
- Can analyze a simple design solution which may include criteria for success and/or appropriate constraints.
- Can interpret data to identify patterns used as evidence to construct a limited explanation.

- Can recognize reasonable questions to make limited predictions.
- Can recognize, develop or use a model that explains some components and relationships within a system.

- Can develop a partial model or use a given model that shows and minimally explains components and relationships.

Novice Performance Level

A student performing at the **Novice** performance level for grade 4 science has a minimal understanding of the three dimensions of the science and engineering concepts and practices incorporated in the Kentucky Academic Standards for science up through grade 4. The student communicates ideas ineffectively or inaccurately, providing little detail and little or no support. Attempts at problem-solving or critical thinking are minimal or inappropriate.

The student at the **Novice** performance level does not demonstrate the knowledge, skills, and abilities to be classified into the **Apprentice** performance level.

Grade 7 Science Performance Level Descriptors

Kentucky Science Assessments

Performance Level Descriptors

Science - Grade 7

Distinguished Performance Level

A student performing at the **Distinguished** performance level for grade 7 science has a comprehensive understanding of the three dimensions of the science and engineering concepts and practices incorporated in the Kentucky Academic Standards for science up through grade 7. The student consistently communicates ideas in a sophisticated and complex manner, using thorough supporting detail and explicit examples. The student reasons and solves problems by using appropriate strategies in an insightful way. Connections between concepts/ideas from different areas of science, when appropriate, are justified and insightful.

The student at the **Distinguished** performance level will demonstrate knowledge, skills, and abilities related to the Kentucky Academic Standards for grade 7 science such as:

- Can develop detailed models that clearly represent the relationships within systems
- Can consistently use, analyze, and evaluate models, data, evidence and claims in insightful ways
- Can make appropriate predictions using patterns within graphical displays
- Can construct and present arguments supported with relevant evidence and reasoning in a sophisticated manner
- Can critique the arguments and reasoning of others
- Can plan a detailed investigation, identifying the appropriate variables and controls

- Can evaluate competing solution designs that meet the criteria and constraints
- Can explain in detail the relationships and interactions between structure and function
- Can understand and apply knowledge and appropriate terminology in a relevant way

Proficient Performance Level

A student performing at the **Proficient** performance level for grade 7 science has a broad understanding of the three dimensions of the science and engineering concepts and practices incorporated in the Kentucky Academic Standards for science up through grade 7. The student usually communicates ideas accurately using clear and appropriate examples, supporting or justifying those ideas with relevant details and evidence. Problem-solving and critical thinking skills are used effectively. Connections between concepts/ideas from different areas of science, when present, are reasonable and appropriate.

The student at the **Proficient** performance level will demonstrate knowledge, skills, and abilities related to the Kentucky Academic Standards for grade 7 science, such as:

- Can develop models that represent most of the relationships within systems
- Can effectively use, analyze, and evaluate models, data, evidence and claims
- Can revise a model to reduce limitations, including correcting errors
- Can identify a pattern within graphical displays to indicate relationships that exist within a system
- Can describe systems in terms of their components, roles, and interactions
- Can construct, use, and present arguments supported with relevant evidence
- Can plan an investigation, identifying the variables and controls
- Can define the problem and design a solution that meets the criteria and constraints
- Can identify the relationship and interaction between structure and function

Apprentice Performance Level

A student performing at the **Apprentice** performance level for grade 7 science has a basic understanding of the three dimensions of the science and engineering concepts and practices incorporated in the Kentucky Academic Standards for science up through grade 7. The student demonstrates some problem-solving and critical thinking skills, but they are not consistently applied. The student communicates ideas in a basic manner, but explanations, solutions or justifications may be unclear or ineffective.

The student at the **Apprentice** performance level will demonstrate knowledge, skills, and abilities related to the Kentucky Academic Standards for grade 7 science such as:

- Can develop models that represent limited or basic relationships within systems
- Can attempt to support claims and/or reasoning, demonstrating inconsistencies and ineffectiveness
- Can identify a basic pattern within graphical displays to indicate limited or partial relationships that exist within a system
- Can partially describe systems in terms of their components, roles, and interactions
- Can construct, use, and present arguments, with partial effectiveness, supported with limited evidence
- Can inconsistently identify variables within a defined investigation
- Can understand constraints and criteria with limited factors

Novice Performance Level

A student performing at the **Novice** performance level for grade 7 science has a minimal understanding of the three dimensions of the science and engineering concepts and practices incorporated in the Kentucky Academic Standards for science up through grade 7. The student communicates ideas ineffectively or inaccurately, providing little detail and little or no support. Attempts at problem-solving or critical thinking are minimal or inappropriate.

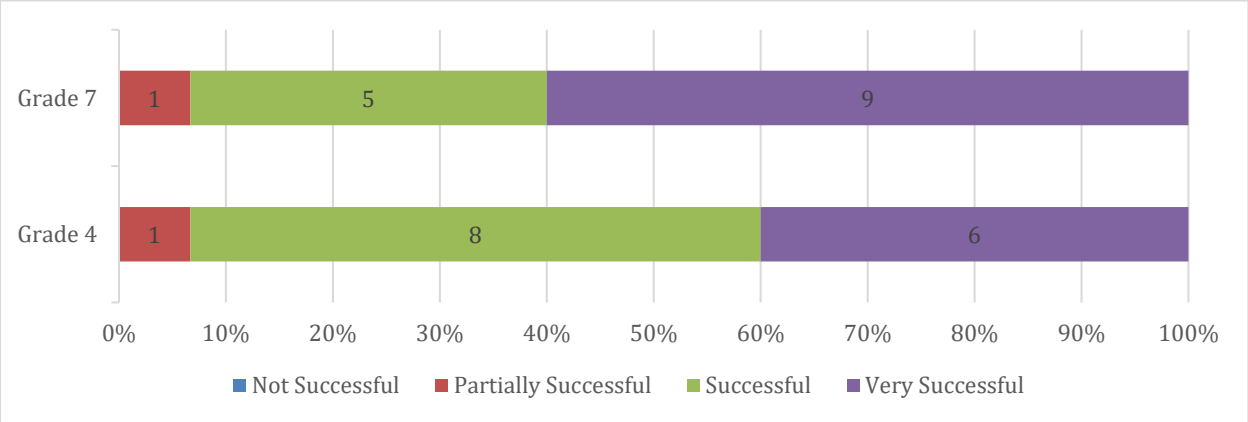
The student at the **Novice** performance level does not demonstrate the knowledge, skills, and abilities to be classified into the **Apprentice** performance level.

Appendix K – Participant Evaluation Results

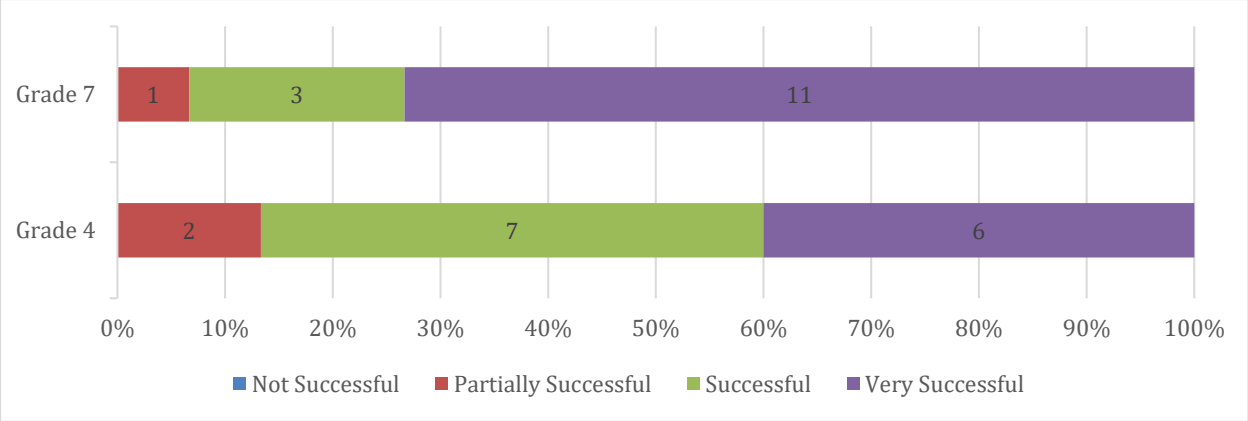
Breakout Session Process Evaluation

Question 1: Select the option that best reflects your opinion about the level of success of the various components of the meeting in which you participated. The activities were designed to help you both understand the process and be supportive of the recommendations made by the committee.

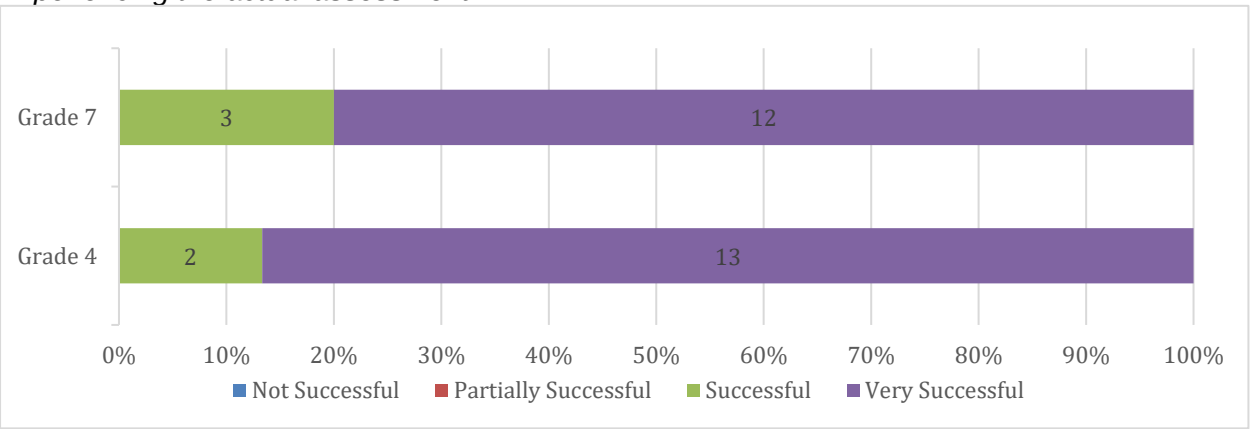
Overview of the Kentucky Science assessments



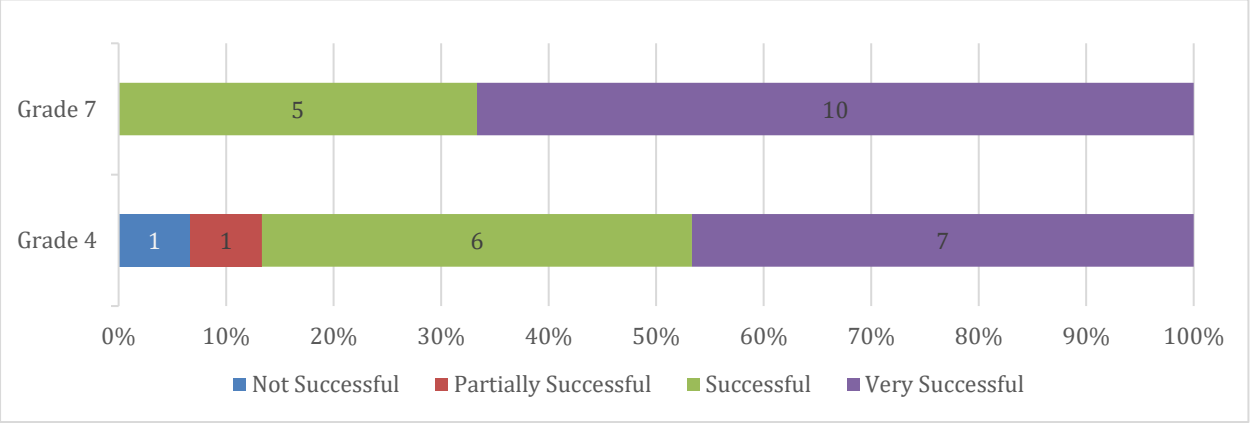
Introduction to the standard setting process



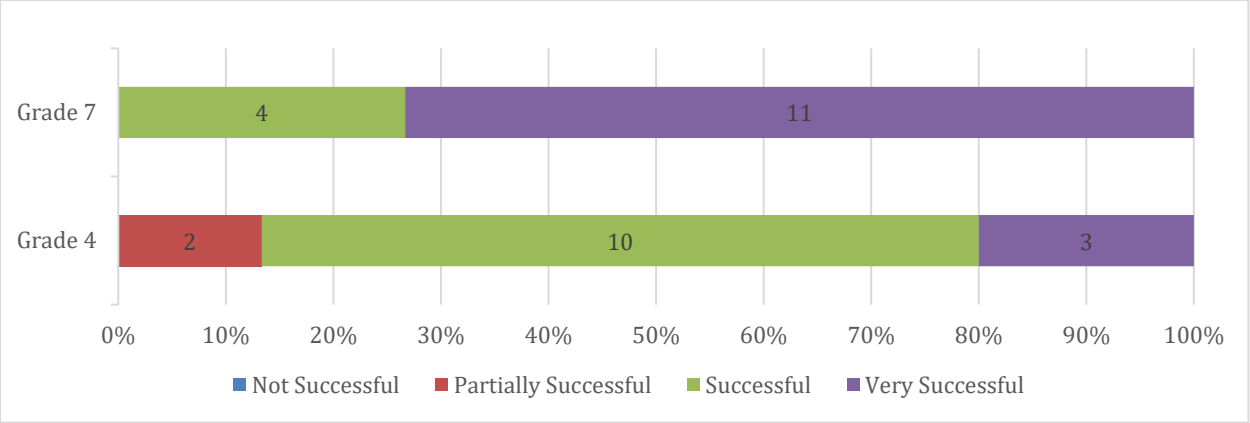
Experiencing the actual assessment



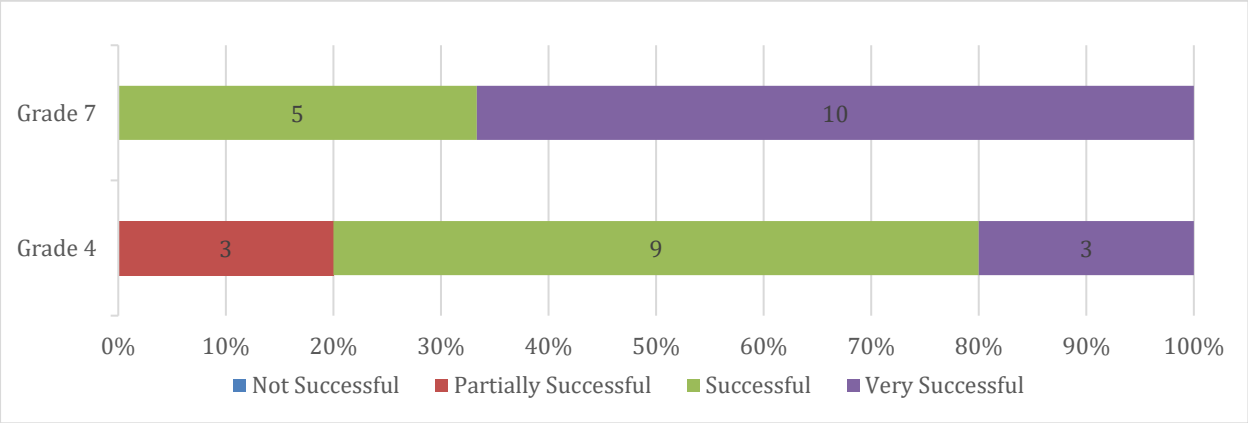
Discussion of the scoring of items on the assessment



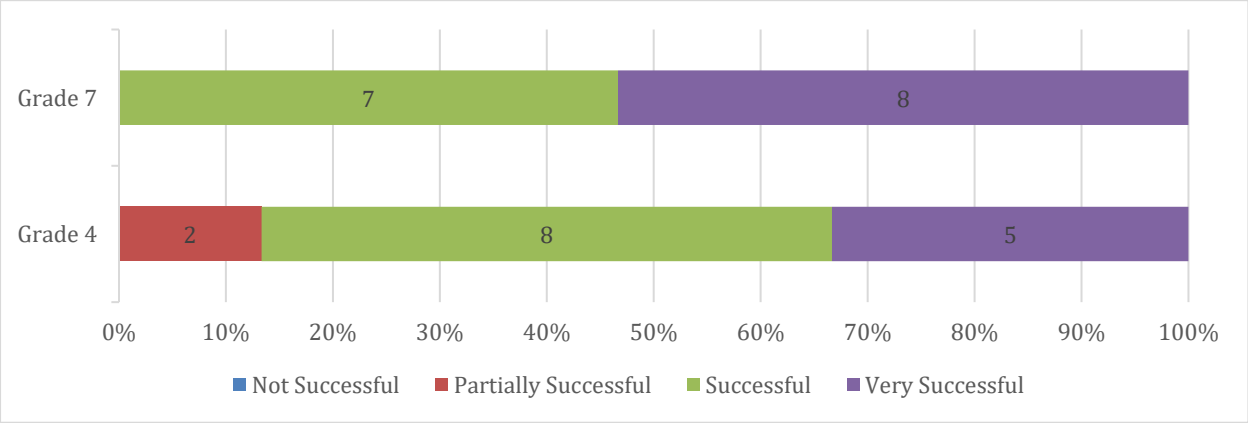
Discussion of performance level descriptors (PLDs)



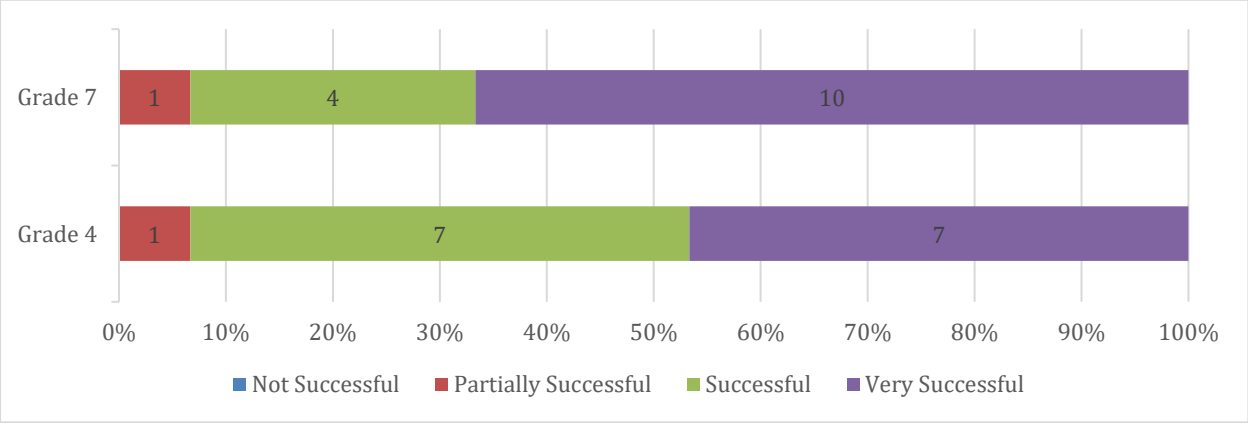
Development and discussion of the borderline performance level descriptors



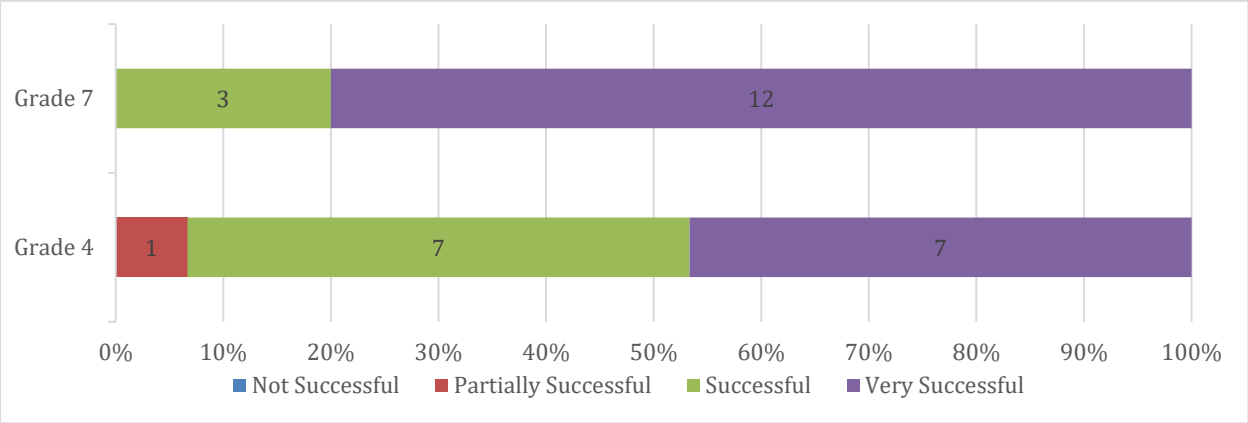
Overview of the standard-setting procedure



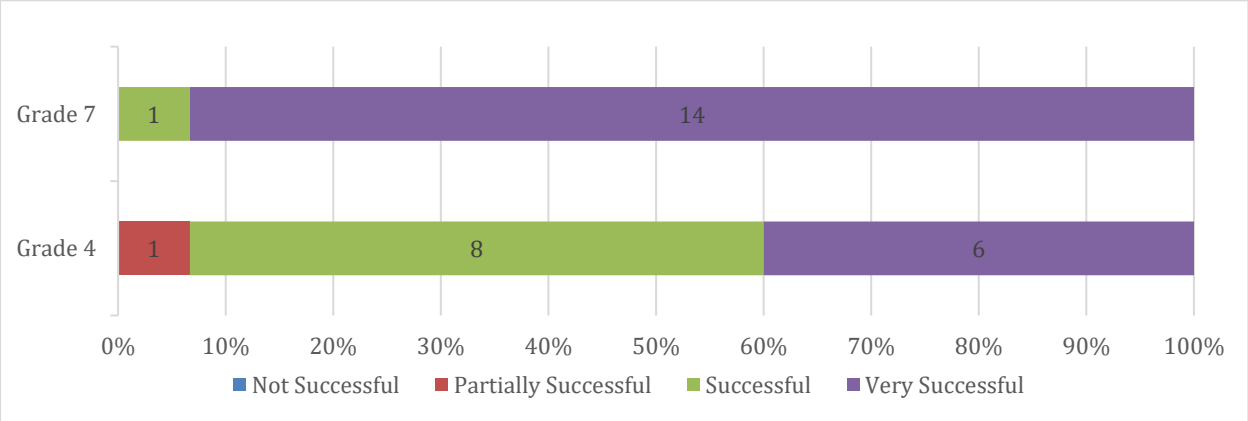
Practice exercise for the standard-setting procedure



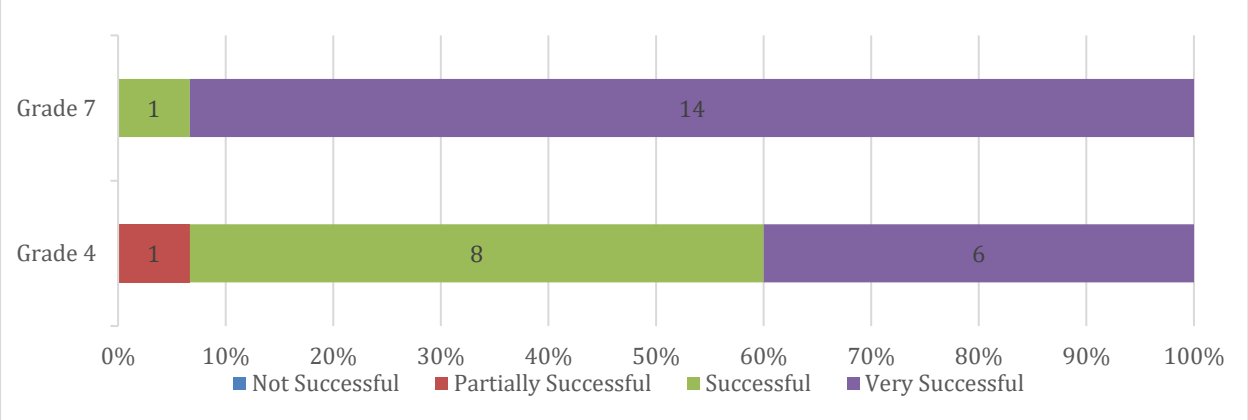
Judgment rounds



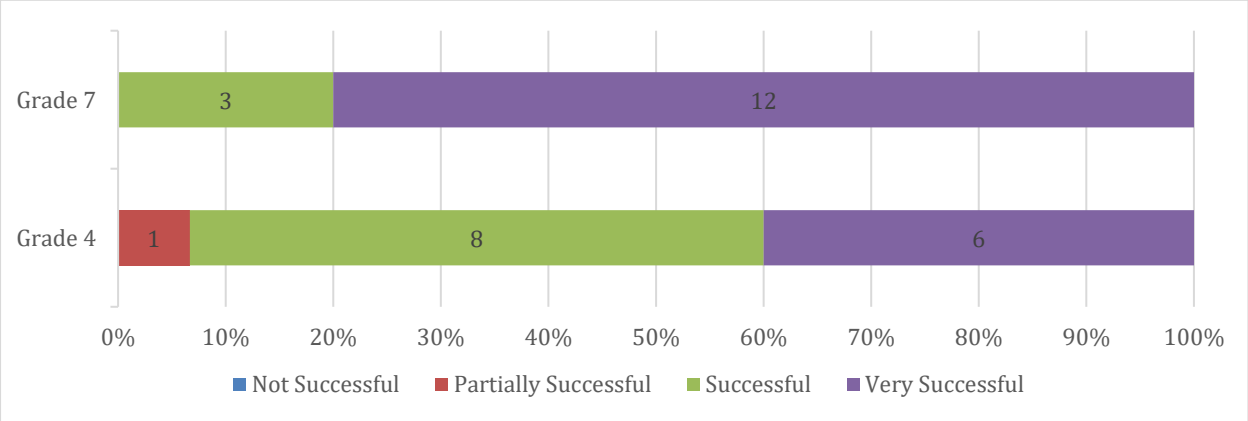
Judgment round feedback - table-level statistics



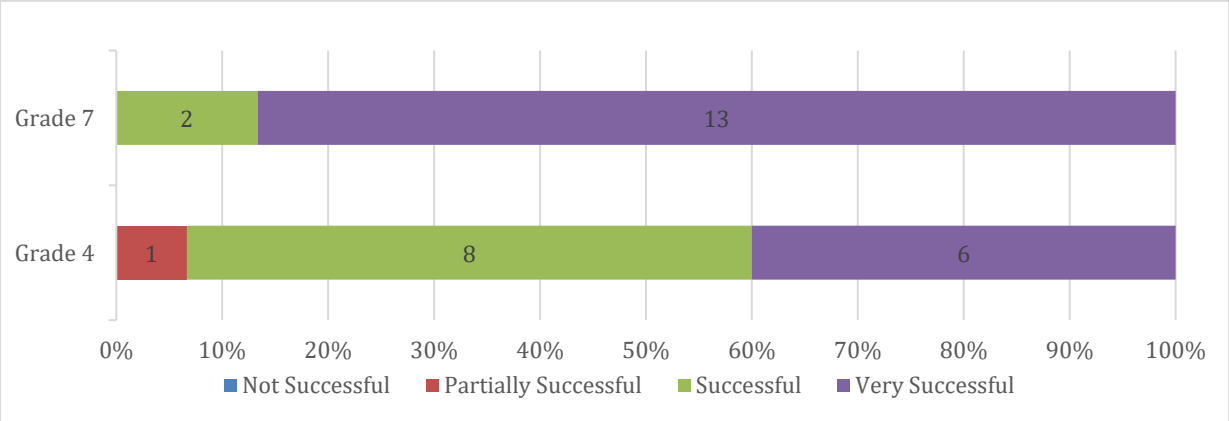
Judgment round feedback - committee-level statistics



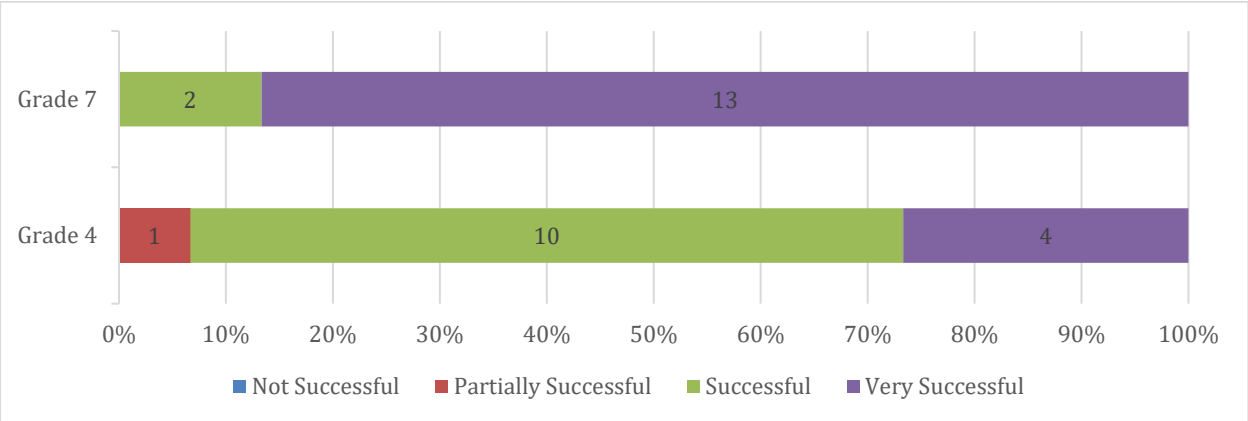
Judgment round feedback - panelist agreement data



Judgment round feedback - impact data

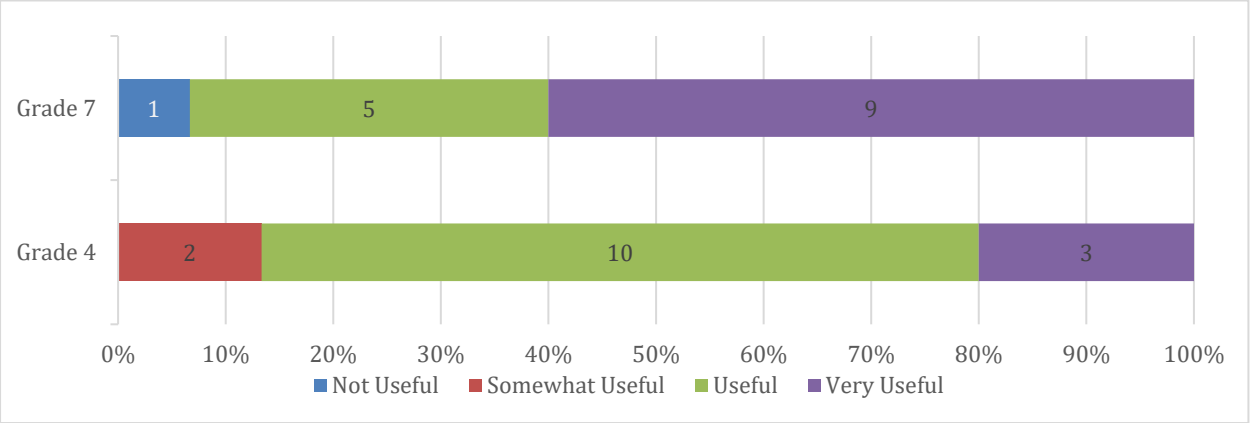


Discussions after each round



Question 2: How useful do you feel the following activities or information were in assisting you to make your recommendations?

Performance Level Descriptors (PLDs)



Borderline performance level descriptors

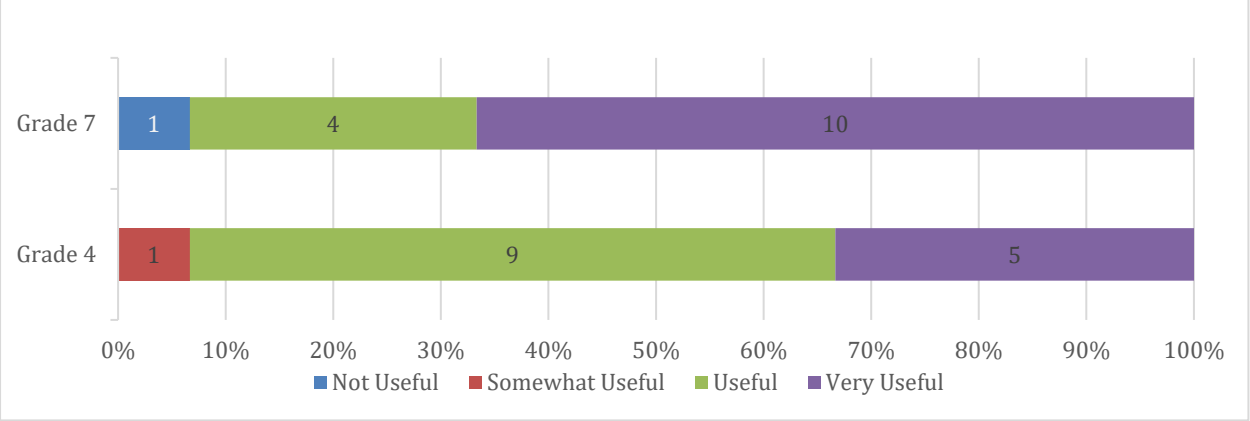
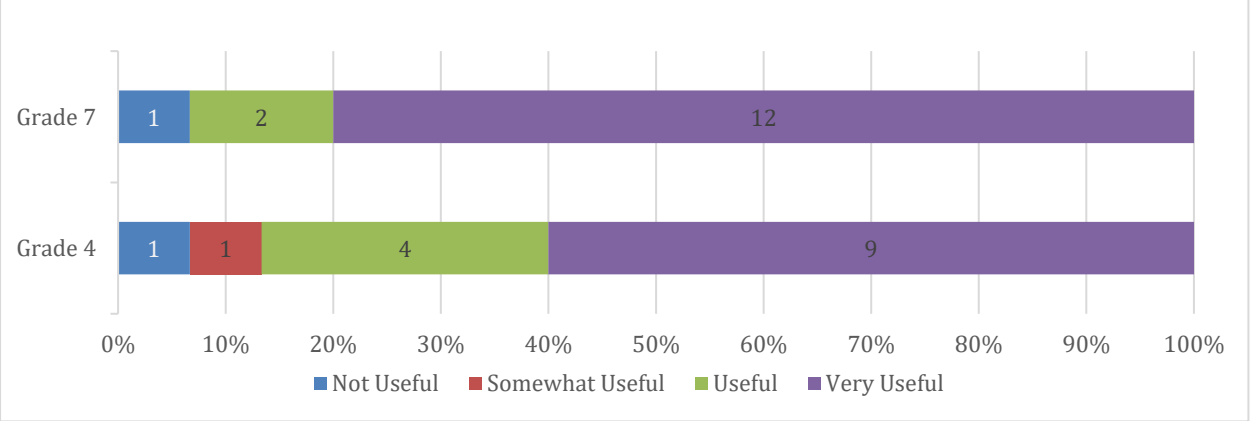
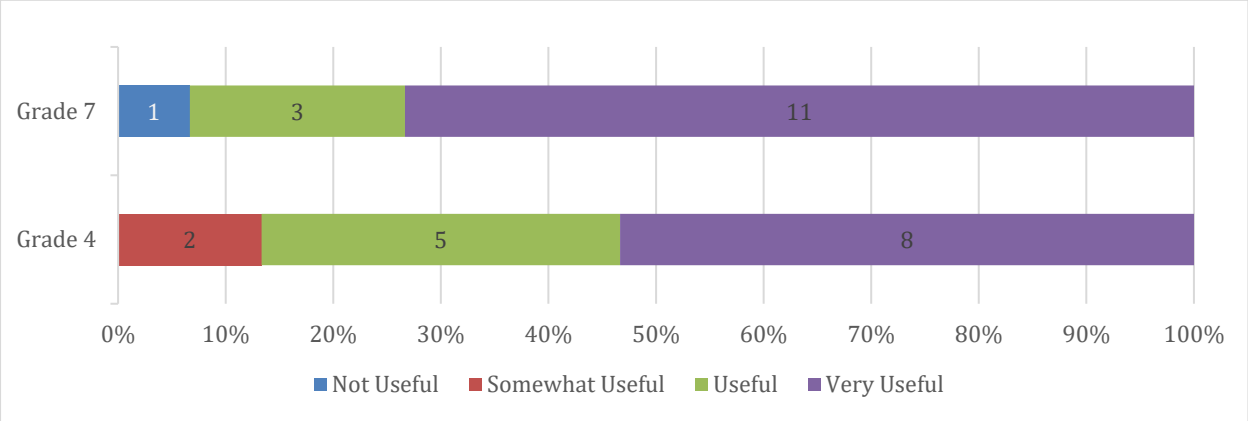


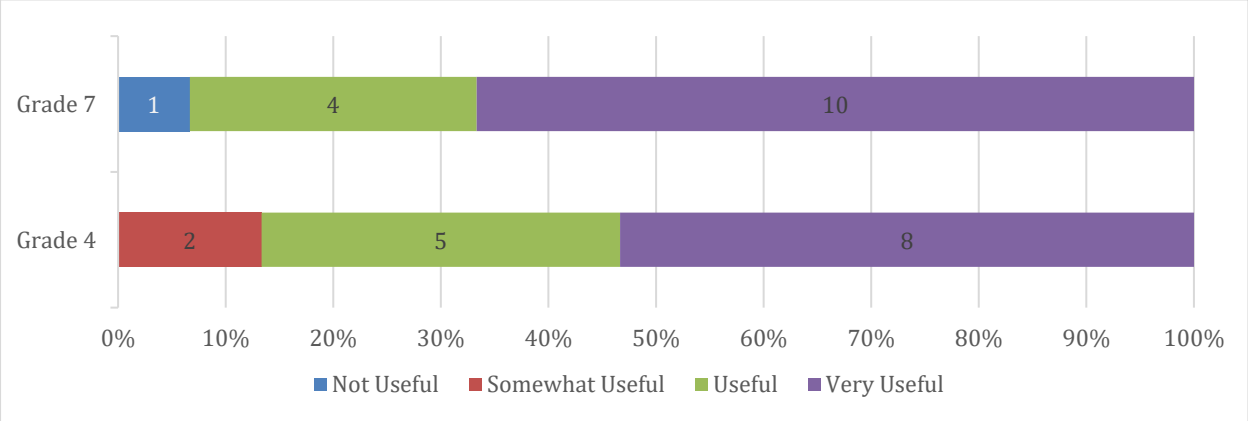
Table-level statistics after Rounds 1 and 2



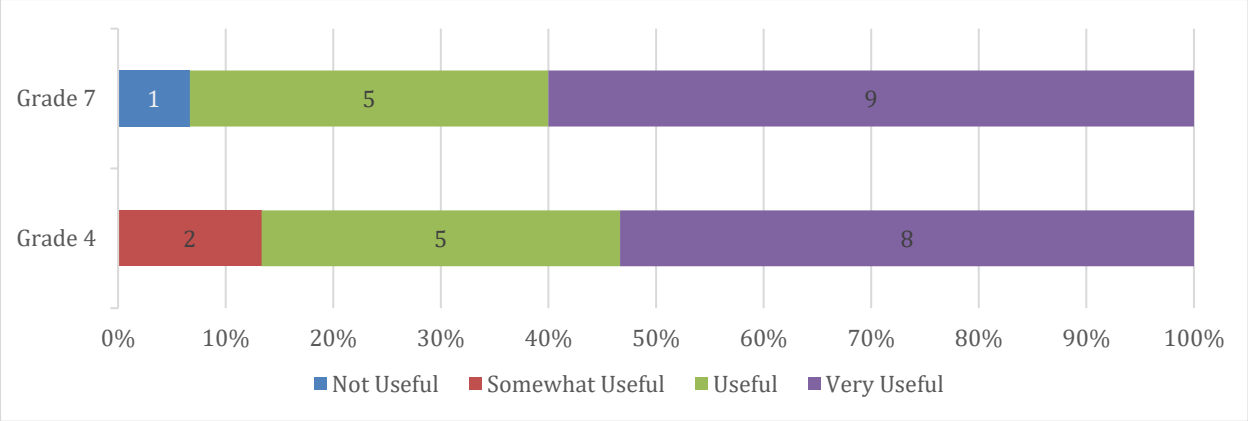
Committee-level statistics after Round 2



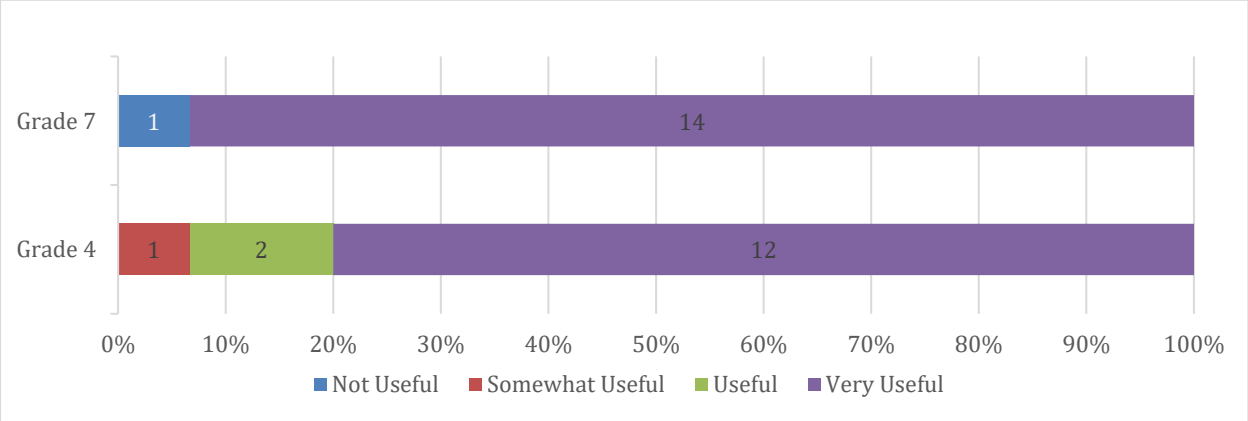
Panelist agreement data provided after Round 1



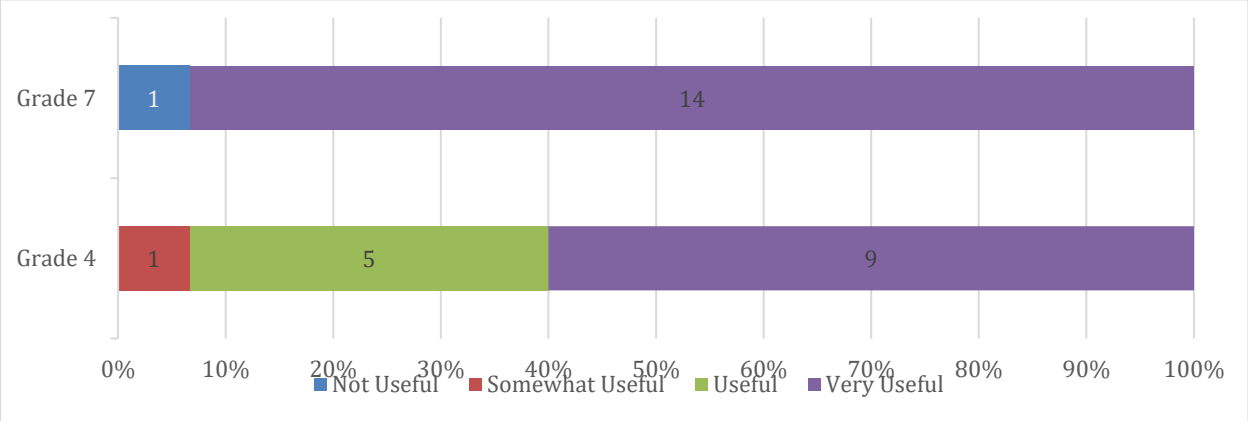
Panelist agreement data provided after Round 2



Impact data after Round 2

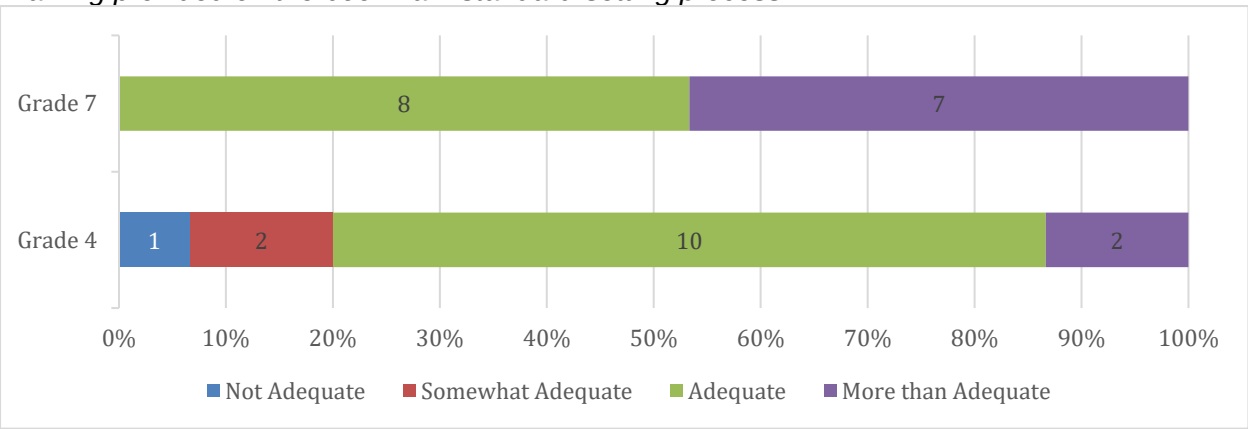


Discussion after each judgment round

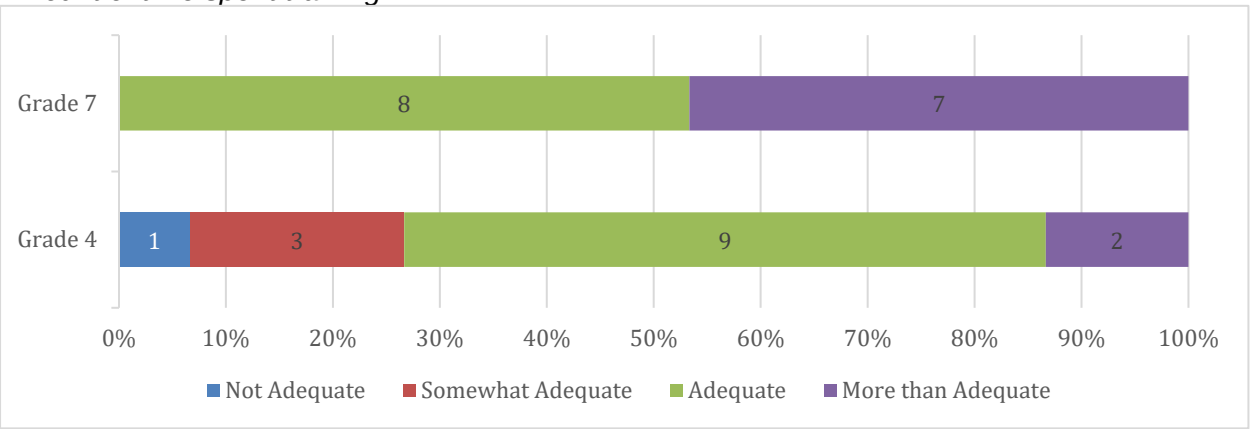


Question 3: How adequate were the following elements of the session?

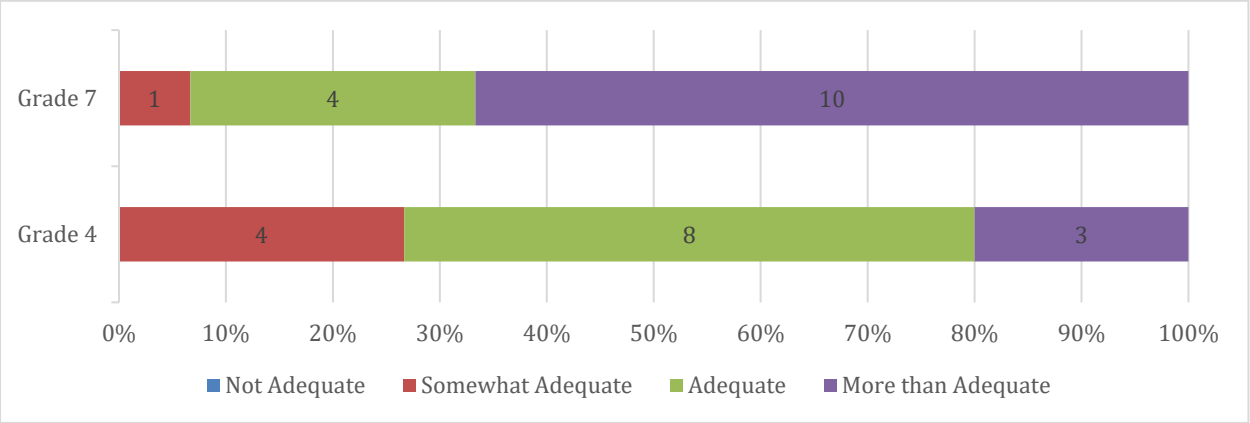
Training provided on the bookmark standard-setting process



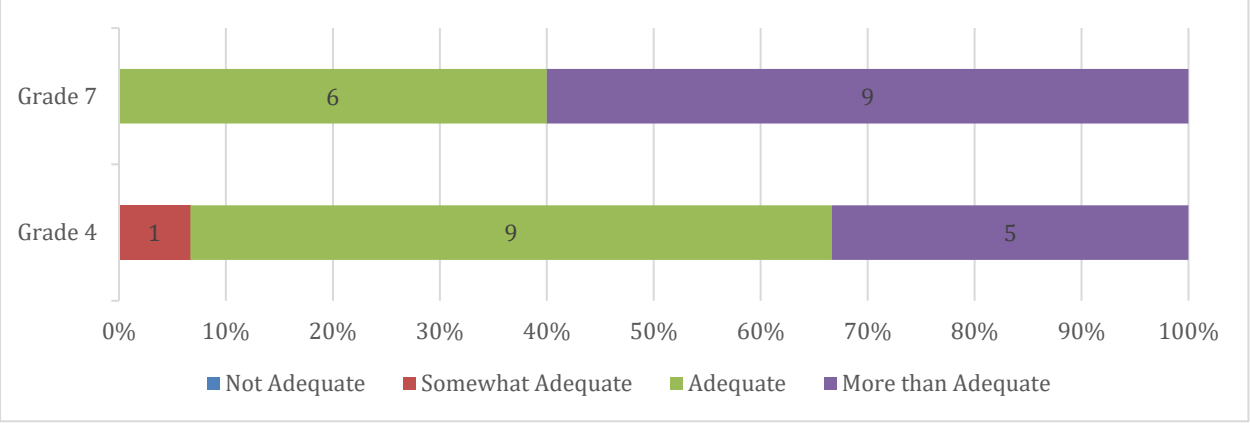
Amount of time spent training



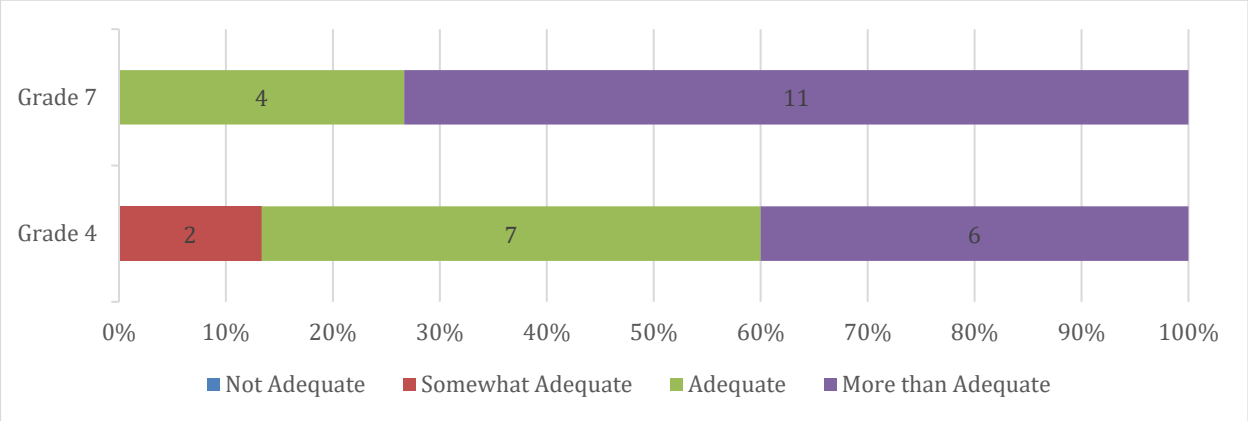
Total amount of time to create and discuss borderline performance level descriptors



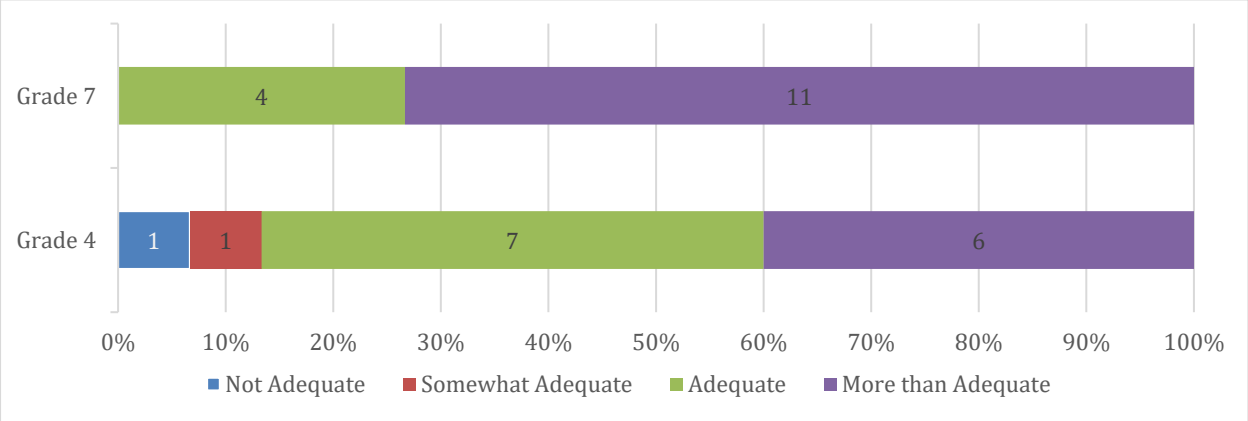
Total amount of time to discuss the practice judgments



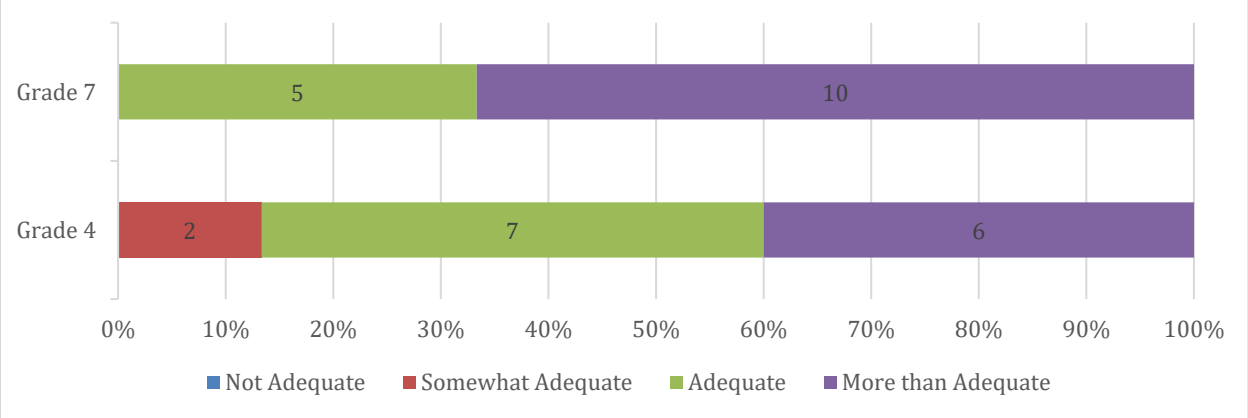
Amount of time to make judgments



Visual presentation of the feedback provided

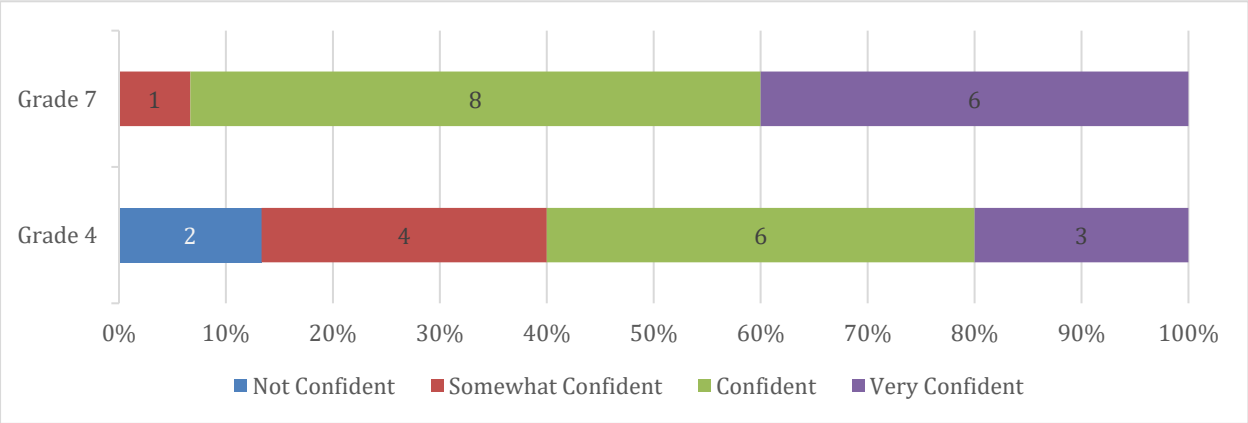


Number of judgment rounds

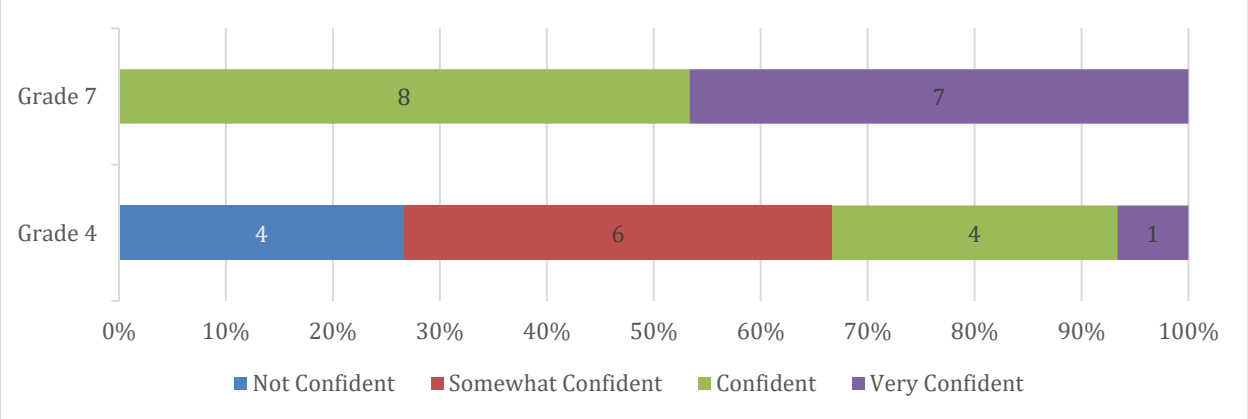


Question 4: How confident do you feel that the final cut score recommendations for grade 4 & 7 science represent appropriate levels of student performance?

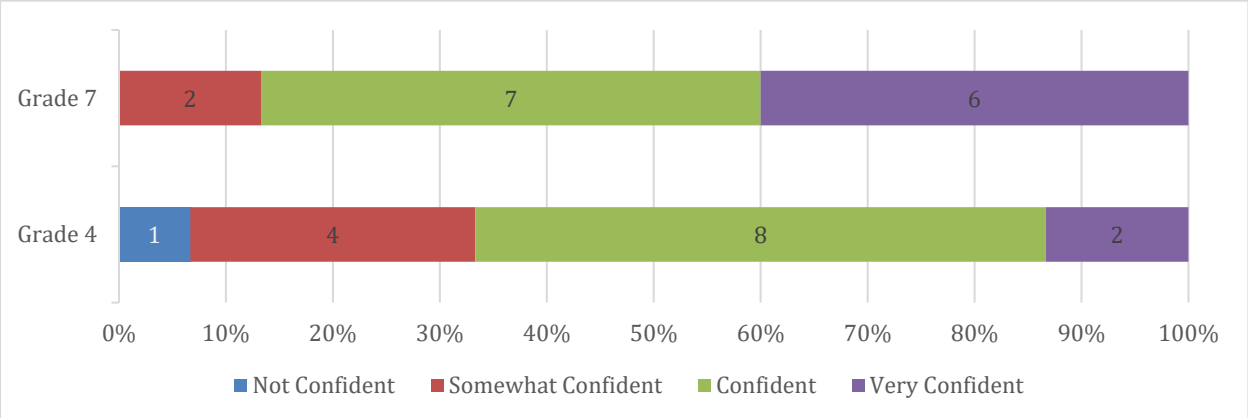
Apprentice



Proficient

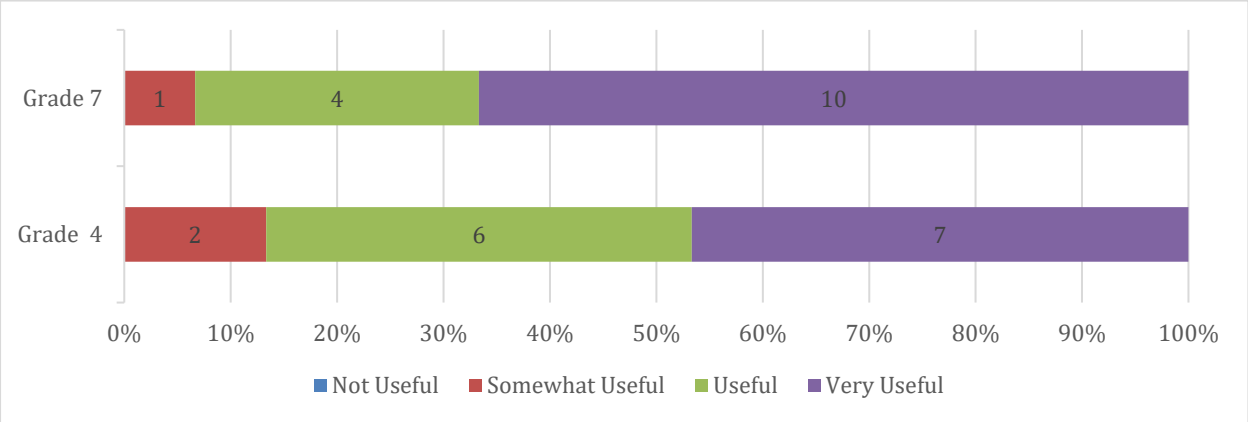


Distinguished

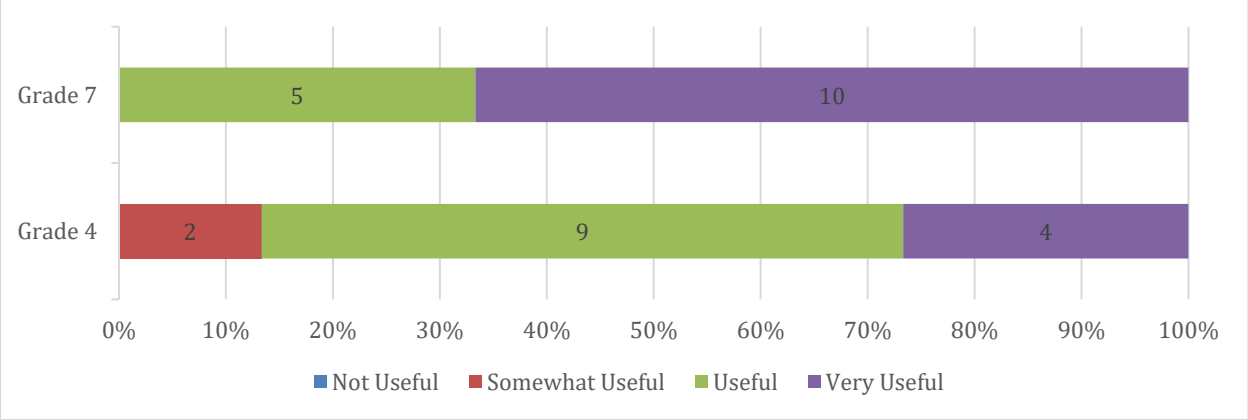


Question 5: How useful do you feel the following activities or information were in assisting you to make your recommendations?

Vertical Articulation Discussion

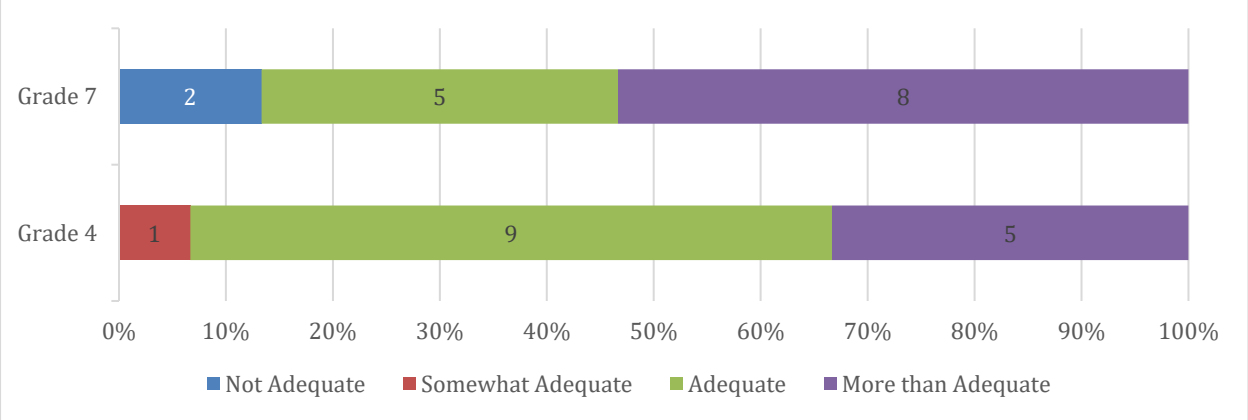


Training on PLD development

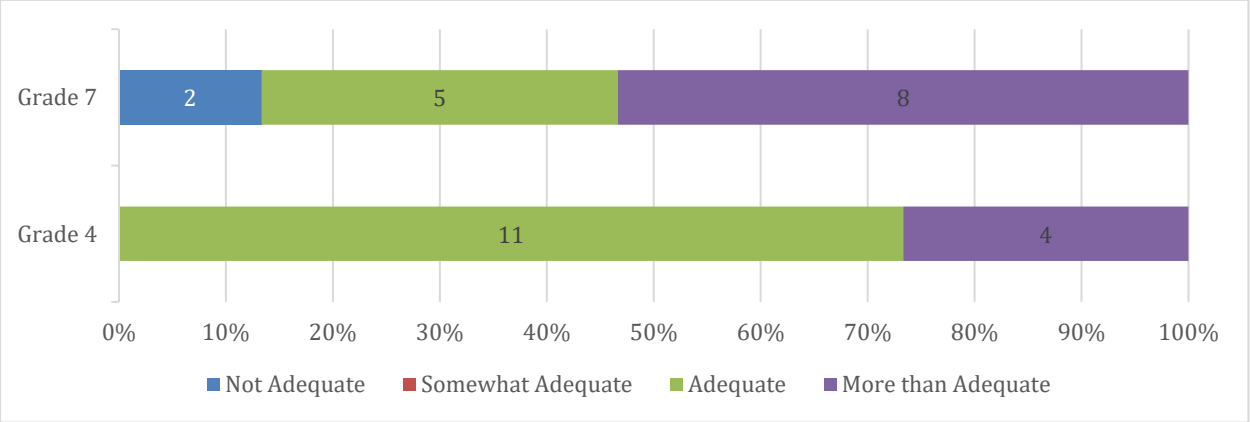


Question 6: How adequate were the following elements of the session?

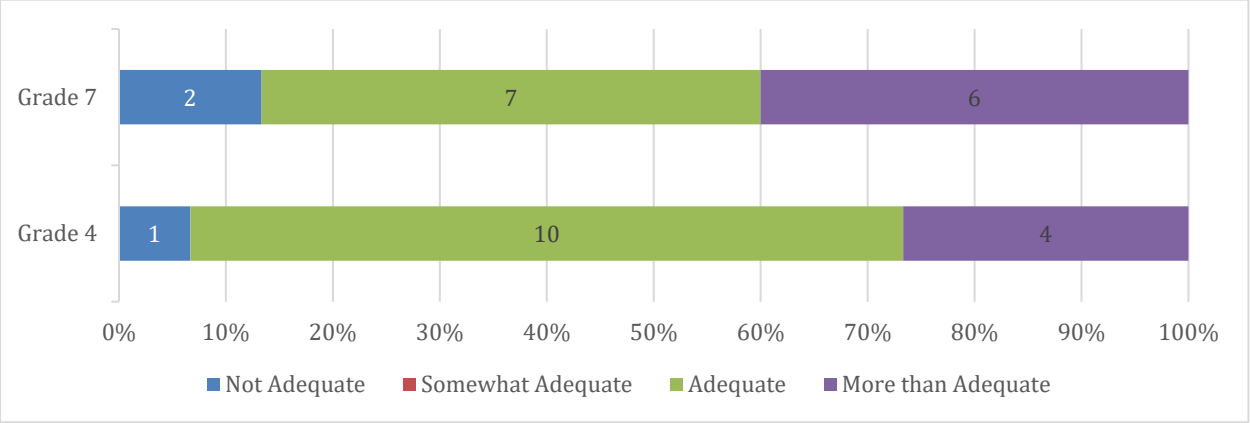
Visual presentation of the vertical articulation information



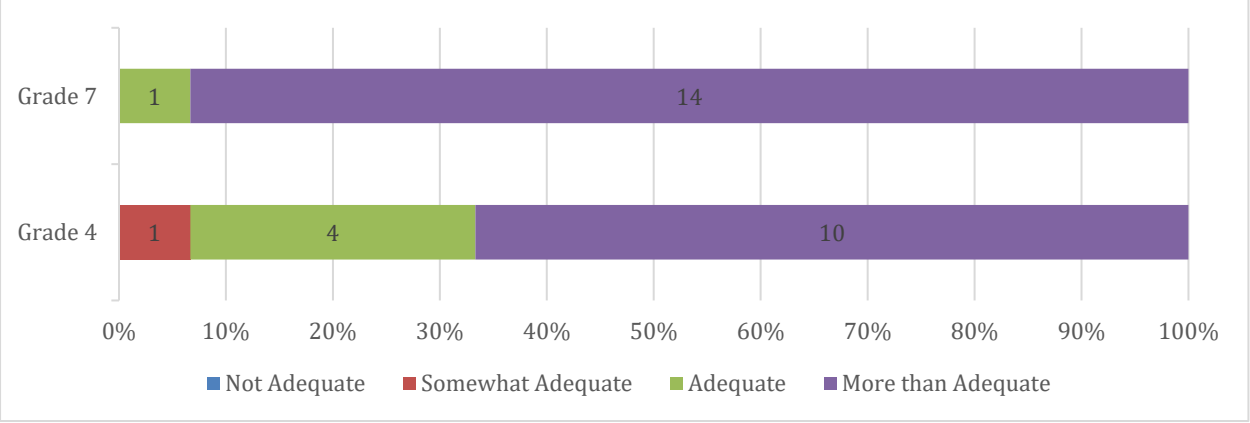
Amount of time table group discussion of PLDs



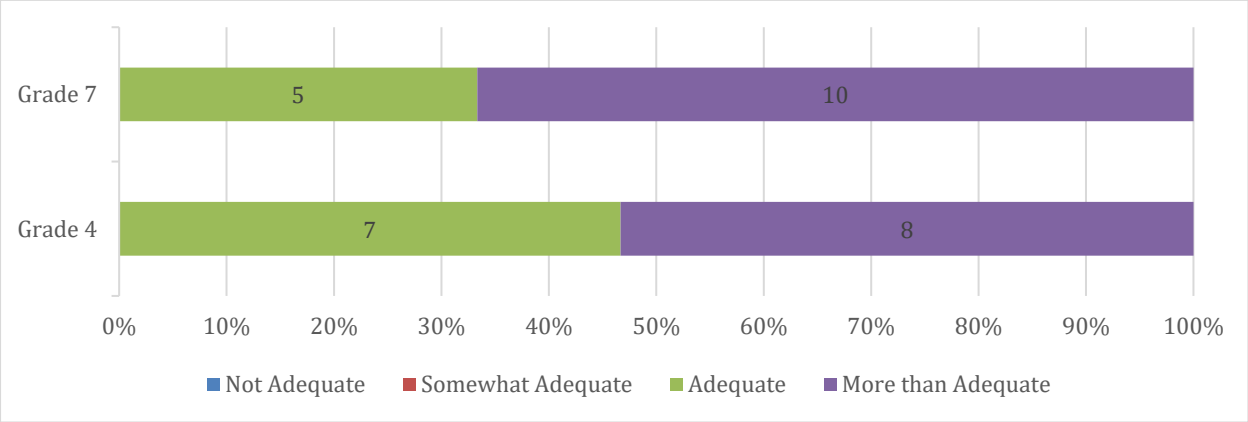
Amount of time whole group discussion of PLDs



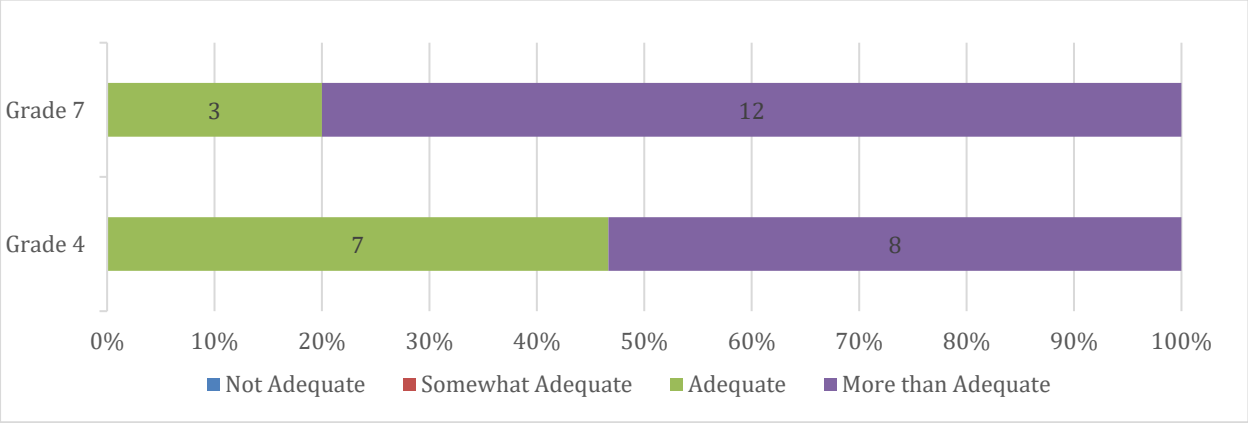
Facilities used for the meeting



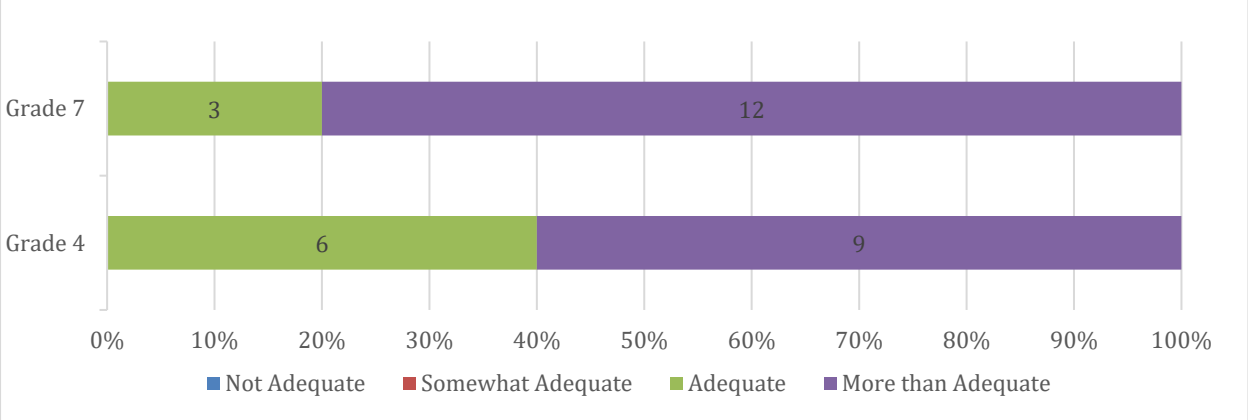
Computers used during the meeting



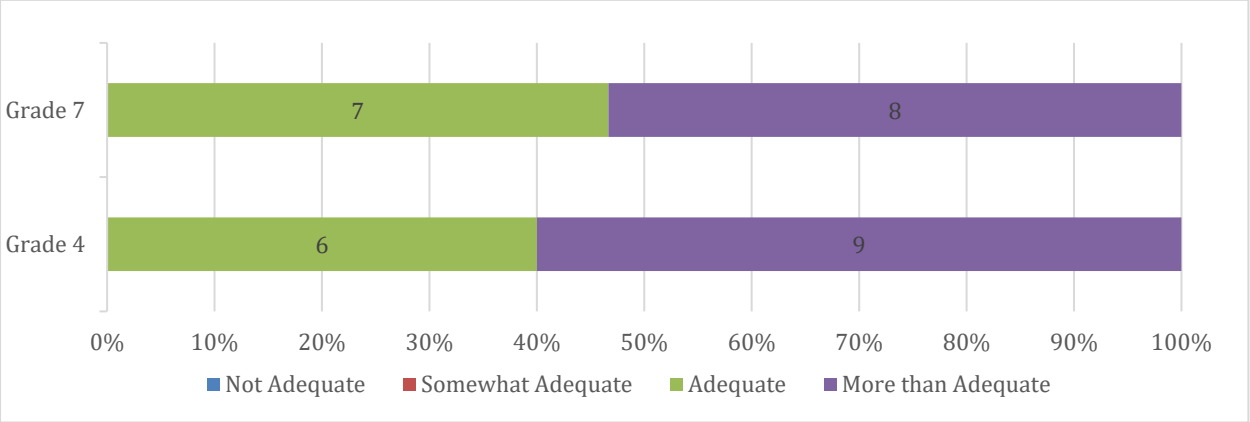
Pearson standard setting website for accessing materials and making judgments



Materials provided in the folder

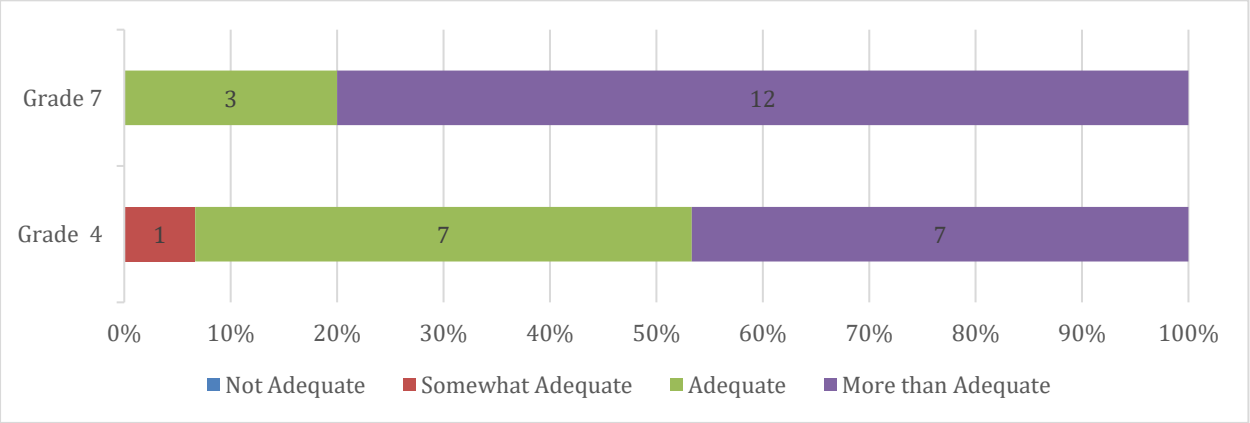


Work space in table groups during the meeting

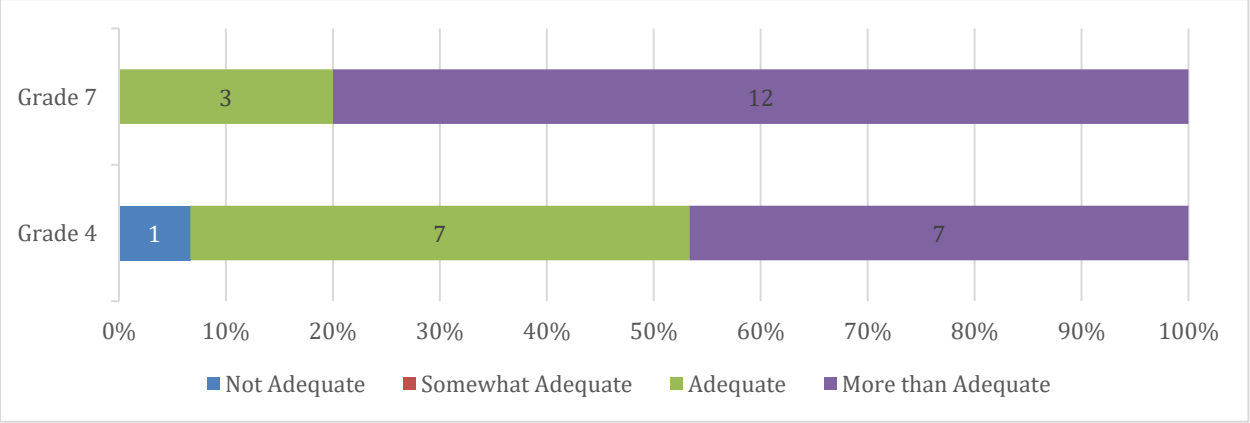


Question 7: Did you have adequate opportunities during the session to:

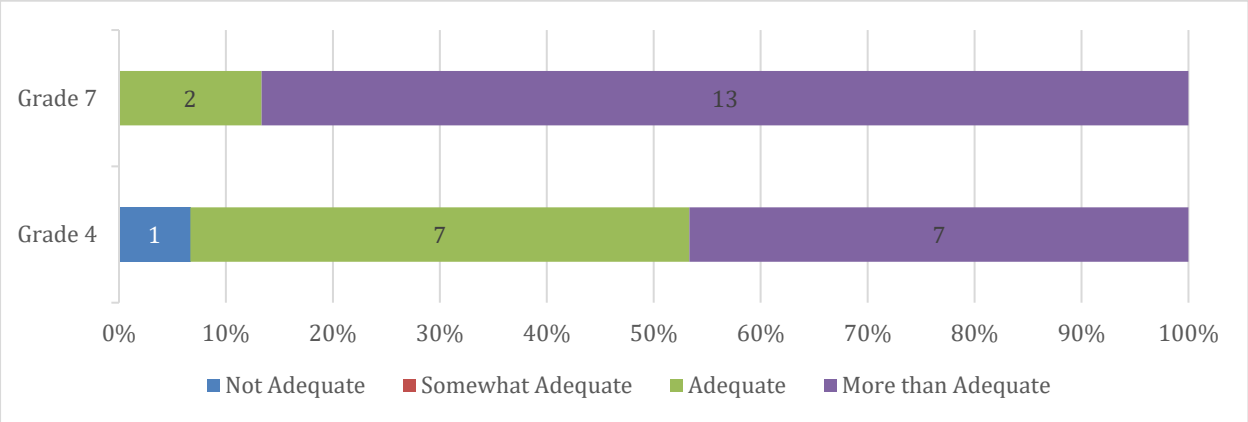
Express your opinions about student achievement levels



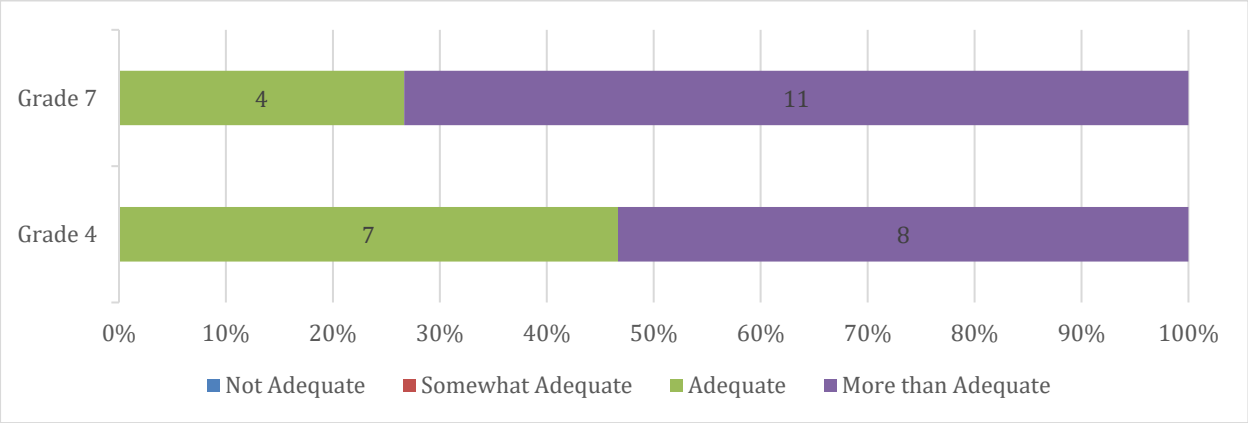
Ask question about the cut scores and how they will be used



Ask questions about the process of making cut score recommendations

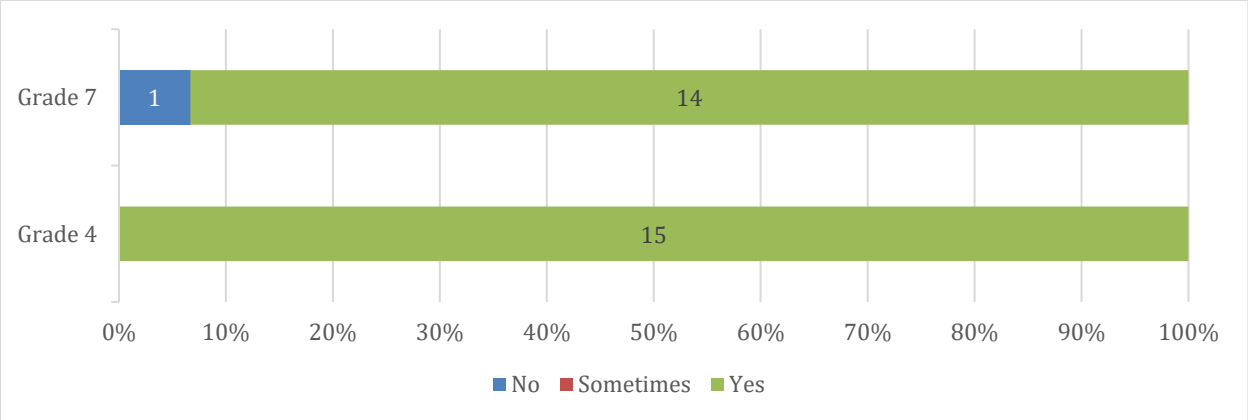


Interact with you fellow panelists

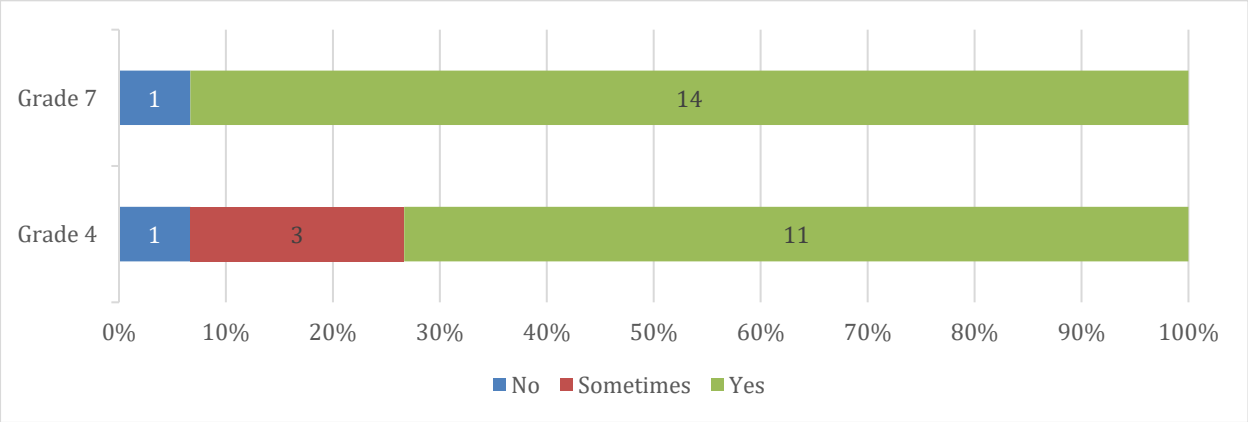


Question 8: Do you believe your opinions and judgments were treated with respect by:

Fellow panelists

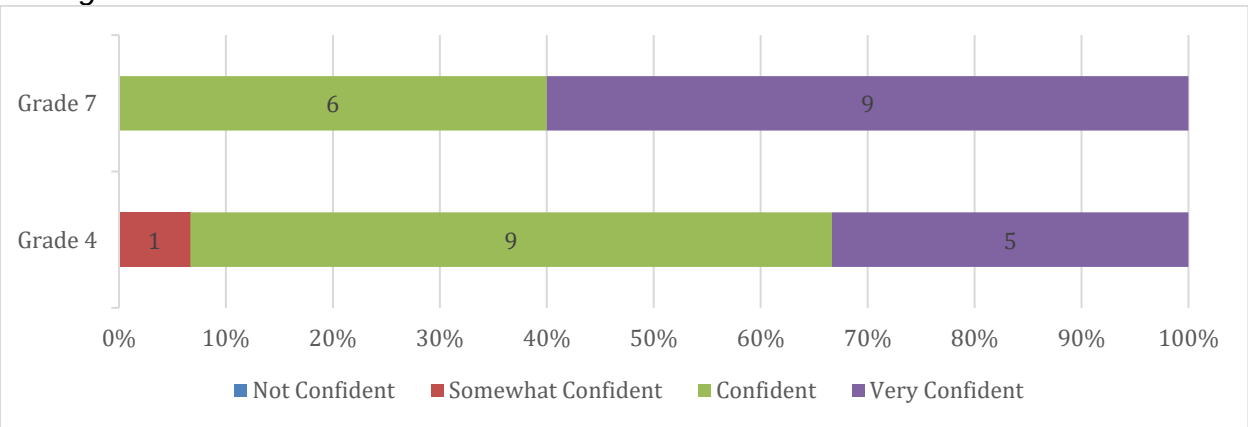


Facilitators



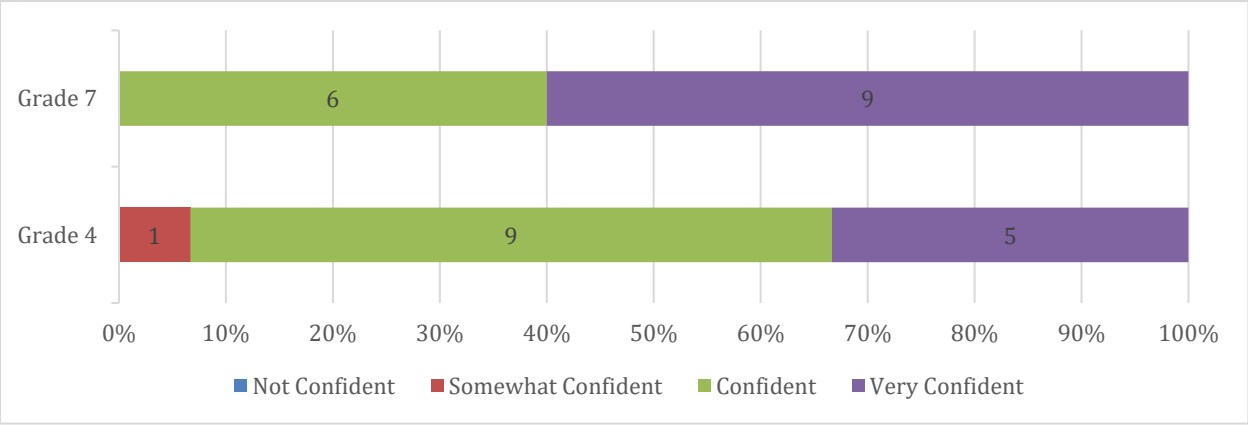
Question 9: How confident do you feel that the performance level descriptors (PLDs) that you developed for grade 4 & 7 science are reasonable for each student performance level?

Distinguished

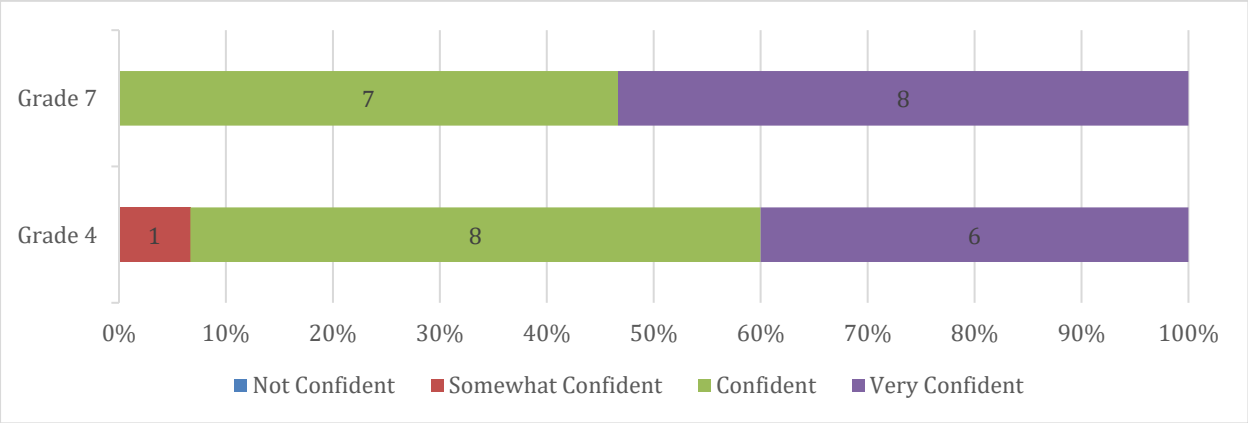


Proficient

Apprentice



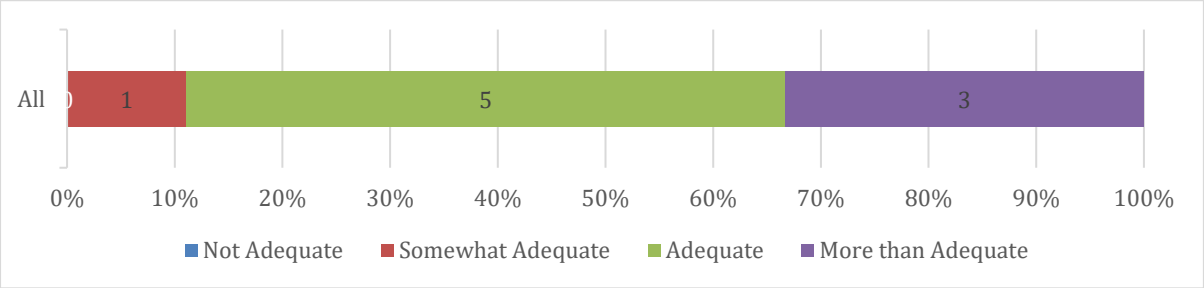
Novice



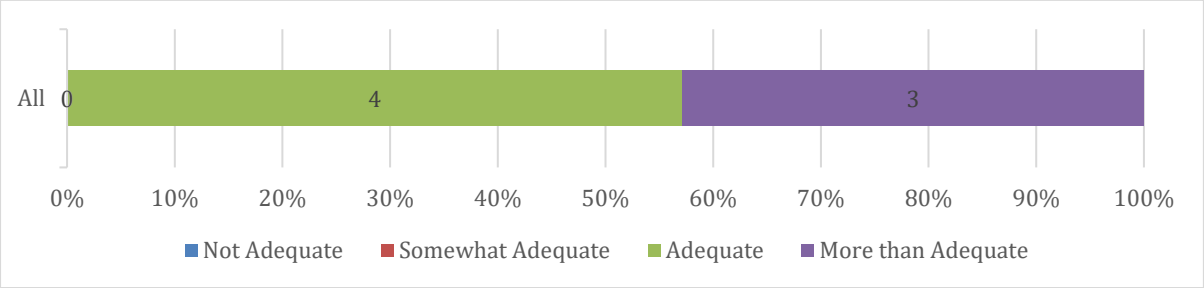
Follow-Up Meeting Process Evaluation

Question 1: How adequate were the following elements of the meeting?

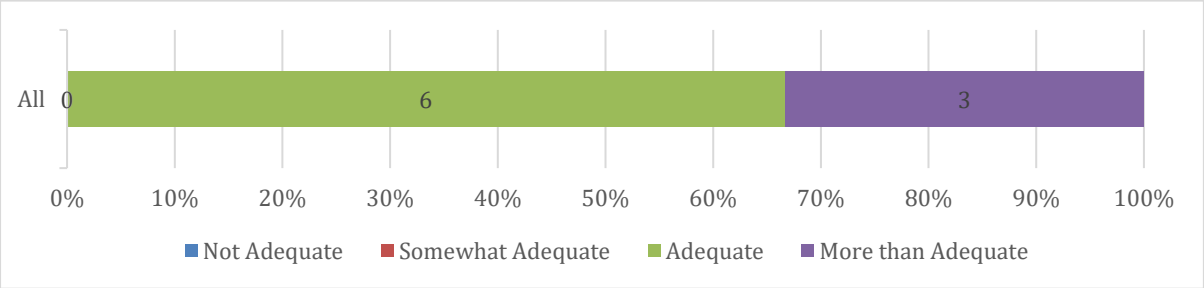
Description of the purpose of the meeting



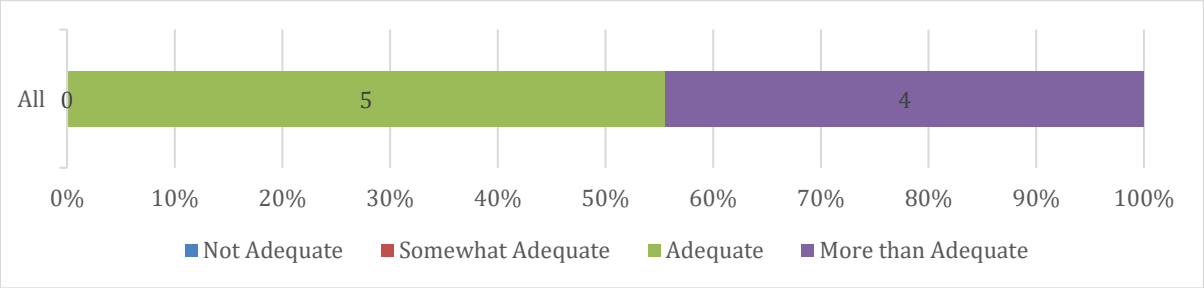
Review of the Standard Setting Recommendations



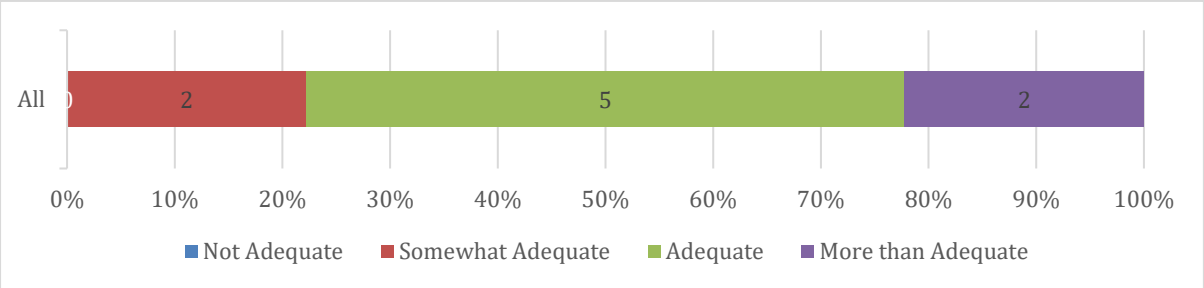
Presentation of the Additional Data



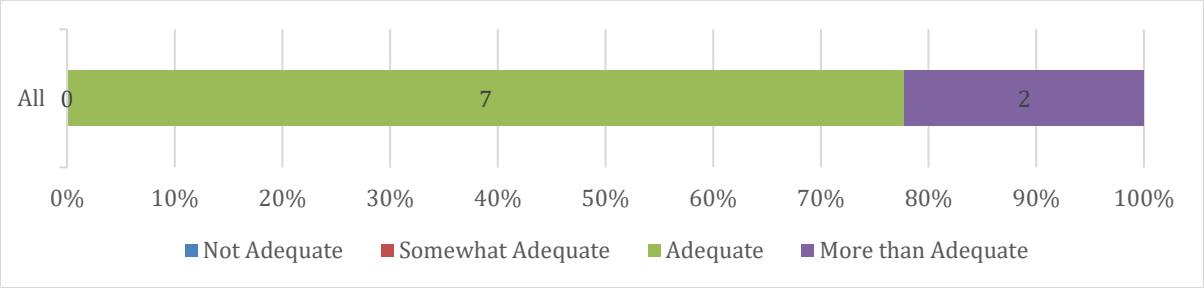
Interactive Spreadsheet



Discussion of Possible Changes to Cut Score Recommendations

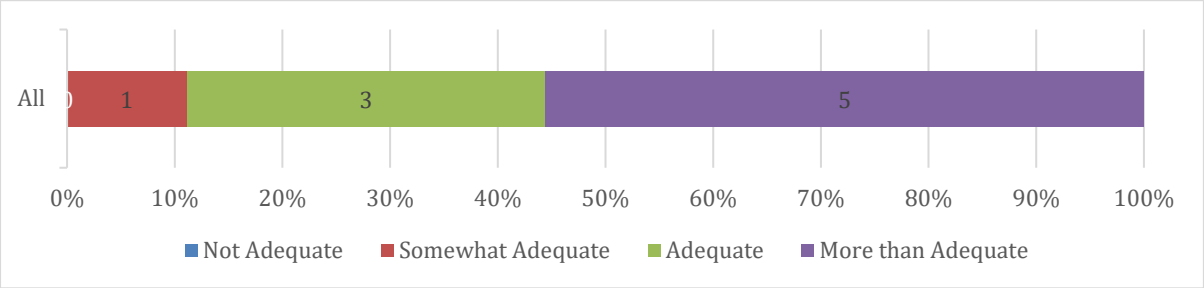


Use of Webex for Facilitating the Meeting

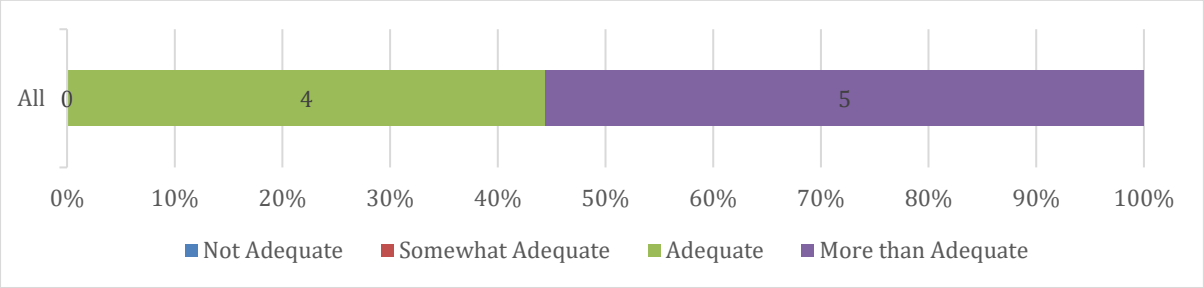


Question 2: Did you have opportunities during the session to:

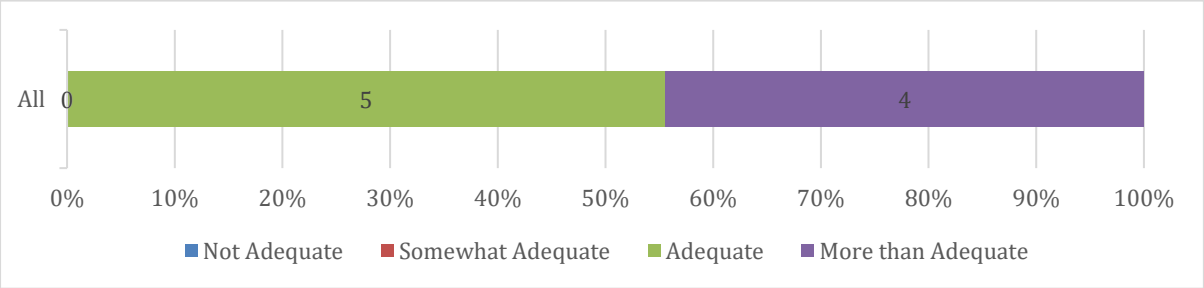
Express your opinions



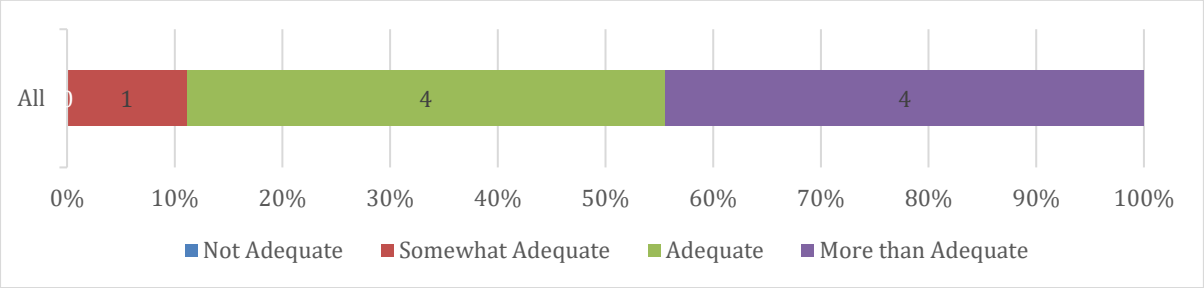
Ask questions about additional data



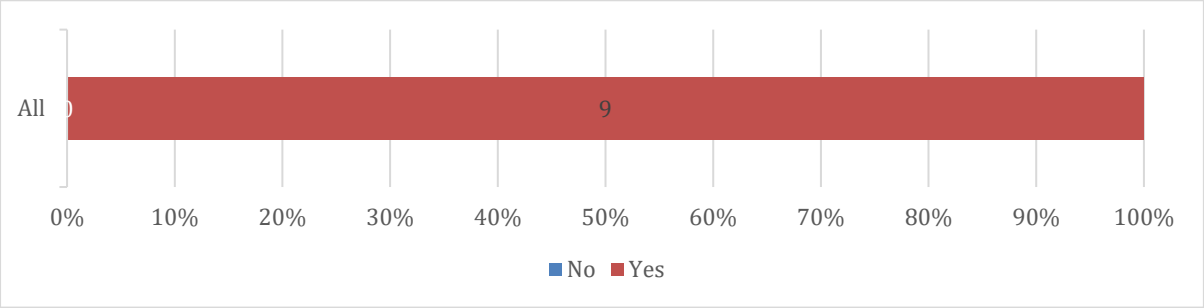
Recommend adjustments to the cut scores



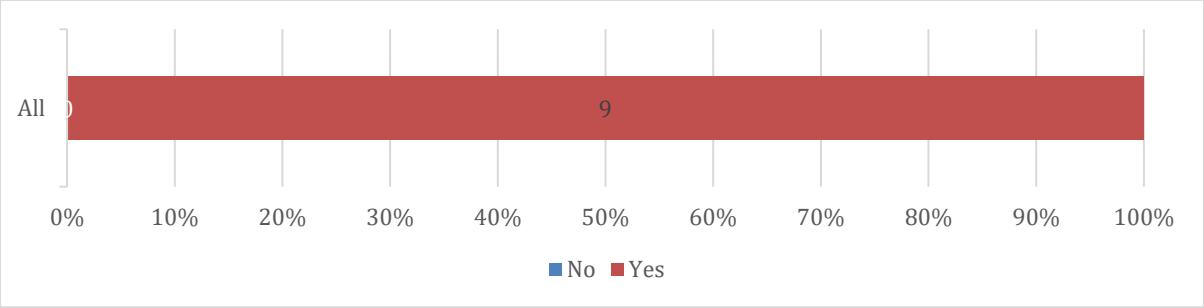
Discuss recommended changes to the cut scores



Question 3: Do you believe your opinions were treated with respect by the other participants?



Question 4: Do you believe your opinions were treated with respect by the facilitator?



Question 5: How confident do you feel that the performance levels defined by the adjusted cut scores are reasonable for each performance level?