

Oregon Extended Assessment
Technical Report on Standard Setting ORExt Science
ORExt Mathematics
ORExt English Language Arts

Submitted to the Oregon Department of Education

June 2015

by

DCE Educational Communications LLC

Executive Summary

In June of 2015, 53 content area and special education experts, representing three subject areas met over the course of three days and were guided through a judgmental decision-making workshop to set the cut scores for Oregon's 2015 Extended Assessments. Oregon's Extended Assessments are designed for students with the most significant cognitive disabilities and were recently revised to align to the state's Essentialized Assessment Frameworks. The Essentialized Assessment Frameworks provide a direct link to the English Language Arts and Mathematics Common Core State Standards, and Oregon's Science Standards as well as the Next Generation Science Standards in Science, for this population. A third-party, neutral observer was present to document and evaluate the proceedings to determine the validity of the resulting cut scores. The documentation that follows, details the logistical and statistical procedures undertaken in preparation for the workshop, describes the procedures followed during the workshop, and documents steps taken after the workshop toward finalizing the cut scores for use by this population. The results of the workshop are included in this document and the validity of the process is affirmed.

Table of Contents

Overview	4
Oregon’s Extended Assessments	5
Method.....	6
Initial Procedures.....	6
Workshop	8
Table 1.	9
Judgment Rounds	10
Table 5.	13
Materials	13
Panel Confidence.....	14
Table 6.	14
Articulation Round and Final Results.....	16
Articulation Round Guidelines.	16
Articulation Round Summary of Changes.....	16
Table 7.	17
Table 8.	17
Table 9.	18
Table 10.	19
Table 11.	20
Conclusion.....	20
References	23
Appendices A-O	24-139

Setting Achievement Standards for Oregon's Extended Assessments – 2015

Overview

In reference to the process of setting assessment cut scores, the American Educational Research Association (AERA, APA, & NCME, 2014) Standards for Educational and Psychological Testing suggests that

“if a judgmental standard setting process is followed, the method employed should be described clearly, and the precise nature and reliability of the judgments called for, should be presented... Documentation should also include the selections and qualifications of standard setting panel participants, training provided, any feedback to participants concerning the implications of their provisional judgments, and any opportunities for participants to confer with one another. Where applicable, variability over participants should be reported.” (p. 108).

In June of 2015, Behavioral Research and Teaching, developers of Oregon's alternate assessments, in collaboration with the Oregon Department of Education conducted a standard setting workshop in Eugene, Oregon, to determine the cut scores that would delineate the achievement categories for the population of students that takes Oregon's Extended Assessments. In addition to proposing cut scores, participants also reviewed and edited the associated Achievement Level Descriptors that provide qualitative descriptions of proficiency in each category. The workshop was conducted using the Bookmarking method of standard setting and was accomplished over the course of three days. Workshop participants recommended cut scores for the Oregon Extended Assessments in three subject areas: Science, Mathematics, and English Language Arts (ELA).

This document summarizes the main components of the standard setting process, and provides information related to the validity of the process in four areas: procedural consistency, internal consistency, panel membership, panel confidence.

The evaluation of procedural consistency examined whether a formal model of standard setting was implemented with integrity to an established procedure. The evaluation of internal consistency examined the function of the test items and the relationship between test items and the content standards (upon which achievement would be based). Panel membership and diversity was reviewed to ensure that the qualifications and perspective of the standard setting panel aligned with those necessary for the judgments required for standard setting. Finally, panelists were surveyed to determine their support of the process and their confidence in the outcomes -- including projected student impact. The cut scores generated from the standard

setting as well as the projected student impact of the cut scores (in terms of percentages of students falling into each of four achievement categories) are included in this review.

The complete document will be submitted to the Oregon Department of Education as part of a body of evidence documenting the validity of the Oregon Extended Assessment achievement standards.

Oregon's Extended Assessments

Oregon's alternate assessment, referred to as Oregon's Extended Assessment (ORExt), is designed to ensure that students in Oregon who have significant cognitive disabilities are exposed to critical, and appropriately stimulating academic content and are included in Oregon's educational accountability system. Oregon's Extended Assessments assess student performance in three subject areas via dichotomously-scored, selected response items that are administered by trained individuals. The assessments were originally developed in 2000 and have undergone at least 4 major revisions (as well as annual refinements) over their 15 years of use by the state of Oregon. The most recent assessments were revised in 2014 and field tested in 2015.

The three subject areas assessed by ORExt are as follows: (1) English Language Arts (ELA) which assesses both Reading and Writing and is taken in grades 3, 4, 5, 6, 7, 8, and 11. ORExt ELA assesses reading standards for literature, informational text, foundational skills, writing, and language, but excludes the assessment of speaking, listening, or literacy in history, social studies, science, and technical subjects. (2) ORExt Mathematics, which is taken in grades 3, 4, 5, 6, 7, 8, and 11 and assesses operations and algebraic thinking, number and operations in base ten, number and operations – fractions, measurement and data, and geometry in grades 3 – 5; ratios and proportional relationships, the number system, expressions and equations, geometry, and statistics and probability in grades 6 – 8, and number and quantity, algebra, functions, modeling, geometry, and statistics and probability in high school. (3) ORExt Science, which is taken in grades 5, 8, and 11 and assesses matter and its interactions, motion and stability: forces and interactions, energy, structure and processes of molecules and organisms, interaction, energy, and dynamics of ecosystems, Earth's place in the universe, Earth's systems, Earth and human activity, and engineering design (ODE, 2015).

Both ORExt ELA and ORExt Mathematics are linked to the Common Core Standards (CCSS) using the Essentialized Assessment Frameworks (EAFs). (The process of “essentializing” standards for students with the most significant cognitive disabilities will be described later in this document.) ORExt Science is linked to Next Generation Science Standards using the EAF. Currently in Oregon, a student with a significant cognitive disability may take the general assessment (with appropriate accessibility supports), the alternate assessment, or a combination of the two. Student eligibility for an alternate assessment is based on the IEP team's decision.

Method

Selection of standard setting method: Bookmarking. The Oregon Department of Education (ODE) in conjunction with Behavioral Research and Teaching (BRT) selected the Bookmarking method of standard setting to set standards for the newly revised ORExts. The Bookmarking Method of standard setting is consistent with the method used for the state's general assessment, and is the method previously used with the state's alternate assessment. The Bookmarking method of standard setting, though based on rigorous statistical procedures necessary to develop the Ordered Item Booklets, is a relatively simple procedure to implement with a large-scale state assessment, and is well-accepted among many states (Cizek, 2007). The bookmarking method is typically used with mixed responses items and vertically scaled items similar to those used in Oregon's tests.

Though there are certain variations to the Bookmarking process, the central process as described by Cizek in 2007 is as follows:

The task presented to participants in a Bookmark standard-setting procedure is straightforward. Using the [Ordered Item Booklet] assembled with one item (or score point) on each page, [panelists] are instructed to indicate the point at which they judge that the borderline or minimally qualified examinee's chances of answering the item correctly (or obtaining the score point) fall below the specified response probability or decision rule. For example, if a 2/3 decision rule is used, participants beginning to work through the OIB would ordinarily judge that the minimally qualified examinee would have better than a 2/3 likelihood of answering items at the beginning of the OIB (i.e., the easiest items) correctly. At some point in the OIB, however, participants would begin to discern that the chances of the minimally qualified examinee answering correctly approach and begin to drop below 2/3. Participants are instructed to indicate the point in the OIB at which the chances of the minimally qualified examinee answering correctly drop below 2/3. They indicate this judgment by placing a page marker—often a self-adhesive note or similar indicator—on the first page in the OIB at which the chance drops below the criterion. That is, the participants are indicating that the items prior to the marker represent content that the minimally qualified examinee would be expected to master at the [Response Probability] or decision rule specified.” (p.175).

Instructions for the full Bookmarking procedure that was followed by BRT and ODE in the June standard setting, are documented in Appendices A and B.

Initial Procedures

The newly developed ORExt in Science, Mathematics, and English Language Arts were developed in 2014 and field tested with students in the Spring of 2015. The revised assessments were updated to: assess students on the Essentialized Assessment Frameworks of the CCSS/ORSci/NGSS, support longitudinal growth models, improve administration, remove

administration functions that had become obsolete (such as the administration of the levels of support assessment), and improve general item functioning. A complete summary of the most recent changes to the assessment is included in Appendix C Summary of changes.

Oregon’s Essentialized Assessment Frameworks. As part of the development of the assessment, Oregon developed a set of alternate content standards based on the essential components of the Common Core State Standards, Oregon Science Standards, and Next Generation Science Standards. These alternate standards were developed to ensure that Oregon’s alternate assessment links to academic content. Almost 200 standards were distilled to under 50 essentialized standards. Each standard was analyzed and reduced to its essential core using a standardized process that is described in Oregon’s Extended Assessment administration manual as follows:

The standards have been “essentialized” by analyzing the content, the intellectual operation being requested, and the delimiters to the content. Structurally, this can be seen in the manner in which standards are written with the content identified by nouns, the intellectual operation by verbs, and the delimiters by either conditional phrases or as placed as the object of the sentence. The essentialization system uses the following conventions: (a) content (nouns) is boxed, (b) intellectual operations (verbs) are underlined (with complex verbs bold), and (c) delimiters (of content or intellectual operations) are italicized. Once the portions of the standard have been appropriately identified, the reduction in depth, breadth, and complexity (RDBC), which is explained below, follows.

The essentialization process involves [the reduction in depth, breadth, and complexity] of the Common Core State Standards (CCSS), Oregon's Science Standards, and the Next Generation Science Standards (NGSS) in order to establish a performance expectation that is relevant and accessible for students who participate in the ORExt, while maintaining the highest possible standards of rigor (the science tests will thus be dual-aligned to both the Oregon Science Standards and the NGSS). Complexity is reduced by: 1) focusing on essential content; 2) simplifying the process verb; and, 3) eliminating inappropriate delimiters. For the ORExt, all essentialized standards were written at three levels of complexity, which feeds the population of the Low, Medium, and High difficulty forms. The essentialized standards that will be assessed on the ORExt are called Essentialized Assessment Frameworks (EAFs) (ODE, 2015).

A flowchart of the standardized process of essentializing Oregon’s content standards is included in Appendix D.

Field testing. Items were operationally field tested with Oregon’s population of students with the most significant cognitive disabilities. Field testing was conducted in all three subject areas: Science (2,011 students), Mathematics (6,364 students), and English Language Arts (6,627

students). Almost six thousand (6,000) items were developed. Any items that failed to function as anticipated after scoring were eliminated from the item pool.

Ordered item booklet (OIB) development. Following field testing, item difficulty and student ability scores were calculated, using Item Response Theory procedures, in preparation for developing the ordered item booklets (OIBs). Student ability level on Oregon's alternate assessment differed by subject area. ORExt ELA student ability ranged from 1.91 (3rd grade) – 2.65 (7th grade) in consecutive grades, whereas ORExt Mathematics student ability ranged from .13 (3rd grade) to .78 (8th grade) in consecutive grades. In consecutive grades, mean item difficulty also varied from test to test. Mathematics mean item difficulty ranged from 0.7 (3rd grade) to 2.22 (8th grade). ELA mean item difficulty ranged from .93 (5th grade) to 2.14 (8th grade).

To develop the OIBs, items representing the full range of assessed items per grade were identified and then placed into booklets in their order of difficulty. The operational test taken by students was 48 items long in each subject area, however, Ordered Item Booklets constructed for the standard setting workshops ranged in length from 50 to 56 items. Appendix E includes more detailed information on OIB length and item difficulty across tests.

Selection of panelists. Each panelist was recruited by the Oregon Department of Education to play a specialized role as part of a subject-area group. Participants were recruited from among Oregon's licensed teachers throughout the year as well as from Oregon's Qualified Assessors (QAs) and Qualified Trainers (QTs) who are individuals trained in Oregon's Extended Assessments. Individuals were also recruited from among Oregon's Content Specialists who are educators who teach in Oregon and also serve the state in the development of educational materials. Panelists were asked to provide information on their: affiliation, degree, licensure, any certifications, and years of experience working with students with significant cognitive disabilities. Panelists were also asked to share their ethnicity and race.

Workshop

Panel Participants. A total of 53 panelists participated in the event. Eleven panelists in ORExt-Science, and 21 panelists were present each day for both ORExt-Mathematics and ORExt ELA.

The panel was highly educated. Over 90% of the panel possessed a Master's degree or higher. Fifty-seven (57%) percent of the panelists had over 11 years of teaching experience. Seventy-six percent (76%) of the panelists had some experience working with students with significant cognitive disabilities with 64% licensed as Special Educators. The panel was overwhelmingly female (87%), overwhelmingly from the Northwest of the state (87%), and overwhelmingly White (83%). No panel member self-identified with Oregon's major minority population (Hispanic). Panelist demographics collected at the workshop are compiled in Appendix F.

Structure of workshop. On each of the three workshop days a group of panelists met representing their specific subject-area. Each day's group of panelists had the same agenda and sequence of activities. One of the primary procedural differences among the three meetings was related to the number of grade levels assessed in that subject. Participants sat at tables in groups by their grade-level of expertise. On day one (ORExt Science), the three tables represented grades 5, 8, and 11. On days two and three (ORExt Mathematics, and ORExt ELA), the seven tables represented each of grades 3 - 8, and 11.

Across all three days/subject areas, each table had a similar configuration consisting of four to five individuals -- a table facilitator and three or four standard setting participants. The table facilitator was assigned by BRT to manage time and materials, keep the discussions focused, and to complete the rating sheets that captured the results from each of the decision-making rounds.

To ensure sufficient expert knowledge of the population, the subject area, the assessment and accountability in all decision-making groups, each grade level group in each subject area was required to be comprised of at least two special educators, and at least one subject-area specialist. The two Oregon special education teachers were present to ensure the panel's judgments included knowledge of the subject area, the population, and the scope and content of the assessment. The Oregon general education teacher(s) at each grade in each subject area was present to ensure the panel's judgments included subject area expertise, familiarity with the general education achievement expectations as they relate to Oregon's educational standards, the CCSS/ORSci/NGSS.

Training and process. Each day's session began with an overall training to ensure that participants understood their role in determining the state's alternate achievement standards, and the rationale for the day's activities. The training provided information on the development of the assessment, its framework, purpose and uses the training materials are included in Appendix G. Participants were provided with the appropriate subject-level materials and instructed on the standard setting procedure. Panelists were trained on the four levels of achievement. Proficiency Levels are referred to as Levels 1-4. Table 1 provides a general description of each performance category as it is currently used in Oregon.

Table 1.
Oregon Alternate Assessment Achievement Categories

Level	Description
1	Students demonstrate limited to no mastery of knowledge and skills related to essentialized standards that do not meet proficiency .
2	Students demonstrate inconsistent or partial mastery of knowledge and skills related to essentialized standards that do not meet proficiency .

- | | |
|---|--|
| 3 | Students demonstrate adept knowledge and skills related to essentialized standards that meet proficiency . |
| 4 | Students demonstrate adept mastery of knowledge and skills related to essentialized standards that exceed the requirements for proficiency . |
-

During training, panelists were instructed to place their bookmarks (sticky notes) on the first item of each category starting with the determination for Level 3 (the level in which a student is deemed minimally proficient). Panelists were next instructed to work on Level 4, the level in which a student is deemed to have exceeded the expectations of the population for the assessment. Finally, panelists were instructed to place their third marker on Level 2 to delineate the point at which a student who is not meeting even the minimal expectations, begins to demonstrate some understanding of the material.

Panelists were guided to place their marker on the first item that a minimally proficient student in that given level would have an 80% chance of getting right in the category. Panelists were asked to jot notes about what made the item they selected more difficult than the previous item. Following the standard setting, these (jotted) notes were used by BRT psychometricians when it was necessary to make articulation adjustments (the full process of post-standard-setting articulation is described later in this document).

Judgment Rounds

Judgment rounds 1 and 2. Discussions occurred in three rounds: an independent round, a consensus round, and a post-impact adjustment round. During the first (independent) round, individuals were asked to review their OIBs independently and to set all three level markers according to their knowledge of the population and the content of the items. During the second round, individuals discussed their round 1 findings with their grade and subject level colleagues at their table and discussed their findings and values to come to a shared conclusion about the placement of the cut points. In these discussions, individuals were required to support their judgments by providing content-driven explanations as to why the particular placement marked a delineation not only between two items, but between two categories. A sample of the types of discourse the individual engaged in is included in Table 2. Additional discussion points are included in Appendix H.

Table 2.*Panelist content considerations during judgment rounds*

Subject	Discussion
Science	Complexity of academic concepts. “Abstract concepts for this population are anything that they cannot experience through physical means, even a term like oxygen may be considered an abstract concept. Gravity, orbit, are all abstract concepts for this population [and render an item more difficult as a result]”.
Mathematics	Level of skill (academic verb) required by the item. “Up until this point there’s just a lot of point and matching and so on”.
Mathematics	Complexity of academic concepts: “Concepts change here. Now they have to know the concepts <i>same</i> , <i>more</i> , and <i>less</i> ”.
Mathematics	Level of skill (academic verb) required by the item. “Even with manipulatives, this item still requires a lot of accurate counting”.
Mathematics	Complexity of academic concepts. “At this point we are starting to talk about a student who could be taking the General Assessment, for example this item is about a clock, whereas this item requires in depth knowledge about fractions.”
ELA	<p>Experience with the item type or content. Individual A: “The length of item is very different from the previous, lots of extra information is provided. My students don’t know most of this information.”</p> <p>Individual B: “Yes, but look, the information that the question is based on is literally provided immediately before the question is asked.”</p>

Judgment round 3. Following the second round, BRT psychometricians calculated impact data for each of the groups to demonstrate the percentages of students that would fall into each of the four levels of achievement based on the cut points. For round 3, groups used this data to make any final adjustments to their cut points in the event that the percentages of students deemed proficient or not proficient were inconsistent across levels or indefensible. Groups were encouraged to maintain a content- and skill-driven discussion (similar to the discussion after round 2) to see whether their cut points would change. Panelists were warned not to use the impact data to simply place students into levels by percentages. Once panels made final (post-impact) changes they were shown a final round of impact data, however, no changes were made after round 3.

Documentation. Participants used different colored sticky notes for each of the rounds (green sticky notes for round 1, blue for round 2, and pink for round 3). Participants marked each sticky

note with the item number that represented the cut point for the performance category. Round 1 and 2 sticky notes were certified with the participant's initials, round 3 sticky notes were certified by the participant's signature. White sticky notes were also provided for participants to use as markers to indicate any general comments they may have made in the OIBs such as thoughts about items, item difficulty, or their decisions. At the end of the final round (round three), a representative from the group was designated as scribe and captured the group's rationale for each of the placed cut scores. These rationales are included in Appendix I. All judgments from independent reviews in round one, consensus reviews in round 2, and post-impact reviews in round 3 were collected by the table facilitators and are included in this report in Appendices J - L. Examples of data collected at each of the rounds for grade 5 Science is included here in Tables 3 – 5.

Table 3.

ORExt Science Grade 5 Rounds 1 – 3 Judgment Results Item (and Item Difficulty) by person/consensus

Grade 5 Science						
	Round 1				Round 2	Round 3
	Person 1	Person 2	Person 3	Person 4	Consensus	Consensus
Level 2	17 (0.986)	9 (0.556)	16 (0.926)	25 (1.536)	17 (0.986)	9 (0.556)
Level 3	21 (1.176)	17 (0.986)	30 (1.676)	31 (1.776)	29 (1.656)	29 (1.656)
Level 4	29 (1.656)	36 (1.956)	37 (2.006)	40 (2.306)	37 (2.006)	46 (2.956)

Table 4.

Grade 5: Impact following Round 2

Level	Percentage
1	30.9
2	8.6
3	6.5
4	53.9

Table 5.
Grade 5: Impact following Round 3

Level	Percentage
1	26.7
2	12.8
3	24.9
4	35.6

Materials

Panelist materials. Each group was provided the following materials:

- A copy of the standard setting procedure Appendices A and B,
- A copy of the training presentation (Appendix G),
- An Ordered Item Booklet (OIB) specific to their grade and subject area, (Individuals were not permitted to remove the OIBs, or the ALDs from the standard setting location.)
- An evaluation survey to share their confidence in the process,
- A background sheet on which they documented their demographic information,
- A copy of the essentialized frameworks, and
- A copy of the Achievement Level Descriptors (ALDs) was provided following the standard setting for the purpose of review and editing.

Ordered Item Booklets consisted of the secure items presented to students in 2014-2015, the language the administrator used to administer the item, the graphics and answer choices that a student was presented in relation to the item, and the correct score associated with the item. An image of the top of an OIB page (with the secure item removed) is shown below.

<i>Oregon Extended Assessment - Grade 3 English Language Arts - 2014-2015</i>					<i>Item Difficulty: -1.514</i>	
Item 1	Option:	A	B	C	Correct	Scoring (0/1)
1 - Here are three pictures. (Point to						

Achievement Level Descriptor review. Oregon's Achievement Level Descriptors (ALDs) were developed by educators at BRT with a panel of Oregon teachers, and approved by the state board of education, in May of 2015. Following the standard setting, standard setting panelists

were also asked to review the ALDs and to make any edits they deemed necessary. Panelists reviewed according to the following questions:

Is the language clear enough to communicate to parents?

Does the definition accurately capture a reasonable expectation for this population?

Is the expectation for this population a sufficiently appropriate parallel to expectations for students taking the general benchmark?

No major changes were made as a result of the review. Participants suggested three universal refinements. One such refinement was to alter and reduce the language at level 1 (the does not yet meet) category, to make it clear that Level 1 did not require, expect, or anticipate, any of the skills listed. A brief summary of the panels' suggested changes as shared to the group is included in Appendix M. Specific changes were noted in hardcopy and submitted to BRT and ODE for adjustments and re-submission to the State Board of Education.

Panel Confidence

Survey. At the end of each day's workshop, panelists completed a survey to capture their sentiments regarding the day's process and outcomes. Panelists were asked to respond to affirmative statements regarding the process and the outcomes and rate their agreement with the affirmative statements as Strongly Agree, Agree, Disagree, or Strongly Disagree. The 15 affirmative statements are listed in Table 6.

Table 6.
Affirmative Statements to Determine Panelist Confidence

Oregon Extended Assessment Standard Setter Evaluation Form - 2015
1. The orientation provided me with a clear understanding of the purpose of the standard setting meeting.
2. The training helped me understand the bookmark method and how to perform my role as a standard setter.
3. Reviewing the ORExt helped me to understand the assessment.
4. The small and large group discussions aided my understanding of the process.
5. There was an equal opportunity for everyone in my group to contribute his/her ideas and opinions.

-
6. I was able to follow instructions and complete the rating sheets accurately.
 7. The discussions after the first round of ratings were helpful to me.
 8. The discussions after the second round of ratings were helpful to me.
 9. The information showing the impact of our cut scores on proficiency percentages was helpful to me.
 10. I am confident about the defensibility and appropriateness of the final recommended cut scores.
 11. The achievement level descriptions were clear and useful.
 12. The time provided for discussions was adequate.
 13. The workshop leaders helped to answer questions and ensure that all input was respected and valued.
 14. The facilities and food service helped create a productive and efficient working environment.
 15. Overall, I am confident that the standard setting procedures allowed me to use my experience and expertise to recommend cut scores for the ORExt.
-

In Science, 100% of participants either strongly agreed or agreed with all 15 of the affirmative statements.

In Mathematics, 95% of participants either strongly agreed, or agreed with all 15 of the affirmative statements, 5% (1 individual) disagreed with statement 11, which read “The achievement level descriptions were clear and useful”.

In ELA 90% of participants either strongly agreed, or agreed with all 15 of the affirmative statements. One individual (5%) disagreed with statement 6, which read: “I was able to follow instructions and complete the rating sheets accurately”. One individual (5%) disagreed with statement 12 that read “The time provided for discussions was adequate.” This participant felt that too much time was provided.

Across all three subject areas, 100% of participants either Strongly agreed, or agreed with statement 10, which read, “I am confident about the defensibility and appropriateness of the final recommended cut scores.” Percentages of panel responses by subject area are included in Appendix N.

Articulation Round and Final Results

Articulation. The day following the standard setting workshop, psychometricians met to review the vertical alignment of the proposed cut scores across grades in the assessed subject area. Articulation is reviewed to make sure that, within each subject area of a vertically scaled test, the cut scores set at a given level for one grade do not exceed the cut scores set at the same level for the next grade. A smooth and intuitive progression is anticipated of the item difficulty in a given level as the grades increase. Of the cut scores set, 12 changes were made to maintain integrity across grades. Cut scores were adjusted in consecutive grades 3 – 8 in Mathematics and ELA.

When adjusting to maintain articulation integrity the following rules were followed to ensure that the fewest changes were made following the panelist's input overall:

Articulation Round Guidelines.

1. Identify the fewest number of steps necessary to bring the scores into articulation: Identify the scores that have the least cascading impact on other grades if changed. In reviewing alignment, isolate any scores (at any of the three cut scores levels) that appear to be outliers when compared to scores at other grades.
2. Follow the same order of adjustment as required by panelists: Start at the proficiency (Level 3) cut point, then evaluate Level 4 cut point, followed finally by the does not yet meet (Level 2) cut point.
3. Whenever possible, revert to a score that the panelists had considered previously with particular primacy to round two judgments (prior to their review following impact data): Reverting to round two was based in maintaining panelists' integrity. Panelists came to their round 2 conclusion based on their content review and only changed it in an attempt to influence the impact data if they found the impact data to be skewed.
4. Use booklets to confirm item changes: Whenever possible select the closest item to the panelist's original item selection while maintaining panelist rationale (which was often written in the booklet).
5. Only stray from the "closest item" rule (5 above) if the closest possible item contributes to creating a gap that further compromises the integrity of the articulation.

Articulation Round Summary of Changes

Science

No changes. Grades are not immediately consecutive and the scale was not vertical because of the gap between grades. In addition, the proportions (impact data) were not significantly different from ELA proportions overall.

Table 7.
Changes Made to Cut Scores in ORExt Mathematics

Grade	Level Adjusted	Previous Item Difficulty (item)	New Item Difficulty (item)	Shift in number of items
4	Level 1 – 2 (Nearly Meets)	-0.994 (5)	-0.734 (6)	1
4	Level 2 – 3 (Meets)	0.676 (25)	0.606 (21)	-4
4	Level 3 – 4 (Exceeds)	2.326 (48)	1.906 (42)	-6
5	Level 3 – 4 (Exceeds)	1.586 (35)	2.016 (41)	6
7	Level 1 – 2 (Nearly Meets)	-0.244 (6)	0.746 (18)	12
7	Level 3 – 4 (Exceeds)	2.776 (50)	2.276(43)	-7

Table 8.
Changes Made to Cut Scores in ORExt ELA

Grade	Level Adjusted	Previous Item Difficulty (Item)	New Item Difficulty (Item)	Shift in number of items
3	Level 3 – 4 (Exceeds)	3.006 (54)	2.776 (52)	2
4	Level 3 – 4 (Exceeds)	2.746 (45)	2.816 (46)	1
5	Level 1 – 2 (Nearly Meets)	0.516 (12)	0.166 (9)	-3
6	Level 2 – 3 (Meets)	1.666 (25)	2.036 (32)	7
6	Level 3 – 4 (Exceeds)	2.976 (45)	3.266 (49)	4
7	Level 1 – 2 (Nearly Meets)	0.386 (3)	0.776 (6)	6

Post Articulation Cut Scores. Tables 9 - 11 document the final cut scores and associated impact by level following the cross-grade articulation review. (Shaded cells are cells in which cut scores were changed from round 3.)

Table 9.*Science Post Articulation Final Recommended Cut Scores and Impact*

	Level 1	Level 2	Level 3	Level 4
Grade 5 cut point (item difficulty)		9 (0.556)	29 (1.656)	46 (2.956)
Grade 5 Impact	26.7%	12.8%	24.9%	35.6%
Grade 8 cut point (item difficulty)		19 (0.956)	36 (2.016)	51 (3.106)
Grade 8 Impact	28.8%	13.7%	15.2%	42.3%
Grade 11 cut point (item difficulty)		5 (0.106)	24 (1.406)	47 (2.856)
Grade 11 Impact	20.8%	10.8%	21.2%	47.2%
Mean Cross Grade Impact	25.43%	12.43%	20.43%	41.7%
SD of Impact	4.15	1.48	4.90	5.8

Table 10.*Mathematics Post Articulation Final Recommended Cut Scores and Impact*

	Level 1	Level 2	Level 3	Level 4
Grade 3 cut point (item difficulty)		6 (-0.764)	16 (0.136)	44 (1.816)
Grade 3 Impact	25.9%	13.9%	44.5%	15.7%
Grade 4 cut point (item difficulty)		6 (-0.734)	21 (0.606)	42 (1.906)
Grade 4 Impact	15.4%	30.5%	34.8%	19.3%
Grade 5 cut point (item difficulty)		8 (-0.664)	22 (0.616)	41 (2.016)
Grade 5 Impact	15.5%	25.6%	45%	14%
Grade 6 cut point (item difficulty)		6 (0.406)	13 (0.846)	37 (2.176)
Grade 6 Impact	32.1%	10.7%	39.1%	18.1%
Grade 7 cut point (item difficulty)		18 (0.746)	22 (0.916)	43 (2.276)
Grade 7 Impact	19.5%	25.3%	39.9%	15.4%
Grade 8 cut point (item difficulty)		5 (0.806)	18 (1.236)	35 (2.566)
Grade 8 Impact	41.9%	13%	38.5%	6.7%
Grade 11 cut point (item difficulty)		6 (0.136)	13 (0.656)	43 (2.206)
Grade 11 Impact	38.2%	11.9%	36.2%	13.8%
Mean Cross Grade Impact	26.93%	18.7%	39.71%	14.71%
SD of Impact	10.78	8.13	3.86	4.07

Table 11.*ELA Post Articulation Final Recommended Cut Scores and Impact*

	Level 1	Level 2	Level 3	Level 4
Grade 3 cut point (item difficulty)		5 (-0.764)	18 (1.316)	52 (2.776)
Grade 3 Impact	12.1%	23.4%	23%	41.5%
Grade 4 cut point (item difficulty)		8 (0.096)	23 (1.346)	46 (2.816)
Grade 4 Impact	15.2%	13.3%	23.6%	48%
Grade 5 cut point (item difficulty)		9 (0.166)	30 (2.006)	47 (3.246)
Grade 5 Impact	17.5%	16.2%	19.3%	47%
Grade 6 cut point (item difficulty)		5 (0.466)	32 (2.036)	49 (3.266)
Grade 6 Impact	19%	13%	23.1%	44.8%
Grade 7 cut point (item difficulty)		6 (0.776)	30 (2.226)	48 (3.636)
Grade 7 Impact	22.4%	12.8%	21.8%	43%
Grade 8 cut point (item difficulty)		5 (1.266)	18 (2.426)	50 (3.646)
Grade 8 Impact	27.3%	14.2%	24.1%	34.5%
Grade 11 cut point (item difficulty)		3 (-0.124)	35 (1.996)	48 (2.736)
Grade 11 Impact	19.5%	17.3%	11.8%	51.5%
Mean Cross Grade Impact	19%	15.74%	20.96%	44.33%
SD of Impact	4.92	3.78	4.34	5.46

Conclusion

Because a Bookmarking standard setting process is, at its heart, based on human judgments, no single piece of information can easily confirm the validity of the standards that result. To determine the validity of the cut scores from Oregon's 2015 standard setting workshop described in this document, a convergence of evidence model was used to evaluate the likelihood of valid outcomes from four perspectives: procedural consistency, internal consistency, panel

membership, and panel confidence in the results. Overall, the process undertaken in Oregon for the ORExt subject area assessments is likely to have resulted in valid outcomes due to soundness in the major procedural areas. Some minor deficits are noted in the summaries below.

Procedural consistency. Procedural consistency was evaluated by a review of: the methods used to set the standards, the integrity to which those responsible for the workshop adhered to the formal procedures, and the rationale used when diversions from formal procedure were necessary. The structure of the workshop, the quality and integrity of the training and materials, as well as the participants' adherence to training guidelines during rounds, contributed to strong procedural consistency of the workshop.

Internal consistency. Internal consistency was evaluated by a review of: the soundness of the initial procedures that went into the essentialization process, the soundness of the OIB development and IRT calculations, the scope of the field testing and associated scoring, and the soundness of the judgments used to guide the post-round articulation. While all internal procedures were carried out with fidelity to the statistical expectations of IRT, the range of item difficulty and student ability did not always fit the expected range of tests on an IRT scale. This likely contributed to some weakness in the internal consistency of the standards. However, the following consideration is an important one: In Oregon, the range of students eligible to take the ORExt is broad. Eligibility criteria currently is provided in the form of broad guidance for IEP team decision-makers and does not require empirical evidence of student ability as eligibility criteria for participation, see Appendix O (ODE, 2015). As a result, the population taking this assessment ranges from students who have difficulty interacting with items in any setting, to students who are close to being (but not quite) able to participate in the general assessment. This range of student skill level has an annual impact on item difficulty scores of Oregon's alternate assessment. Cut scores were made for this year's test with panelist knowledge that the tests (particularly ELA) would require additional, more difficult items in the coming years and that eligibility criteria for the assessment may be more stringent in future test populations.

Panel membership. Panel membership was evaluated by: a review of the diversity and expertise of the panel. As noted, the panels were highly educated with over 90% of the panel possessing a Master's degree or higher. The majority of the panel had had experience working with the population of students with significant disabilities, and while 64% had a special education license, decisions were balanced by the presence of general educators familiar with the expectations of the general population. The panel diversity was low, particularly racial/ethnic diversity, gender diversity and regional diversity. No panel member self-identified with Oregon's major minority population (Hispanic). However, the concentrations of educator gender, and regional representation aligned loosely with proportions of educators in the state. It is not clear how different the cut scores would have been if there had been greater racial diversity in the panel. The educational level of the panel and the quality of the training (with a focus on the specialized needs of the population of students with significant cognitive disabilities) may

mitigate any variance resulting from the panel, however, future panel membership would benefit from greater diversity.

Panel Confidence. Panel confidence was measured via survey following the final round of the decision-making. Panelists had an opportunity to discuss their rationale with colleagues, work toward consensus, and adjust decisions after a review of the impact data. Following the workshop, panelists had full confidence in the standards they had set for the population. Only three of the 53 panelists deviated from agreement to affirmative statements about the process. None of the 3 disagreements impacted the individuals' confidence in the outcome.

References

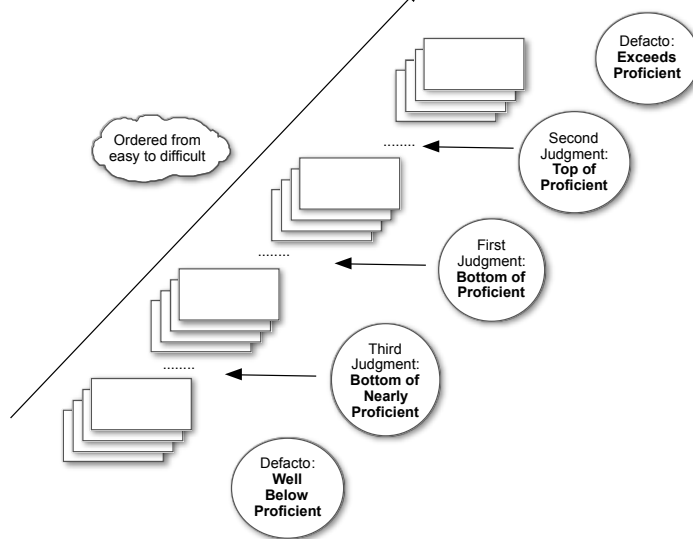
- AERA, APA, & NCME (2014) Standards for Educational and Psychological Testing.
Washington, DC: AERA.
- Cizek, G. J. (2007). The Bookmark Method. In G. J. Cizek & M.B. Bunch (Eds.), Standard
Setting: A Guide to Establishing and Evaluating Performance Standards on Tests.
Thousand Oaks, CA: Sage Publications.
- Oregon Department of Education (2015). Oregon Extended Assessment Administration Manual.
Retrieved from
http://www.ode.state.or.us/teachlearn/testing/admin/alt/ea/updates/orext_adminman_14-15.pdf

APPENDIX A

Standard Setting – Oregon Extended (Alternate Assessment)

In mid-June, performance standards for the extended assessment in grades 3-8 and 11 will be set in Science (June 15), English-Language Arts (June 16), and Mathematics (June 17). A bookmark method will be used. Teachers will be grouped into teams of 3-4 individuals at each grade level (and content area) with one of them designated a team leader. Each team will have 1-2 special education and 1-2 general education teachers. Each team will be given a handout that reflects the purpose (to organize performance scores into four groups consisting of Exceeds Proficient, Proficient, Nearly Proficient, and Well Below Proficient) and an agenda of times and activities for the day.

The day will begin with an overview of the test, its design and alignment with and coverage of the common core state standards. This overview will then focus on specific items and their technical specifications. Finally, teachers will be provided a description of the population for whom the test is designed. By mid-morning (10 AM), teachers will be given specific directions on the standard setting process and provided test booklets that will present each item with its corresponding standard and ordered on theta (item difficulty). They will be provided 90 minutes to make **individual judgments** of the item that corresponds to the bottom score for Proficient. Then they will select a top score for Proficient. All items above are de facto Exceeds Proficient. The third judgment will be the bottom score for Nearly Proficient. All items below this will be de facto Well Below Proficient. Teachers will complete a brief evaluation scale on the certainty of their judgments. During a working lunch, each team will share their individual results (in the same order as they made them) and both (a) discuss the reasons for making the judgment and (b) **reach consensus**. Score sheets from each team will identify the cut scores (items) identified individually and collectively. Each group will complete a brief evaluation of their certainty in sorting students into these groups. They will also develop a proficiency level descriptor that depicts the unifying conception of skills for that proficiency category.



All score demarcation-evaluation sheets will be collected as groups finish their judgments and the score value for each cut point will be entered into a master data base with all scores from the operational assessment. A report will be generated that reflects the percentages of students within each proficiency category, which will be shared with the group. They will be asked to make any **adjustments to their judgments** (in the same order as their individual judgments). These final score sheets will be used to articulate standards across grades in a post standard-

setting analysis; this step ensures that proficiency is smoothed out across grades 3-8 with no spurious changes. These standards will be forwarded for submission to the State Board.

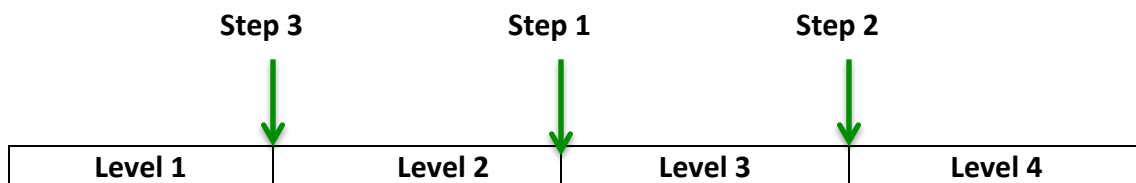
APPENDIX B
Oregon Extended Assessment – Standard Setting 2015
Preparatory Guidance and Resources

Thank you for agreeing to participate in the Oregon Department of Education's (ODE's) standard setting for the 2014-15 Oregon Extended Assessments (ORExt). The ORExt is Oregon's Alternate Assessment based on Alternate Achievement Standards (AA-AAS). You will be helping us define the AAS portion of that acronym, by setting cut scores and recommending Achievement Level Descriptors (ALDs).

These meetings will be held on the campus of the University of Oregon, Eugene, OR, in HEDCO Education Room 220. An interactive map is available at the following link: <http://map.uoregon.edu>. HEDCO is located in the southwest part of campus, in the lower-left hand side of the map. If you plan to pay for parking, we encourage you to find a spot in the parking lot right outside of the HEDCO building. Parking is \$1.25 per hour. The closest parking that is free requires a walk. You can park south of East 24th Street or west of Pearl Street if that is the option you prefer.

The standard setting meetings will be held on June 15 (Science), 16 (Mathematics), and 17 (English language arts), 2015. All meetings will begin promptly at 9:00 AM and run through 4:00 PM. Snacks, drinks, and a working lunch will be provided to all participants, as will all necessary materials.

We are looking forward to a series of smooth standard setting meetings, but making them run smoothly will require a bit of preparation on the part of all standard setters. We are providing this overview to orient you to your task. We will be using a bookmark method for our standard setting process and you will be working in grade level teams composed of two special education experts and one general education expert. You will receive ordered item booklets, where each item is rank ordered based on difficulty from low to high. The process will result in four performance levels, which requires that we select three cut points. Here is a graphic display of this for you, where performance levels are boxes in the table and the bookmarks, which eventually are transposed into cut scores, represented by the three green arrows:



The first step is to place a bookmark between the items that separate performance between Level 2 and Level 3. This decision separates those who are *proficient* from those who are *nearly proficient*. You'll ask yourself the following question, "Would a minimally proficient student have less than a 50% chance of responding to this item correctly?" We will then ask a similar question for differentiating between Levels 3 and 4

(Step 2), as well as Levels 1 and 2 (Step 3). There will be three rounds of these judgments. The first round you will complete independently. The second round you will perform as a group, aimed at building grade level group consensus. The third round will give you an opportunity to look at impact data and see if there are any needed adjustments.

In order to make appropriate judgments, you must be familiar with the student population, their levels of academic functioning, and the test format. As a final step, you will review the ALDs that correspond to the performance levels you set and make any recommended edits you deem needed.

We will focus on the topics mentioned above during the standard setting training, but we ask that you review the following resources prior to attending the meetings, unless you have done so already in your role as a Qualified Assessor or Qualified Trainer for the ORExt.

Please register for the ORExt Training and Proficiency website by logging into <https://or.k12test.com> and following the registration process (begin by selecting one of the three registration links and follow the instructions). Once registered, we ask you to review all of the information presented within the following Training sections on the website:

1. Updates for 2014-15
2. Rules for Administration (please watch all videos)
3. Administering the Content Prompts (please watch all videos)

Reviewing these resources should help you understand the student population, as well as how we approach AA-AAS here in Oregon. We expect that this preparation will take you approximately one hour to complete and is part of your expected compensation. If you are a real glutton for punishment and would like to know even more about the bookmark standard setting process, we recommend Chapter 10 of the Cizek & Bunch text, available for free at this link: http://www.sagepub.com/upm-data/13067_Chapter10.pdf.

Please contact Dan Farley at dfarley@uoregon.edu or 1-800-838-3163 if you have any questions about the or.k12test.com website or the content of this notice. Any questions regarding compensation can be directed toward Brad Lenhardt at ODE at brad.lenhardt@state.or.us or 503-947-5755. Thank you for your dedication to Oregon's students with significant cognitive disabilities. We are really looking forward to working with you in June!

APPENDIX C

Similarities and Differences Between the Former and New OREx

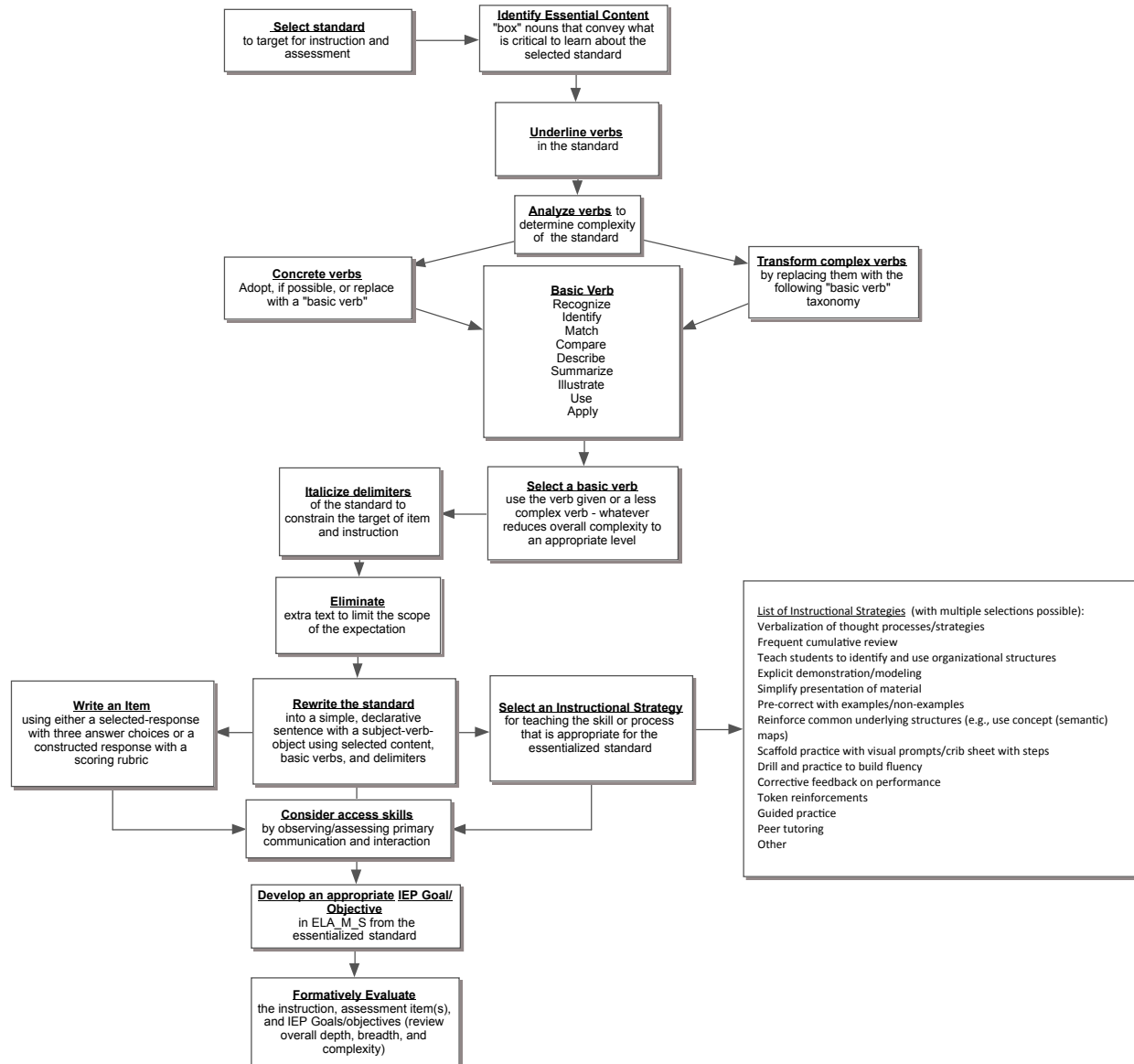
While much remains the same as we implement a new assessment, such as test security requirements, training requirements, and roles and responsibilities of test administrators, the test itself is changing a great deal. We have developed two tables that demonstrate how the approaches taken in the new assessment are the SAME as the former assessment, as well as areas in which the two assessments are DIFFERENT. The tables are provided below.

SAME for Former OREx and New OREx	
<ul style="list-style-type: none"> • Qualified Trainers (QTs) and Qualified Assessors (QAs) are trained annually in terms of test administration 	
<ul style="list-style-type: none"> • QTs and QAs must pass annual proficiency examinations 	
<ul style="list-style-type: none"> • Administered only by proficient QTs/QAs 	
<ul style="list-style-type: none"> • QTs and QAs must also receive test security training and sign a Test Administrator Assurance of Test Security form 	
<ul style="list-style-type: none"> • Ongoing training, materials, and proficiency testing provided online via or.k12test.com website 	
<ul style="list-style-type: none"> • Performance-based assessment tasks are being used with selection type items to ensure students have access for responding 	
<ul style="list-style-type: none"> • Results used for Annual Measurable Objectives (AMO) determinations 	
<ul style="list-style-type: none"> • Assessment available only to Students with Significant Cognitive Disabilities (SWSCDs) 	
<ul style="list-style-type: none"> • Available in Large Print & Braille (when ordered with sufficient notice) 	
<ul style="list-style-type: none"> • Two month test window (late February to late April) 	
<ul style="list-style-type: none"> • Test Materials include a Scoring Protocol booklet, with all administration directions and scoring documentation, as well as a Student Materials booklet, containing the item prompts and student answer choices for each item 	
<ul style="list-style-type: none"> • Students in grades 11-12 can participate in the OREx to satisfy Essential Skills requirements 	

DIFFERENT	
Former OREx	New OREx
<ul style="list-style-type: none"> • Content Area Grades or Grade Bands Tested Reading: 3-5, 6-8 & 11 Writing: Grade 11 Mathematics: 3, 4, 5, 6, 7, 8, & 11 Science: 5, 8, & 11 	<ul style="list-style-type: none"> • Content Area Grades Tested English language arts (combined Reading and Writing): 3, 4, 5, 6, 7, 8, & 11 Mathematics: 3, 4, 5, 6, 7, 8, & 11 Science: 5, 8, & 11
<ul style="list-style-type: none"> • Two versions: Standard /Scaffold administrations 	<ul style="list-style-type: none"> • One version, patterned after the Scaffold administration

DIFFERENT	
Former ORExt	New ORExt
<ul style="list-style-type: none"> • Linked to the Common Core State Standards in Reading, Writing, and Mathematics 	<ul style="list-style-type: none"> • Linked to the Common Core State Standards (CCSS) in English language arts (not including literacy across domains, nor speaking & listening standards), and Mathematics using Essentialized Assessment Frameworks
<ul style="list-style-type: none"> • Linked to the Oregon Science Standards in Science 	<ul style="list-style-type: none"> • Linked to the Oregon Science Standards and Next Generation Science Standards in Science using Essentialized Assessment Frameworks
<ul style="list-style-type: none"> • NOT vertically scaled across grades 	<ul style="list-style-type: none"> • Vertically scaled in ELA and Mathematics across Grades 3-8
<ul style="list-style-type: none"> • Accessibility options 	<ul style="list-style-type: none"> • Expanded accessibility options, including reading of answer choices for <u>all</u> items, even in ELA, unless the item is tied to the Reading Foundations domain
<ul style="list-style-type: none"> • Levels of Independence (LOI) score, resultant from the 10-item LOI test, determines the level of support that the Assessor can provide during test administration (based on 4-level, level of independence scale) 	<ul style="list-style-type: none"> • In 2016, Levels of Independence score (LOI), result from the 5-item LOI test and used for informational purposes only. The Assessor determines the level of support to provide on each item (based on 4-level, level of independence scale, along with a 4-level communication level scale). An option exists for more items to be administered to learn about the student's primary communication and response systems
<ul style="list-style-type: none"> • Five possible scoring options for each item (I = Inappropriate, D = Too Difficult, 0 = incorrect, 1 = partially correct, 2 = fully correct) 	<ul style="list-style-type: none"> • Two possible scoring options for each item (0 = incorrect, 1 = correct)
<ul style="list-style-type: none"> • Available in paper/pencil format only 	<ul style="list-style-type: none"> • Eventually available for administration with computer or tablet (e.g., iPad, electronic) and/or paper/pencil format

DIFFERENT	
Former ORExt	New ORExt
<ul style="list-style-type: none"> • No Placement Test 	<ul style="list-style-type: none"> • In 2016, a 15-item Placement Test is used to determine which of three forms a student should be administered: Low, Medium, or High difficulty
<ul style="list-style-type: none"> • No leveled test forms provided 	<ul style="list-style-type: none"> • In 2016, three different, leveled test forms for each grade level and content area assessed, with students taking either the Low, Medium, or High difficulty assessment based upon their Placement Test score; each leveled test will be composed of 25 items, for a total operational test of 40 items (15 Placement Test items + 25 content prompt items)
<ul style="list-style-type: none"> • No curricular resources provided 	<ul style="list-style-type: none"> • In 2016, Curriculum and Instruction resources will be provided
<ul style="list-style-type: none"> • No Individualized Education Program (IEP) development resources provided 	<ul style="list-style-type: none"> • In 2016, Individualized Education Program (IEP) development resources provided, including the development of goals and objectives that are linked to standards, as well as Present Levels of Academic and Functional Performance (PLAAFPs)



APPENDIX E

Table E-1.

Science OIB distribution

Grade	OIB Item count	Lowest Item Difficulty	Highest Item Difficulty	Mean Item Difficulty
Grade 5	50	-0.344	4.066	1.39
Grade 8	56	-0.424	3.866	1.39
Grade 11	54	-0.514	5.036	1.39

Table E-2.

Mathematics OIB distribution

Grade	OIB Item count	Lowest Item Difficulty	Highest Item Difficulty	Mean Item Difficulty
Grade 3	55	-1.574	2.946	0.7
Grade 4	54	-1.424	3.206	1.05
Grade 5	56	-1.634	3.866	1.57
Grade 6	56	0.046	4.426	1.82
Grade 7	55	-0.914	3.946	1.46
Grade 8	54	0.596	4.306	2.22
Grade 11	54	-1.514	3.166	1.37

Table E-3.
ELA OIB distribution

Grade (OIB item count)	OIB Item count	Lowest Item Difficulty	Highest Item Difficulty	Mean Item Difficulty
3	55	-1.514	3.336	1.24
4	50	-1.124	4.026	1.35
5	50	-1.054	4.066	.93
6	50	-0.134	3.396	1.63
7	50	-0.034	4.396	1.83
8	50	0.666	3.646	2.14
11	50	-0.624	2.986	1.39

APPENDIX F

Table F-1.

Total number of Participants by Gender

Note: A total of 53 individuals participated, however, one individual in the mathematics workshop did not return his/her background worksheet. Calculations in the following tables are based on total attendance of 53.

	Science (%)	Mathematics (%)	ELA (%)	Cross Workshop Total (%)
Male	0	3 (14%)	3 (14%)	6 (13%)
Female	11 (100%)	18 (86%)	18 (86%)	47 (87%)
Total	11	21	21	53

Table F-2.

*Participant Race**

	Science	Mathematics	ELA	Cross Workshop Total
White	11 (100%)	15 (71%)	18 (86%)	44 (83%)
Asian	0	2 (10%)	2 (10%)	4 (8%)
Native Hawaiian Pacific Islander	0	2 (10%)	0	2 (4%)
Two or more races	0	1 (4%)	0	1 (2%)
None listed	0	0	1 (4%)	1(2%)

(*Only races in attendance at the workshop are listed)

Table F-3.*Highest Degree Earned*

	Science	Mathematics	ELA	Cross Workshop Total
Ph.D.	1 (9%)	1 (5%)	0	2 (4%)
Masters	8 (73%)	19 (90%)	21 (100%)	48 (90%)
Bachelors	2 (18%)	0	0	2 (4%)

Table F-4.*Number of Years Working with Students with Significant Cognitive Disabilities*

	Science	Mathematics	ELA	Cross Workshop Total
0 or infrequently	4 (37%)	5 (24%)	2 (10%)	11 (21%)
1-3	0	1 (5%)	4 (19%)	5 (9%)
4-6	2 (18%)	2 (10%)	3 (14%)	7 (13%)
7-10	3 (27%)	4 (19%)	4 (19%)	11 (20%)
11-14	0	4 (19%)	8 (38%)	12 (23%)
15+	2 (18%)	4 (19%)	0	6 (11%)

Table F-5.*Number of Years Teaching*

	Science	Mathematics	ELA	Cross Workshop Total
1-3	2 (18%)	1 (5%)	0	3 (6%)
4-6	0	2 (10%)	3 (14%)	5 (9%)
7-10	2 (18%)	5 (24%)	7 (33%)	14 (26%)
11-14	2 (18%)	5 (24%)	4 (19%)	11 (21%)
15+ (Max 40, 28, 31 respectively)	5 (46%)	7 (33%)	7 (33%)	19 (36%)

Table F-6.*Participants' Geographical Region within Oregon*

	Science	Mathematics	ELA	Cross Workshop Total
Northwest	10 (91%)	19 (90%)	18 (86%)	47 (87%)
Northeast	1 (9%)	1 (5%)	1 (5%)	3 (6%)
Central		0	1 (5%)	1 (2%)
Southwest		1 (5%)	0	1 (2%)
Southeast		0	1 (5%)	1 (2%)

Table F-7.*Educational License of Participants*

	Science	Mathematics	ELA	Cross Workshop Total
Special Education License	6 (55%)	14 (66%)	14 (66%)	34 (64%)
Standard Teaching License	5 (45%)	6 (29%)	7 (33%)	18 (34%)

Setting Alternate Achievement Standards for Oregon's Extended Assessment: 2014-15

Science, Math, & ELA Standard Settings
June 15 -17, 2015

Oregon Department of Education
Behavioral Research and Teaching - University of Oregon

Welcome

9:00 – 9:45 AM

- Welcome!!
- Tell us about yourself
 - Name
 - District/school
 - Education experience
 - Interests



Handouts

- Housekeeping

- Agendas (light pink)
- Background Information (light green)
- Confidentiality Form

- Standard Setting Documents

- Essentializing standards decision tree
- Content Standards (Essentialized Standards)
- Achievement Level Descriptors
- Bookmarking Documents
 - Rating Sheets (light blue)
 - Standard Setter Evaluation form (light yellow)
- Ordered-item Booklets (OIBs)

Housekeeping

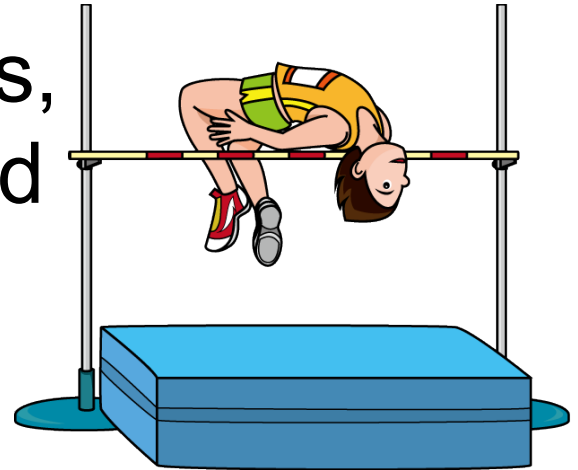
- Make sure that you signed in, please!
- Remuneration from ODE
 - Extended Assessment sub grants (EGMS)
- Confidentiality Form
- Background Info
 - Please record your degree attainment (e.g., B.A., M.A.T)
- Please silence cell phones
- Feel free to leave the room whenever needed, but please do so without disturbance

Orientation to The Task of Setting Standards



Purpose

- To determine the cut scores and achievement level descriptors that aptly define minimally proficient students, as well as those nearing and exceeding proficiency
- In other words, you will be setting the bar to which students will be compared



Outcomes

Quantitative

- Set cut scores that delineate which proficiency category best describes student performance at each level

Qualitative

- Establish achievement level descriptors (ALDs) for Oregon's Extended Assessment (ORExt) based on:
 - state content standards
 - the population assessed
 - the assessment in use (i.e. to determine the minimum expectations for students with significant cognitive disabilities on the state's accountability assessment – how good is good enough?)

Requirements for Your Role

- Minimum

- Knowledge of the population
- Knowledge of the assessment
- Knowledge about accountability

- Ideal

- Advocate for the population
- Advocate for the assessment
- Advocate for accountability

What is Not Needed

- Edits or feedback on the assessment content, scoring, or administration
- Judgments about the relevance of the assessment
- Judgments about the philosophy of accountability or the current statewide assessment system

Table Discussion Rules

- Listen actively and attentively.
- Ask for clarification if you are confused.
- Do not interrupt one another.
- Critique ideas, not people.
- Take responsibility for the quality of the discussion.
- Build on one another's comments; work toward shared understanding.
- Do not monopolize discussion.
- Speak from your own experience, without generalizing.
- If you are offended by anything said during discussion, acknowledge it immediately.
- Consider anything that is said at standard setting is strictly confidential.

Relevant Background Information



How did we get here?

9:45-10:30 AM

- Orientation to the student population
- Orientation to Alternate Assessments based on Alternate Achievement Standards (AA-AAS)
- Essentialization of content standards
- Item/Test Development Process
- Review of the ordered item booklets

Students with Significant Cognitive Disabilities (SWSCDs)

- National Survey Results – Student Attention

Description	%
Generally sustains attention for teacher-directed instruction	36.1
Demonstrates fleeting attention for teacher-directed instruction	52.8
Demonstrates little or no attention for teacher-directed instruction	10.9

Students with Significant Cognitive Disabilities (SWSCDs)

● National Survey Results - Mathematics

Description	ELEM Meets > 80% of the time	MIDDLE Meets > 80% of the time	HIGH Meets > 80% of the time
Sorts objects by common properties (e.g., shape, size, color)	53%	59%	63%
Adds or subtracts by joining or separating groups of objects	36%	44%	48%
Forms groups of objects for multiplication or division	5%	12%	17%
Multiplies and/or divides using numerals	4%	9%	13%

Students with Significant Cognitive Disabilities (SWSCDs)

● National Survey Results - ELA

Description	ELEM % who meet	MIDDLE % who meet	HIGH % who meet
Does not read any words when presented in print or Braille	22	19	18
Reads only a few words or up to pre-primer level	23	16	13
Primer to 1 st grade reading level	28	18	14
1 st grade to 2 nd grade reading level	17	19	15
Above 2 nd grade level to 3 rd grade level	8	18	21
Above 3 rd grade reading level	2	10	19

Video of Student Population of Oregon Extended Assessments

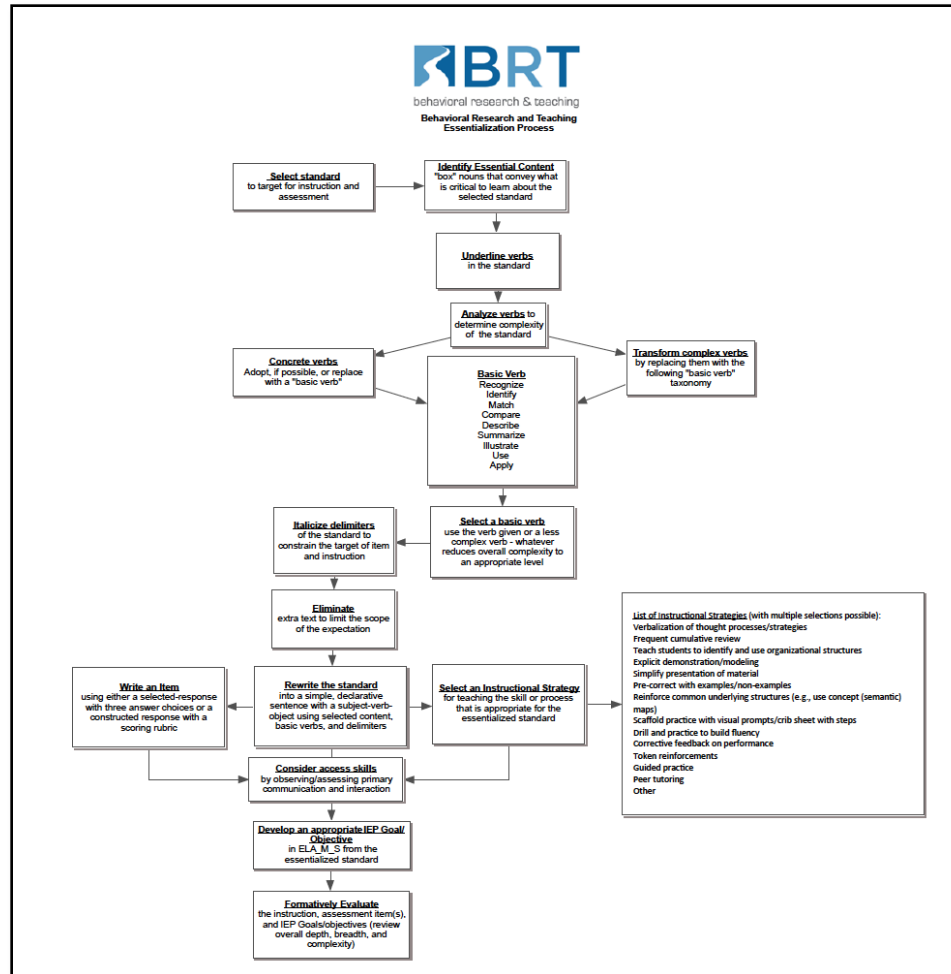


- The lowest functioning students in our schools
- Require intensive and pervasive support across all settings (e.g., home, school, community)
- Often require a full-time educational assistant for support at school
- SWSCDs are typically eligible for special education services due to Intellectual Disabilities, Multiple Disabilities, or Severe Autism

Essentializing Oregon's CCSS

- Select standard
- Code using essentialization system
- Reduce depth, breadth, and complexity by:
 - transforming complex verbs
 - limiting scope of content/verbs
 - eliminating extra text
- Generate the essentialized standard

Essentialization Flowchart



Save your eyes & check your handout

ORExt Items

- Items are linked to OR Content Standards by the Essentialized Standards
- Essentialized Standards are systematically reduced in:
 - Depth
 - Breadth
 - Complexity
(RDBC)

Item Development

- Content standard review (instructional priorities; test blueprint)
- Item writing (iterative process, including judgments from OR teachers)
 - Content review
 - Bias review
 - Alignment study
- Field testing
- Standard setting

What is the OR Extended Assessment?

- Extended Assessment = Oregon's alternate assessment based on alternate achievement standards (AA-AAS)
- The alternate assessment is a statewide accountability assessment designed for students with significant cognitive disabilities

What are Alternate Achievement Standards?

Cut scores



Achievement level descriptors

OR Statewide Assessment Options

- General Assessment (Smarter Balanced, OAKS)
 - With/Without accommodations
- ORExt
 - Embedded system of supports (level of support)
 - Universal design approach with multiple access options
 - Item difficulty rules of thumb
 - More content = more difficult
 - More challenging cognitive tasks = more difficult
 - More steps involved = more difficult
 - More prerequisite knowledge required = more difficult

Administration and Format

- Individually administered
- Substantially accommodated
- Flexible administration, using level of support that student requires to access item
- Items administered in standardized fashion
- Scoring is (0/1 = wrong/correct)
- One version
- Three levels of item difficulty
 - Low
 - Medium
 - High
- Universal design for assessment built in (e.g., low difficulty items have icons, simplified language, lower cognitive demand)

The ORExt Is

- RDBC; increased in terms of accessibility
- Designed to assess student academic knowledge and skills that are linked to grade level content standards
- Aligned to essentialized standards
- Administered in same grades as SBAC & OAKS
- Three content areas with grade-level administrations
 - English language arts (Grades 3, 4, 5, 6, 7, 8, & 11)
 - Mathematics (Grades 3, 4, 5, 6, 7, 8, & 11)
 - Science (Grades 5, 8, & 11)

What is the Purpose of the OR Extended Assessment?

- An accountability assessment is an indicator from states that informs the federal government whether or not students are being challenged with (and exposed to) critical content
- An accountability assessment holds states accountable for giving all students an opportunity to demonstrate their knowledge and skills

Four Performance Levels

- **Level 4:** similar to the former *Exceeds* level, for students whose performance is superior
- **Level 3:** similar to the former *Meets* level, for students who are consistently performing at expected levels
- **Level 2:** similar to the former *Nearly Meets* level, for students whose performance is not consistent enough to match proficiency expectations
- **Level 1:** similar to the former *Does Not Yet Meet* level, for students with extremely limited performances

What are the consequences of *Level 3* or *4* achievement?

- An IEP team will use a variety of information sources to make decisions for any student
- An IEP team may use success on the Extended Assessment **as part of a body of evidence** to inform the assessment decisions for the following year or to adjust instructional approaches for the student.
- A student who achieves a **Level 3 or 4** performance on an ORExt Assessment can count toward a school's Annual Measurable Objective (AMO) federal report for performance in a statewide assessment for that year and will provide the federal government with information about student success based on AA-AAS (1% Rule)

What are the consequences of *Level 1* or *2* achievement?

- Variety of information sources as **part of a body of evidence** to decide to:
 - Alter instruction to incorporate some of the content
 - Reassess the student in the coming year
 - Adjust instructional approaches for the student
 - Take no action and continue to provide the individualized instruction as they have done
- If the student **performs at Level 1 or 2** and took the minimum number of items required, the student may still count toward a school's AMO federal report for participation for statewide assessment for that year

Standard Setting



Educational Standard Setting

10:30-11:00 AM

- A process that allows a group of experts to make judgments regarding what a student should know in order to be a member of a given performance (achievement) category

Standard Setting Outcomes

- Quantitative value associated with minimal membership (Cut score)
- Qualitative definition of Achievement Level Descriptors (ALD) per category

What is the **Primary Question** When Setting Achievement Standards?

- How much does a student need to know in a given content area (e.g., Science) to be considered minimally competent?
 - What does that look like when represented quantitatively? (cut score)
 - What does that look like when described in words? (ALDs)

How are alternate achievement standards set?

- Variety of methods
- Variety of procedures
- Method and procedure are based on the nature of the data
- Bookmarking (Item mapping) process will be employed today

Bookmarking Standard Setting

- Items placed in order of difficulty using item response theory (IRT) calibration
- Using the order of difficulty suggested by these calibrated values, panelists mark the spot in the specially- constructed, ordered-item-booklet (OIB) to indicate where **the student just entering that category is expected to have an 80% change of responding to the item successfully**

ORExt Assessment Science

Outcomes: Participation (2013-14)

Oregon Students Total	Science Assessment				
	Grade 5 (42,649)	Grade 8 (43,522)	Grade 11 (42,633)	Total (128,804)	
Participating in alternate assessment against alternate standards	760 (1.8%)	642 (1.5%)	502 (1.2%)	1,904	1.5%

Source: Oregon Statewide Assessment data and <http://www.ode.state.or.us/search/page/?id=3225>.

Extended Assessment Science

Outcomes: Performance (2013-14)

Oregon Students who took AA-AAS	Science Assessment				
	Grade 5	Grade 8	Grade 11	Total (Number/Percent)	
Proficient or above in alternate assessment against alternate standards	433 (57%)	520 (81%)	136 (27%)	1,089	57%

Source: Oregon Statewide Assessment data and <http://www.ode.state.or.us/search/page/?id=3225>.

ORExt Assessment Math

Outcomes: Participation (2013-14)

Oregon Students Total	Mathematics Assessment									
	Grade 3 (42,649)	Grade 4 (42,858)	Grade 5 (42,752)	Grade 6 (42,449)	Grade 7 (43,202)	Grade 8 (43,522)	Grade 11 (42,633)	Total (300,065)		
Participating in alternate assessment against alternate standards	934 (2.2%)	944 (2.2%)	957 (2.2%)	889 (2.1%)	794 (1.8%)	740 (1.7%)	520 (1.2%)	5,778	1.9%	

Source: Oregon Statewide Assessment data and <http://www.ode.state.or.us/search/page/?id=3225>.

Extended Assessment Math

Outcomes: Performance (2013-14)

Oregon Students who took AA-AAS	Mathematics Assessment								
	Grade 3 (934)	Grade 4 (944)	Grade 5 (957)	Grade 6 (889)	Grade 7 (794)	Grade 8 (740)	Grade 11 (520)	Total (5,778)	
Proficient or above in alternate assessment against alternate standards	281 (30%)	241 (25.5%)	179 (18.7%)	84 (9.4%)	180 (22.7%)	172 (23.2%)	73 (14.0%)	1,210	20.9%

Source: Oregon Statewide Assessment data and <http://www.ode.state.or.us/search/page/?id=3225>.

ORExt Assessment Reading

Outcomes: Participation (2013-14)

Oregon Students Total	Reading Assessment									
	Grade 3 (42,649)	Grade 4 (42,858)	Grade 5 (42,752)	Grade 6 (42,449)	Grade 7 (43,202)	Grade 8 (43,522)	Grade 11 (42,633)	Total (300,065)		
Participating in alternate assessment against alternate standards	1,153 (2.7%)	1,088 (2.5%)	1,043 (2.4%)	888 (2.1%)	782 (1.8%)	681 (1.6%)	539 (1.3%)	6,174	2.1%	

Source: Oregon Statewide Assessment data and <http://www.ode.state.or.us/search/page/?id=3225>.

Extended Assessment Reading

Outcomes: Performance (2013-14)

Oregon Students who took AA-AAS	Reading Assessment									
	Grade 3 (1,153)	Grade 4 (1,088)	Grade 5 (1,043)	Grade 6 (888)	Grade 7 (782)	Grade 8 (681)	Grade 11 (539)	Total (6,174)		
Proficient or above in alternate assessment against alternate standards	885 (76.8%)	765 (70.3%)	728 (69.8%)	457 (51.2%)	546 (69.8%)	398 (58.4%)	336 (62.3%)	4,115	66.7%	

Source: Oregon Statewide Assessment data and <http://www.ode.state.or.us/search/page/?id=3225>.

Setting Cut Scores



General Process

- Three judgment rounds per grade level
 - Round 1: Individual Judgments
 - Round 2: Consensus building
 - Round 3: Evaluation of outcomes with impact data

Process

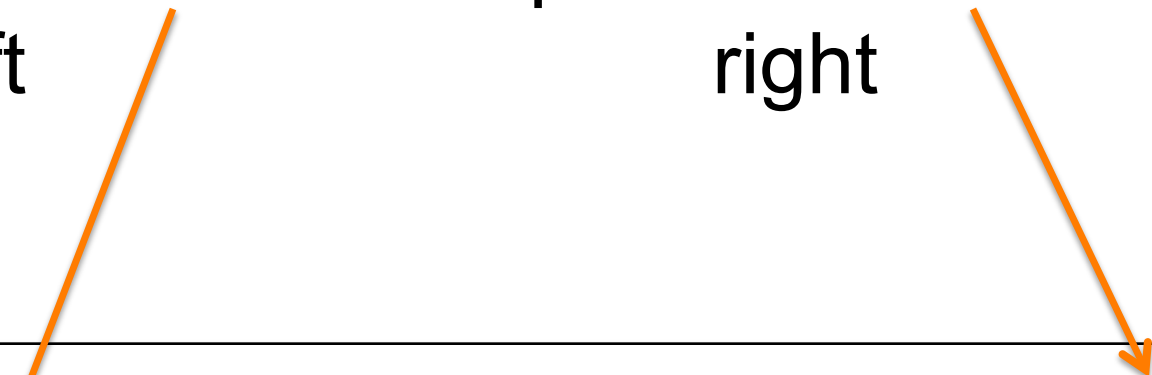
- All portions of the standard setting will be conducted in grade-level groups
- Each Group has a Table Facilitator
- Standard setters (Oregon Teachers)
 - Two special educators
 - One general educator
- BRT Table Facilitator
 - Manages time and materials
 - Keeps discussion focused
 - Takes notes

Materials

- Grade level Oregon Essentialized Standards packets
- Grade level ordered-item booklets:
 - Scoring rubrics are within the item text
 - Item difficulties are recorded on each page
- Individual rating sheet
 - Write observations regarding item difficulty
 - Record the three items that separate the four performance levels
- Overall process evaluation sheets

Structure of the OIBs

- Easiest item in front
- Item numbers top left
- Most difficult item at the end
- Item difficulty top right




Oregon Extended Assessment - Grade 3 English Language Arts - 2014-2015				Item Difficulty: -1.514	
Item 1	Option:	A	B	C	Correct Scoring (0/1)
1 - Here are three pictures. (Point to					

Items in Booklet

- Booklets contain all information that teachers used at the top – *Scoring Protocol*
- Booklets contain all information that was presented to students in the middle and bottom – *Student Materials*

Bookmarking Expectations

- Color-coded
 - Round 1= **Green** post-its
 - Round 2= **Blue** post-its
 - Round 3= **Pink** post-its
- Write the item number and your initials on the post-it, so there can be no confusion regarding which item you intend to mark (booklets are 2-sided)



Please sign the pink post-it note, which is your final recommendation.

Round 1: Item level considerations

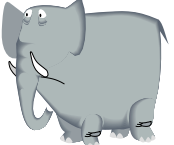
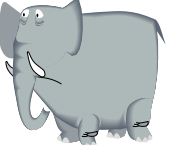
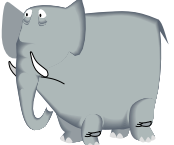
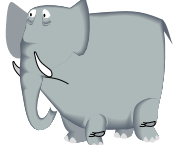
- ***Individual*** judgments
 - What makes this item more difficult than the one before it? Capture this information in summary for use in Round 2.
 - What knowledge, skills, and abilities must be applied correctly to respond to this item?
 - Record the item numbers on your blue rating sheets throughout Rounds 1, 2, and 3

Round 1: Process

11:00-12:00 PM

- Panelists work **independently** to determine the location for the three items that separate the four categories of performance
 - Please do not discuss item difficulty
 - Procedural questions will be answered
- Place 3 post-it notes to represent four categories
- Use **green** post-it notes to mark the location, record the item number, and initial the post- it note
- **Categories**
 - Level 4
 - Level 3
 - Level 2
 - Level 1

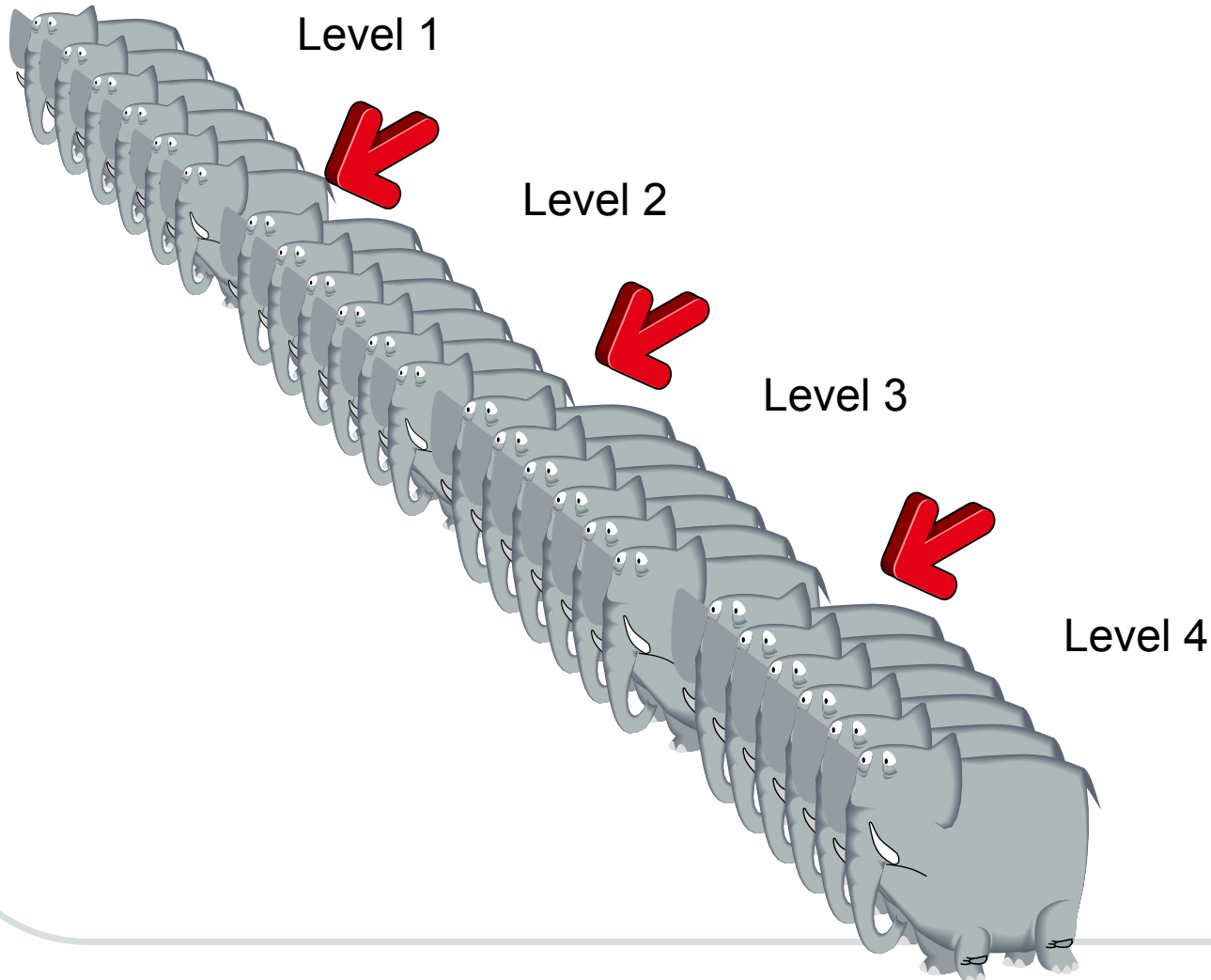
Relationship Between Categories and Cut scores

Level 1	Level 2	Level 3	Level 4
Extremely limited to no performance of knowledge and skills	Inconsistent performance of knowledge and skills	Consistent performance of knowledge and skills	Superior performance of knowledge and skills
			

Meaning of a Cut Score

- Items at the bookmark indicate that students have mastery of all previous items (likely to know all the correct responses) and therefore meet the minimum requirements of category membership
- Remember, individuals within a category will display a range of scores

Placing the Bookmarks



Bookmarking Decision Rules

- Place your first post-it on the item that you believe a **student just entering the proficient category** has an 80% chance to answer correctly.
- Place your second post it on the item that you believe a **student just entering the superior category** has an 80% chance to answer correctly.
- Place your third post-it on the item that you believe **student who is just entering the nearing proficiency category** has an 80% chance to answer correctly.

Round 2: Group Consensus

12:00-1:00 PM

- Table leader consolidates scores on Excel spreadsheet and discusses range of values with participants
- Consider only the range of possibilities suggested by the group and discuss the possible outcomes based on a definition of the category label
- Use your descriptions of what makes a score more difficult than the preceding score to assist with the decision-making
- Make new bookmark selections that capture your new judgment
- Use **blue** post-it notes to mark the location, record the item number, and initial the post- it note

The Scale of the test (hypothetical)

Students by Ability

300 304 309 450



Items by Difficulty

1

2

3

...

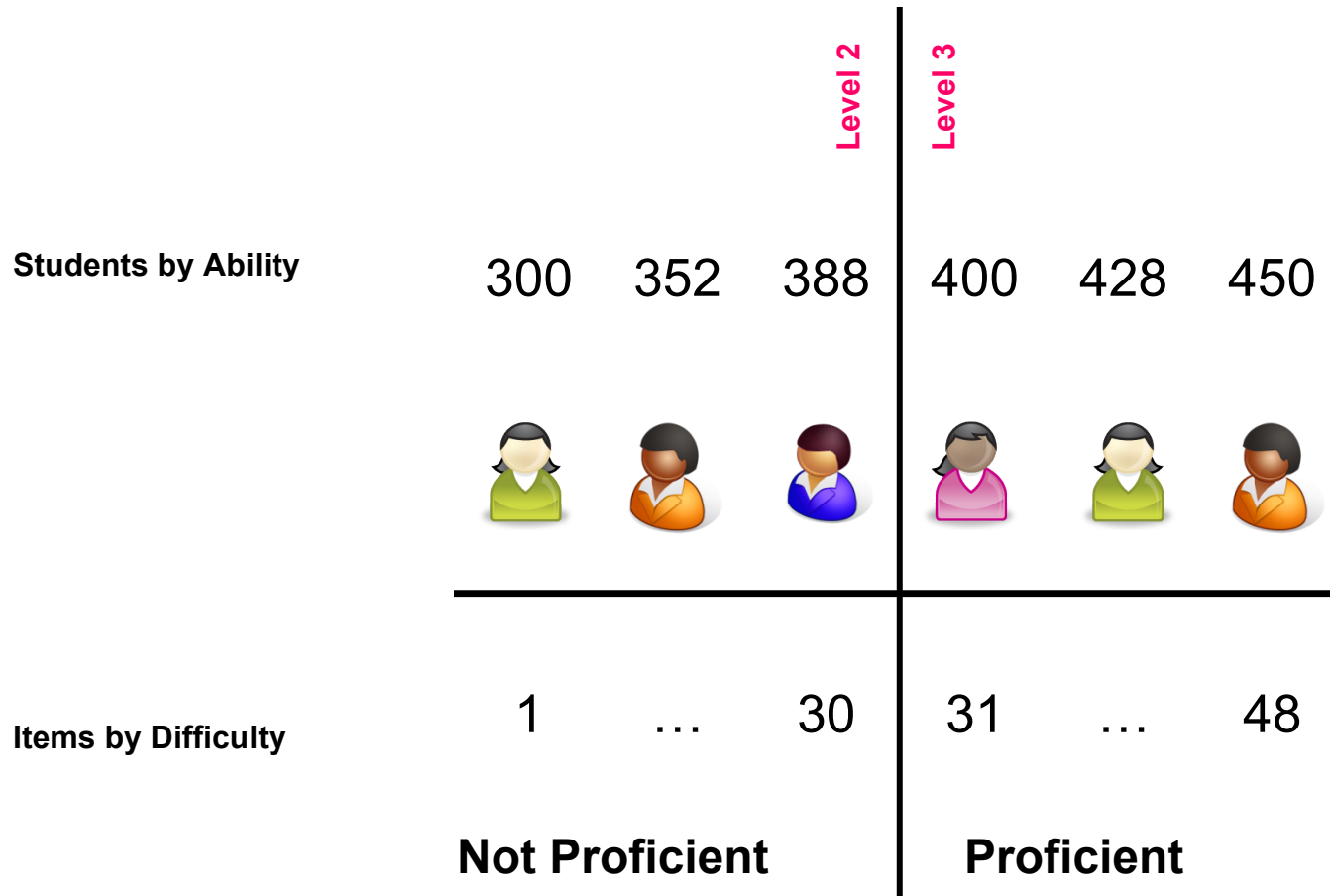
...

...

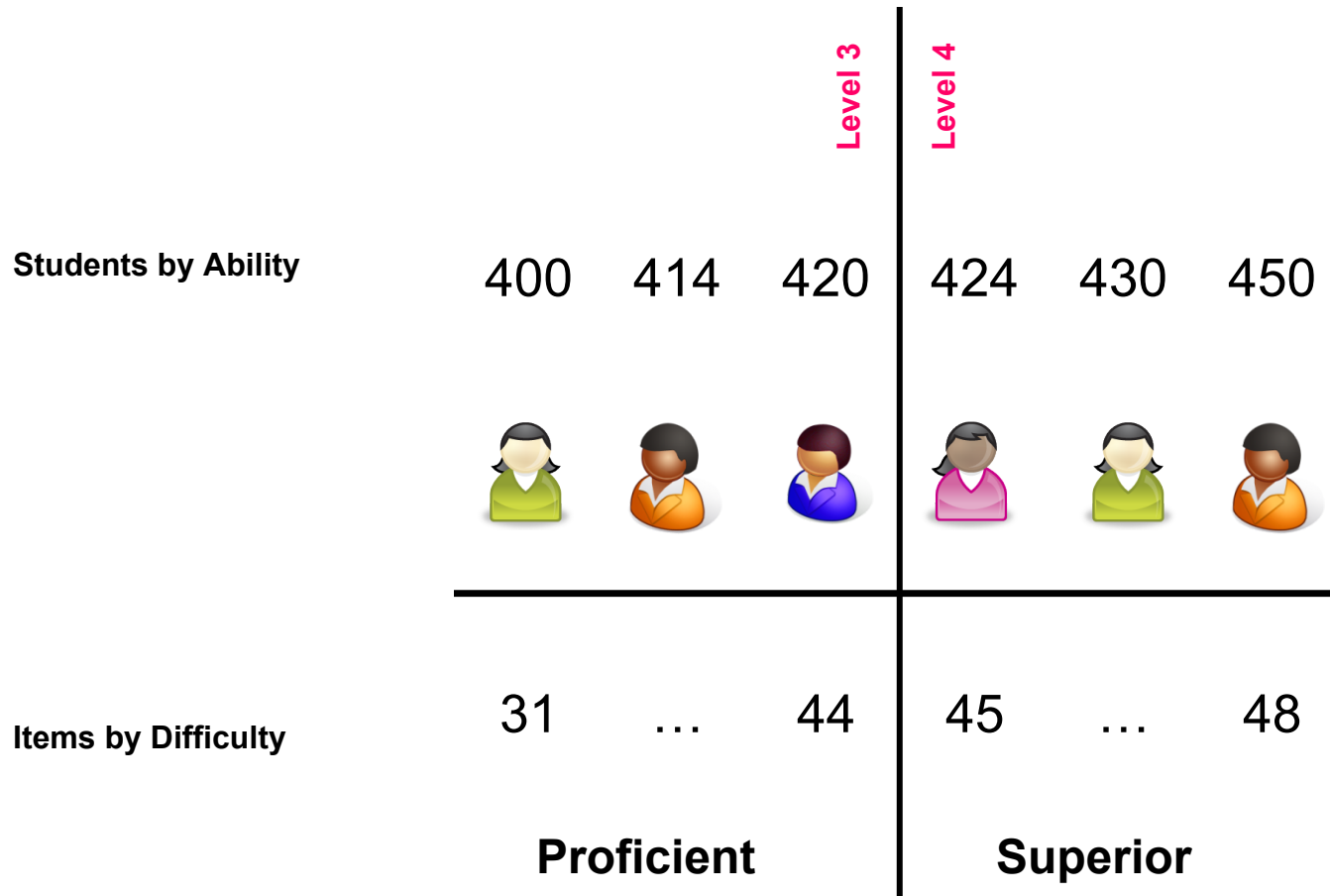
...

48







Test Scale – First Decision



Test Scale – Second Decision



Test Scale – Third Decision

	Level 1			Level 2		
Students by Ability	300	312	325	350	367	388
						
Items by Difficulty	1	...	10	11	...	30
	Extremely Limited			Inconsistent		

Summary

- Items are ranked according to difficulty
- Student ability is ranked on the same scale (higher student abilities are associated with success on higher difficulty items)
- We use the item difficulties associated with the selected item to generate impact data

Break

12:45 – 1:15 PM

Round 3: Data-based Decisions

1:15 -2:30 PM

- Facilitation Team presents impact data based on Round 2 outcomes
- Individuals may adjust bookmarks from Round 2
- With impact data in mind, the group discusses the effectiveness of the proposed cut score
- Discussion between grade level groups with data (high points, justifications, sticky spots, and resolution)
- Use **pink** post-it notes to mark the location, record the item number, and sign the post-it note

Impact Data Review

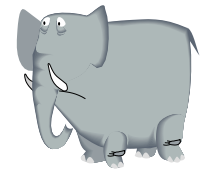
- Round 1 & 2 Complete
- Review percentages of students who would be placed in each performance level based on the selected cut scores
- Revise cut scores, if needed, for final determinations

How do we know it's "right"?

- Look for a reasonable pattern that is cogent and defensible (arguable)
- Do the data appear to progress reasonably?
- Are the results consistent with, or an improvement on prior proficiency percentages?
- ELA & Math
 - Do the data spike/dip at any one grade?
 - Outcomes should be well-articulated across grades and cohesive within subjects

Sample Results (Hypothetical)

	Nearly Meets	Meets	Exceeds
Hypothetical Cutscore (Median)	10	35	45



Impact data

	Does not yet meet	Nearly meets	Meets	Exceeds
Hypothetical Percentages	15%	15%	60%	10%

Establishing Achievement Level Descriptors



Achievement Level Descriptors vs. Content Standards

2:30 – 3:30 PM

- Achievement Level Descriptors (ALDs): Concise statements of the performance required for a student to demonstrate mastery of the content (by level or category)
- Content Standards: Minimum descriptions of what students are expected to learn by subject area, by grade. Minimum that teachers should be teaching.

Achievement Level Descriptor Overview

- ALDs describe what students know and can do based on their performance on statewide assessments in the various content areas.
- The ALDs are based on a sampling of a larger set of testable content outlined in the Oregon Content Standards (RDBC) and give a concise yet general description of what most students know and can do within a particular level of achievement.
- Students who score at or within a particular level of achievement possess the bulk of the abilities described at that level and generally have mastered the skills described in the preceding achievement levels.

ALD Categories

- ALDs for each subject area are developed to establish the minimum scores required for:
 - Level 4
 - Level 3
 - Level 2
 - Level 1

Activity

Read through the ALDs and consider the student you would consider minimally competent in this area (in light of the RBDC of the standards). Answer the following questions:

1. Is this language clear enough to communicate student performance to parents?
2. Does the definition accurately capture a reasonable expectation for this population, at this grade, in keeping with the grade level content standards (RBDC)?
3. Is the expectation for this population a sufficiently appropriate parallel to expectations for students taking the general benchmark assessment?
4. Suggested edits? Please record on your hard copies and flag for us with a **white** post-it note

Summary

3:30 – 4:00 PM

- Results across grades
- Impact results across grades
- ALD discussion
- **Please fill out your yellow Standard Setter Evaluation form and give it to your Table Facilitator**
- State Board of Education adoption
- Use of Cut scores and ALDs for AMO determinations and score reports

References

- CTB Standard Setting Handbook 2005 CTB/McGraw-Hill LLC
- Cizek, G. J. (Ed.). (2012). *Setting performance standards: Foundations, methods, and innovations*. New York: Routledge.
- Smarter Balanced Achievement Level Descriptors:
<http://www.smarterbalanced.org/achievement-levels/>
- First Contact Census Handout, Dynamic Learning Maps, 2013
http://dynamiclearningmaps.org/sites/drupal.dynamiclearningmaps.org/files/documents/First_Contact_Handout_8_6_13.pdf

Questions?

- Brad Lenhardt, Monitoring and Assessment Specialist at Brad.Lenhardt@state.or.us
- Dan Farley, Behavioral Research & Teaching at dfarley@uoregon.edu
- Gerald Tindal, Behavioral Research & Teaching at gerald.tindal@mac.com

Safe Travels & Happy Summer!



APPENDIX H

Table H-1.

Panelist content considerations during judgment rounds

Subject	Discourse
Mathematics	<p>Prerequisite Skills. Cutting at [X] point [earlier] would introduce the concept of proportion sooner, students in this grade are clearly getting this concept.</p> <p>Unintended interactions with item. The question asks [X], but students may be drawn to [Y] in the graphics/or among the response options.</p> <p>General guidance. Do not want to be driven by the percentages.</p> <p>Complexity of academic concepts. Looking at triangles (vs. thinking about how fast cars drive) should be easier.</p> <p>General guidance. Consider difficulty but look at how different the item difficulty increases, it may just be by a fraction of a point.</p>
ELA	<p>Unintended interactions with item. My students would see this question about going outside and rather than select “recess” they would select the response that matches their context. In this case to raise the topic of going outside will immediately cue my student to select lunch because it is part of their reality [“lunch” is among the answer options].</p> <p>Unintended interactions with item. The prompt includes the word “bees” several times. For that reason alone, my student will gravitate to the word “bees” in the response options.</p> <p>Experience with the item type or content. In this item the word “write” changes tense several times, “write”, “writes”, “wrote”. My students are not familiar with changing tenses. For this reason this item is in a more difficult category. In addition, the student is required to make an inference about authorship.</p> <p>Unintended interactions with item. Graphic (image) may be clear or may paint a specific image, but it sometimes elicits a different response or key the student in to a single/specific word on the assessment.</p>

APPENDIX I

Table I-1.

Science Group Final Rationale

Grade Level	Science Group Rationale
5	<p>Level 2 was set at item 9 because entry-level science knowledge begins at that item and the vocabulary is becoming increasingly more advanced.</p> <p>Level 3 was set at item 29 because this question separates proficient and incorporates more of the science content standards (i.e. gravity which is a more complex concept).</p> <p>Level 4 was set at item 46 which is where more complex questioning and concepts begin.</p>
8	<p>Level 2 was set at number 19 which is the end of more concrete concepts simple vocabulary.</p> <p>Level 3 was set at number 36. The vocabulary changed, and students have to distinguish the difference between a set of organs. No obvious outliers.</p> <p>Level 4: Number 51, increased vocabulary and science concepts, including genetics.</p>
11	<p>Level 2: For level of academic vocabulary, [this population has] little to no life experience with the content and/or access to the background knowledge. For some questions, distractors were also taken into consideration. The next item in next section difficulty level increased. This population lacks receptive and verbal communication and focuses on life skills.</p> <p>Level 3: As far as content, students at this level understand and comprehend basic science content and vocabulary. Students have had modified science and are improving in increasing their understanding of the content.</p> <p>Level 4: These students have a more complex thought process and can access a general education class with support and are able to process the content at a higher understanding and can generalize it better.</p>

Table I-2.*Mathematics Group Final Rationale (3- 5)*

Grade Level	Mathematics Group Rationale
3	<p>Following round 3 our group felt the level 1 percentage was too high for kids taking the test (39%) so we lowered the number of items in level 1. We picked the point at which students needed to start interacting more as the [defining] level difference.</p> <p>Level 3 (item 16) started where multiplication and fractions came into the test which is heart of third grade standards so we felt students successful at those items should meet.</p> <p>Level 4 was ultimately set at item 44 which is where two-step problem solving begins. [When students are successful at two-step problems] IEP teams may consider SBAC an option.</p>
4	<p>Level 1: The impact of the data is 15.4. We felt the complexity went from number based ten and numeracy knowledge in the ones and identifying a geometric shape to –</p> <p>Identifying double digit numbers [which is the start of Level 2], shape recognition and labeling a shape with a symbol (right angle, item 8) adding, using standard measurement, recognizing the definition of same, half, equation, all together, heavier, inches, and feet.</p> <p>Level 3: Students can perform level 1 and 2 well, and identify skip counting, identify $y=mx+b$ in item 38 solve mixed numbers, identify area, round numbers, use systems of inequality, and recognize and apply the term “half” in a story problem.</p> <p>Level 4: Can perform levels 1 – 3 proficiently and show division without scaffolding, solve multiplication problems and solve applied math problems with $\frac{1}{4}$.</p>
5	<p>First we determined that students who are minimally proficient would be capable of consistently recognizing numbers 0 – 20, match numbers 0 – 6, using visual supports add numbers, recognize the total number of objects and complete patterns +2, +5. Students could be inconsistent analyzing a table (pictograph), subtract with visual cues, and understand volume and fractions.</p> <p>We adjusted our cut-scores for Exceeded because we determined that item #33 had decent visual support of item #34 involved counting by 10 – a skill many students do learn. The adding of fractions without visuals was the cut off for superior (exceeded) item #35.</p> <p>The cut off for Nearly meets was done at the point that students had to not just match numbers, but had to understand the question read in order to get the right answer. Students looked at shapes to determine the number of angles. This required more cognitive ability than the simple matching of numbers. We set this cut off at # 8.</p> <p>After looking at our impact data we changed the cut off for nearly meets to #8 instead of #13 so that we could catch more students at the nearly meets level. The first time we did impact data, our “does not meet” was slightly larger than our “nearly meets”.</p>

Table I-3.*Mathematics Group Final Rationale (6 – 8)*

Grade Level	Mathematics Group Rationale
6	<p>Level 2: We chose item 6 because it was a jump to addition instead of just identification which is more of a complex skill.</p> <p>Level 3: We chose Item 13, because of the increased variables to analyze and manipulate, the scale of count-by (2s), ignoring distracting or irrelevant information, and the complexity of the visual information.</p> <p>Level 4: Item 37 showed a significant jump in complexity, number of variables, abstract of time, symbols, multi-step problems. Students that could answer these questions correctly should be considered for the SBAC with supports.</p>
7	<p>We feel confident and comfortable with the cut scores we arrived at in our OIB. We used the impact data after round two to increase the percentage of students reaching proficiency in Round 3. In addition, we paid more attention to the item difficulty in round three to inform our decision-making process.</p>
8	<p>We feel confident our cut scores accurately reflect the student population taking the assessment. We looked at large shifts of the complexity of the content to determine our cut scores. After reviewing the ALDS, we feel that our cut scores align well. The impact scores helped us revise our cut scores, putting them in alignment with the ALDs.</p>
11	<p>Level 2: (item 7) introduces difficult vocabulary “hexagon”.</p> <p>Level 3: (item 13) introduces the x axis vocabulary. Finding a point on a line is more difficult than matching two symbols.</p> <p>Level 3: (item 43) first multi-step items. Dealt with division and fractions and line segments.</p>

Table I-4.*ELA Group Final Rationale (3 – 5)*

Grade Level	English Language Arts Group Rationale
3	<p>Level 2: [Following the review of our impact data] we wanted to decrease the proportion of students in Level 1, so we lowered our item number from 8 to 5. Question 6 marked a change to word functions as opposed to identification.</p> <p>Level 3: We kept 18 the same because it is a shift to an academic skill (decoding).</p> <p>Level 4: We adjusted from item 52 to 54 to decrease the proportion in level 4. It was also an increase in academic skill (longer sentences, higher vocabulary to decode).</p>
4	<p>Level 2: Bookmark was placed at item 8 because content moved from identification with pictures, student was required to perform independently, the item difficulty jumped from .04 to .09, and it was a move to more expressive language.</p> <p>Level 3: Bookmark was placed at item 23 because there were no visuals, no answer choices read to the student, it required the student to draw from their own experience using vocabulary, and the item difficulty increased from 1.326 to 1.346.</p> <p>Level 4: Bookmark was placed at item 45 because the item required the student to have knowledge of pronouns, the answer was not contained within the item or the answer, and item difficulty [increased to] 2.746 whereas all the previous items were in the 2.5 range.</p>
5	<p>Between levels 1 and 2: Variety of picture choices.</p> <p>Between levels 2 and 3: Visual supports drop off and a student who is not proficient would need visuals. Literal, auditory skills, pictures are provided. Prior knowledge may help a student.</p> <p>Between levels 3 and 4: Main idea begins in Level 3 because it is a higher order concept. Language increased, content and vocabulary. Higher order thinking strategies were involved. Item 44 requires writing skills, vocabulary knowledge, and letter discrimination.</p> <p>Levels 4 change – students are required to read. Vocabulary, different forms between question and answer. Moved Level 3 to where L4 was. Item 47 not an emphasizing word. 48 – 50 (ordinal, multi-step). Language complexity. Dialogues, length, multi-sentence choices (all “R”).</p>

Table I-5.*ELA Group Final Rationale (6 – 8)*

Grade Level	English Language Arts Group Rationale
6	<p>Level 2 entry item number 5. Introduces visual discrimination, girls all look similar. Segue quickly into handwriting.</p> <p>Level 3 entry is item 25. Text is not repetitive of subject like previous questions. Requires discrimination between 3 boys' names that all begin with same letter, similar in length.</p> <p>Level 4 entry is item 45. Science content/vocabulary introduced. Text in questions becomes much longer on next question.</p>
7	<p>Our group chose item 3 for our level 2 cut because it asked students to trace, which requires them to be more independent than just pointing or matching. The following items required a number of skills including inferring, summarizing, and working without picture cues. We chose item 30 for our cut score. We decided to be proficient in the 7th grade you should be able to read K- 1 text. From there forward, the writing had no model and a variety of skills were represented, as opposed to basic listening comprehension only.</p> <p>We chose item 48 for our level ¾ cut because this item required you to think about and apply what was read instead of just careful reading. We had previously chosen 45, but the impact data showed a large number of students exceeding. We chose a new method to increase the cut score. We reached consensus and are confident in our results.</p>
8	<p>Levels 2 – 3: At item 18. Students have to read independently (~2nd grade text). Significant jump in difficulty, from 17 to 18. Contains decoding and comprehension. 58-60% kids will meet or exceed. Decided not to look at low items for cut.</p> <p>Level 3 – 4: Item 50. Percentage of students in exceed was 37% at item 48. Needs to show understanding of word tenses “write, wrote, written”. Distractor of multiple names (Ginny first). Three distractor sentences were presented after the clue and before the question.</p> <p>Level 1 – 2: Item 5. Too many students level 1 when item 7 was the cut off. Answer [needs to be] based on text, not belief that fish, cat make better pets. Some bias for dog owners, i.e. more likely to get it correct if the student is a dog owner.</p>

Table I-6.*ELA Group Final Rationale (11)*

Grade Level	English Language Arts Group Rationale
11	<p>Between Levels 1 and 2: Started with 5, but ended with item 3. Started with 5 due to issues of some students (struggling with) tracing, and reading tasks were similar with support, concept and length. (We always were dubious about 3) because the Flag is so abstract, and not very much support [was provided by the] pictures. Went back after getting impact data and looked at 3 again. Talks about stars, stripes. Not everyone knows what they are. Talks about colors, but pictures are black and white.</p> <p>Between Level 2 and 3 (item 35):</p> <p>Required inference. The answer was not readily in question. Required student to apply word to choices and see what worked. Impact data for this choice still allowed 63% of the students to pass.</p> <p>Between 3 and 4 (item 48—previously 43):</p> <p>We felt that pictures distracted rather than supported. Had to understand what the word “face” meant in this context.</p> <p>We changed it to 48 based on the impact data. Another jump of skills required, understanding verb tenses, helping verb (not an action word), items 49 and 50 had a lot of distracting information in both readings. Items 49 and 50 content was abstract.</p>

APPENDIX J

Determination of Cut Scores via judgmental rounds (including review of Impact data)

Table J-1.

ORExt Science Grade 5 Rounds 1 – 3 Judgment Results Item (and Item Difficulty) by person/consensus

Grade 5 Science						
	Round 1				Round 2	Round 3
	Person 1	Person 2	Person 3	Person 4	Consensus	Consensus
Level 2	17 (0.986)	9 (0.556)	16 (0.926)	25 (1.536)	17 (0.986)	9 (0.556)
Level 3	21 (1.176)	17 (0.986)	30 (1.676)	31 (1.776)	29 (1.656)	29 (1.656)
Level 4	29 (1.656)	36 (1.956)	37 (2.006)	40 (2.306)	37 (2.006)	46 (2.956)

Table J-2.

Grade 5: Impact following Round 2

Level	Percentage
1	30.9
2	8.6
3	6.5
4	53.9

Table J-3.*Grade 5: Impact following Round 3*

Level	Percentage
1	26.7
2	12.8
3	24.9
4	35.6

Table J-4.*ORExt Science Grade 8 Rounds 1 – 3 Judgment Results Item (and Item Difficulty) by person/consensus*

Grade 8 Science						
	Round 1		Round 2		Round 3	
	Person 1	Person 2	Person 3	Person 4	Consensus	Consensus
Level 2	19 (0.956)	15 (0.856)	36 (2.016)	19 (0.956)	26 (1.526)	19 (0.956)
Level 3	25 (1.426)	27 (1.556)	43 (2.416)	48 (2.616)	36 (2.016)	36 (2.016)
Level 4	35 (1.916)	43 (2.416)	44 (2.446)	50 (2.716)	44 (2.446)	51 (3.106)

Table J-5.*Grade 8: Impact following Round 2*

Level	Percentage
1	36
2	6.5
3	5.6
4	51.9

Table J-6.*Grade 8: Impact following Round 3*

Level	Percentage
1	28.8
2	13.7
3	15.2
4	42.3

Table J-7.*ORExt Science Grade 5 Rounds 1 – 3 Judgment Results Item (and Item Difficulty) by person/consensus*

Grade 11 Science					
	Round 1			Round 2	Round 3
	Person 1	Person 2	Person 3	Consensus	Consensus
Level 2	21 (1.216)	11 (0.756)	5 (0.106)	8 (0.526)	5 (0.106)
Level 3	28 (1.556)	23 (1.256)	22 (1.216)	24 (1.406)	24 (1.406)
Level 4	37 (2.126)	37 (2.126)	38 (2.246)	38 (2.246)	47 (2.856)

Table J-8.*Grade 11: Impact following Round 2*

Level	Percentage
1	24
2	6.8
3	12.7
4	55.7

Table J-9.*Grade 11: Impact following Round 3*

Level	Percentage
1	20.8
2	10.8
3	21.2
4	47.2

Table J-10.*Average Across Grade Levels: Impact following Round 3*

Level	Percentage	SD
1	25.43	4.15
2	12.43	1.48
3	20.43	4.90
4	41.7	5.82

APPENDIX K

Table K-1.

ORExt Mathematics Grade 3 Rounds 1 – 3 Judgment Results Item (and Item Difficulty) by person/consensus

Grade 3 Mathematics					
	Round 1			Round 2	Round 3
	Person 1	Person 2	Person 3	Consensus	Consensus
Level 2	14 (0.006)	15 (0.096)	24 (0.566)	15 (0.096)	6 (-0.764)
Level 3	28 (0.896)	28 (0.896)	32 (1.146)	28 (0.896)	16 (0.136)
Level 4	44 (1.816)	42 (1.696)	42 (1.696)	42 (1.696)	44 (1.816)

Table K-2.

ORExt Mathematics Grade 4 Rounds 1 – 3 Judgment Results Item (and Item Difficulty) by person/consensus

Grade 4 Mathematics					
	Round 1			Round 2	Round 3
	Person 1	Person 2	Person 3	Consensus	Consensus
Level 2	10 (-0.464)	11 (-0.344)	5 (-0.994)	5 (-0.994)	5 (-0.994)
Level 3	29 (0.956)	29 (0.956)	35 (1.436)	29 (0.956)	25 (0.676)
Level 4	42 (1.906)	49 (2.526)	48 (2.236)	48 (2.236)	48 (2.236)

Table K-3.

ORExt Mathematics Grade 5 Rounds 1 – 3 Judgment Results Item (and Item Difficulty) by person/consensus

Grade 5 Mathematics					
	Round 1			Round 2	Round 3
	Person 1	Person 2	Person 3	Consensus	Consensus
Level 2	29 (1.206)	18 (0.206)	13 (-0.124)	13 (-0.124)	8 (-0.664)
Level 3	30 (1.326)	31 (1.326)	22 (0.616)	22 (0.616)	22 (0.616)
Level 4	43 (2.176)	40 (1.986)	32 (1.356)	32 (1.356)	35 (1.586)

Table K-4.

ORExt Mathematics Grade 6 Rounds 1 – 3 Judgment Results Item (and Item Difficulty) by person/consensus

Grade 6 Mathematics					
	Round 1			Round 2	Round 3
	Person 1	Person 2	Person 3	Consensus	Consensus
Level 2	13 (0.846)	23 (1.356)	13 (0.846)	16 (0.946)	6 (0.406)
Level 3	24 (1.486)	33 (2.066)	24 (1.486)	24 (1.486)	13 (0.846)
Level 4	37 (2.176)	43 (2.626)	37 (2.176)	37 (2.176)	37 (2.176)

Table K-5.

ORExt Mathematics Grade 7 Rounds 1 – 3 Judgment Results Item (and Item Difficulty) by person/consensus

Grade 7 Mathematics					
	Round 1			Round 2	Round 3
	Person 1	Person 2	Person 3	Consensus	Consensus
Level 2	18 (0.746)	19 (0.776)	29 (1.396)	18 (0.746)	6 (-0.244)
Level 3	28 (1.386)	28 (1.386)	36 (1.826)	30 (1.506)	22 (0.916)
Level 4	43 (2.276)	38 (1.916)	50 (2.776)	43 (2.776)	50 (2.776)

Table K-6.

ORExt Mathematics Grade 8 Rounds 1 – 3 Judgment Results Item (and Item Difficulty) by person/consensus

Grade 8 Mathematics					
	Round 1			Round 2	Round 3
	Person 1	Person 2	Person 3	Consensus	Consensus
Level 2	19 (1.316)	19 (1.316)	9 (0.916)	10 (0.926)	5 (0.806)
Level 3	28 (2.306)	26 (2.096)	18 (1.236)	26 (2.096)	18 (1.236)
Level 4	43 (2.906)	35 (2.566)	30 (2.336)	35 (2.566)	35 (2.566)

Table K-7.

ORExt Mathematics Grade 11 Rounds 1 – 3 Judgment Results Item (and Item Difficulty) by person/consensus

Grade 11 Mathematics					
	Round 1			Round 2	Round 3
	Person 1	Person 2	Person 3	Consensus	Consensus
Level 2	9 (1.206)	21 (1.376)	11 (0.566)	19 (1.206)	6 (0.136)
Level 3	29 (1.646)	29 (1.646)	27 (1.616)	29 (1.646)	13 (0.656)
Level 4	44 (2.216)	44 (2.206)	37 (1.846)	43 (2.206)	43 (2.206)

Table K-8.

Mathematics Impact following Round 2

	3	4	5	6	7	8	11
Level 1	39%	15.4%	21.5%	46.9%	40.7%	44.9%	62%
Level 2	23.3%	40.7%	17.1%	16.8%	19.4%	39%	11.7%
Level 3	20.9%	30.2%	26.8%	18.1%	24.6%	9.5%	12.5%
Level 4	16.8%	13.7%	34.6%	18.1%	15.4%	6.7%	13.8%

Table K-9.

Mathematics Impact following Round 3

	3	4	5	6	7	8	11	Mean	SD
Level 1	25.9%	15.4%	15.5%	32.1%	19.5%	41.9%	38.2%	26.93%	10.78
Level 2	13.9%	32.2%	25.6%	10.7%	25.3%	13%	11.9%	18.94%	8.55
Level 3	44.5%	38.7%	33.1%	39.1%	46.8%	38.5%	36.2%	39.56%	4.69
Level 4	15.7%	13.7%	25.8%	18.1%	8.5%	6.7%	13.8%	14.61%	6.32

Table K-10.*Mathematics Impact following Articulation*

	3	4	5	6	7	8	11	Mean	SD
Level 1	25.9%	15.4%	15.5%	32.1%	19.5%	41.9%	38.2%	26.93%	10.78
Level 2	13.9%	30.5%	25.6%	10.7%	25.3%	13%	11.9%	18.7%	8.13
Level 3	44.5%	34.8%	45%	39.1%	39.9%	38.5%	36.2%	39.71%	3.86
Level 4	15.7%	19.3%	14%	18.1%	15.4%	6.7%	13.8%	14.71%	4.07

APPENDIX L

Table L-1.

ORExt English Language Arts Grade 3 Rounds 1 – 3 Judgment Results Item (and Item Difficulty) by person/consensus

Grade 3 English Language Arts					
	Round 1			Round 2	Round 3
	Person 1	Person 2	Person 3	Consensus	Consensus
Level 2	6 (-0.304)	8 (0.076)	5 (-0.764)	8 (0.076)	5 (-0.764)
Level 3	18 (1.316)	18 (1.316)	11 (0.166)	18 (1.316)	18 (1.316)
Level 4	52 (2.776)	52 (2.776)	36 (2.166)	52 (2.776)	54 (3.006)

Table L-2.

ORExt English Language Arts Grade 4 Rounds 1 – 3 Judgment Results Item (and Item Difficulty) by person/consensus

Grade 4 English Language Arts					
	Round 1			Round 2	Round 3
	Person 1	Person 2	Person 3	Consensus	Consensus
Level 2	14 (0.656)	18 (1.016)	6 (-0.024)	14 (0.656)	8 (0.096)
Level 3	23 (1.346)	23 (1.346)	18 (1.016)	23 (1.346)	23 (1.346)
Level 4	36 (2.026)	33 (2.026)	23 (1.346)	31 (1.916)	45 (2.746)

Table L-3.

ORExt English Language Arts Grade 5 Rounds 1 – 3 Judgment Results Item (and Item Difficulty) by person/consensus

Grade 5 English Language Arts					
	Round 1			Round 2	Round 3
	Person 1	Person 2	Person 3	Consensus	Consensus
Level 2	11 (0.346)	16 (0.816)	1 (-1.054)	9 (1.066)	12 (0.516)
Level 3	18 (0.986)	25 (1.576)	12 (0.516)	16 (0.816)	30 (2.006)
Level 4	44 (3.016)	36 (2.556)	26 (1.666)	25 (1.576)	47 (3.246)

Table L-4.

ORExt English Language Arts Grade 6 Rounds 1 – 3 Judgment Results Item (and Item Difficulty) by person/consensus

Grade 6 English Language Arts					
	Round 1			Round 2	Round 3
	Person 1	Person 2	Person 3	Consensus	Consensus
Level 2	11 (0.856)	7 (0.596)	19 (1.296)	12 (0.916)	5 (0.466)
Level 3	18 (1.216)	25 (1.666)	27 (1.836)	25 (1.666)	25 (1.666)
Level 4	28 (1.876)	45 (2.976)	33 (2.116)	38 (2.506)	45 (2.976)

Table L-5.

ORExt English Language Arts Grade 7 Rounds 1 – 3 Judgment Results Item (and Item Difficulty) by person/consensus

Grade 7 English Language Arts					
	Round 1			Round 2	Round 3
	Person 1	Person 2	Person 3	Consensus	Consensus
Level 2	3 (0.386)	3 (0.386)	15 (1.336)	6 (0.776)	3 (0.386)
Level 3	30 (2.226)	23 (1.796)	30 (2.226)	30 (2.226)	30 (2.226)
Level 4	47 (3.606)	40 (2.746)	40 (2.746)	45 (3.066)	48 (3.636)

Table L-6.

ORExt English Language Arts Grade 8 Rounds 1 – 3 Judgment Results Item (and Item Difficulty) by person/consensus

Grade 8 English Language Arts					
	Round 1			Round 2	Round 3
	Person 1	Person 2	Person 3	Consensus	Consensus
Level 2	5 (1.266)	7 (1.596)	9 (1.726)	7 (1.596)	5 (1.266)
Level 3	18 (2.426)	21 (2.556)	18 (2.426)	18 (2.426)	18 (2.426)
Level 4	39 (3.106)	30 (2.886)	43 (3.196)	48 (3.596)	50 (3.646)

Table L-7.

ORExt English Language Arts Grade 11 Rounds 1 – 3 Judgment Results Item (and Item Difficulty) by person/consensus

Grade 11 English Language Arts					
	Round 1			Round 2	Round 3
	Person 1	Person 2	Person 3	Consensus	Consensus
Level 2	8 (0.416)	16 (0.996)	15 (0.966)	5 (0.146)	3 (-0.124)
Level 3	27 (1.606)	25 (1.466)	30 (1.816)	35 (1.996)	35 (1.996)
Level 4	43 (2.386)	38 (2.246)	46 (2.536)	43 (2.386)	48 (2.736)

Table L-8.

English Language Arts (ELA) Impact following Round 2

	3	4	5	6	7	8	11	Mean	SD
Level 1	19.4%	20.5%	17.5%	22.4%	22.4%	30%	21.6%	21.97%	3.95
Level 2	16.1%	8%	4.4%	6.1%	12.8%	11.4%	15.2%	10.57%	4.52
Level 3	23.0%	9.7%	7.7%	10.6%	11.3%	21.6%	5.3%	12.74%	6.84
Level 4	41.5%	61.9%	70.4%	60.9%	53.5%	37%	57.9%	54.73%	11.80

Table L-9.

English Language Arts (ELA) Impact following Round 3

	3	4	5	6	7	8	11	Mean	SD
Level 1	12.1%	15.2%	19.4%	19.0%	19.5%	27.3%	19.5%	18.86%	4.68
Level 2	23.4%	13.3%	14.2%	9.5%	15.7%	14.2%	17.3%	15.37%	4.28
Level 3	26.7%	23.6%	19.3%	17.6%	21.8%	24.1%	11.8%	20.7%	4.97
Level 4	37.7%	48.0%	47.0%	53.9%	43.0%	34.5%	51.5%	45.09%	7.09

Table L-10.*English Language Arts Impact following Articulation*

	3	4	5	6	7	8	11	Mean	SD
Level 1	12.1%	15.2%	17.5%	19%	22.4%	27.3%	19.5%	19%	4.92
Level 2	23.4%	13.3%	16.2%	13%	12.8%	14.2%	17.3%	15.74%	3.78
Level 3	23%	23.6%	19.3%	23.1%	21.8%	24.1%	11.8%	20.96%	4.34
Level 4	41.5%	48%	47%	44.8%	43.0%	34.5%	51.5%	44.33%	5.46

APPENDIX M

Mathematics ALD review panelists verbal summary

Grade 3: Make it clearer that level 1 is not where you want to be. The amount of text/narrative in level 1 makes it seem as though it is the desired category.

Grade 4: Use of the term “essentialized” may be too unfamiliar to parents. Remove the word “far” from the description of [far] exceeds. Will parents understand the “Base 10” language. Use unit squares instead of square feet or inches.

Grade 5: Remove some language from level 1.

Grade 6: No major edits. Review of ALDs was confirmatory regarding their standard setting decisions.

Grade 7: Recommend putting percentages prior to fractions so that percentages appear in level 3 while fractions appear in level 4.

Grade 8: No edits. Review of ALDs was confirmatory regarding their standard setting decisions.

Grade 11: Minor edits. Review of ALDs was confirmatory regarding their standard setting decisions.

ELA ALD Review panelist verbal summary

Grade 3: Minor typos. Some writing progression across key points needed to be made.

Grade 4: Minor typos.

Grade 5: Define “short and medium” sentences. Align the wording between levels 1 and 4. In some cases they are not consistent. Remove the word “extremely” from the phrase “extremely limited” in level 1 as it may have a negative impact on parents.

Grade 6: Language of ALDs was broader than the language we reviewed in the OIBs.

Grade 7: Define the term “medium sentences”. For level 4 add “or more” to the end of any expectation that limits the number of sentences a student will read or interact with.

Grade 8: Minor typos. Is increasing the length of sentences enough to justify getting into a higher level?

Grade 11: Recommend using the term “proficient” instead of mastery. Add “or more” when referencing expectations regarding sentences. Recommend students read to themselves rather than being read to.

APPENDIX N

Table N-1. *Evaluation Statements and Percentage of Panelists per Response*

		SCI (11)			Math (21)			ELA (20)		
	Statement	SA %	A%	D%	SA %	A%	D%	SA %	A%	D%
1	The orientation provided me with a clear understanding of the purpose of the standard setting meeting.	100%			81%	19%		80%	20%	
2	The training helped me understand the bookmark method and how to perform my role as a standard setter.	82%	18%		81%	19%		80%	20%	
3	Reviewing the ORExt helped me to understand the assessment.	55%	45%		81%	19%		90%	10%	
4	The small and large group discussions aided my understanding of the process.	91%	9%		71%	29%		95%	5%	
5	There was an equal opportunity for everyone in my group to contribute his/her ideas and opinions.	91%	9%		81%	19%		100%		
6	I was able to follow instructions and complete the rating sheets accurately.	64%	36%		81%	19%		90%	5%	5%
7	The discussions after the first round of ratings were helpful to me.	91%	9%		81%	19%		90%	10%	
8	The discussions after the second round of ratings were helpful to me.	91%	9%		90%	10%		90%	10%	
9	The information showing the impact of our cutscores on proficiency percentages was helpful to me.	91%	9%		81%	19%		90%	10%	
10	I am confident about the defensibility and appropriateness	73%	27%		82%	18%		85%	15%	

	of the final recommended cut scores.									
11	The achievement level descriptions were clear and useful.	55%	45%		52%	43%	5%	60%	40%	
12	The time provided for discussions was adequate.	82%	18%		82%	18%		90%	5%	5%
13	The workshop leaders helped to answer questions and ensure that all input was respected and valued.	100%			86%	14%		100%		
14	The facilities and food service helped create a productive and efficient working environment.	91%	9%		71%	29%		80%	20%	
15	Overall, I am confident that the standard setting procedures allowed me to use my experience and expertise to recommend cut scores for the ORExt.	91%	9%		81%	19%		95%	5%	

SA = Strongly Agree

A = Agree

D = Disagree

SD = Strongly Disagree (not shown in this image)

APPENDIX O

Selecting the Assessment

General Assessment

In general, an IEP team should consider the General Assessment with or without accommodations if...

Student

- Performs at or around grade level, based on progress monitoring data
- Academic difficulties are “mild to moderate” and can typically be addressed by using simplified language
- Reading is within two to three grades of enrolled level
- Academic difficulties primarily surround one content area (more commonly, reading) but performance in other subject areas is similar to grade-level peers

Instruction

- Student primarily receives instruction from grade level materials
- Student relies on accommodations only to access instructional information

Oregon Extended Assessment

In general, an IEP team should consider administration of the ORExt if...

Student

- Performs well below grade level, based on progress monitoring data
- Academic difficulties are generalized (to all subject areas) and are significant.
- Reads significantly below enrolled grade level or does not read.
- Performance is significantly impacted by the disability
- May have mobility and receptive and expressive language difficulties that are generalized and significant.

Instruction

- Student primarily receives instruction from specialized, and/or functional, materials that are significantly reduced in depth, breadth, and complexity.
- Student relies on accommodations and/or modifications to access instructional information.

To support IEP teams in their deliberations around assessment selection decisions the Oregon Department of Education has created the following instrument:

Assessment Decision Making Tool

Question 1: What is the student's minimum reading level?

- a) Student reads at grade level

Elementary	Middle	High
1 grade below	1 grade below	1 grade below
2 grades below	2 grades below	2 grades below
3+ grades below	3 grades below	3 grades below
	4+ grades below	4+ grades below

Question 2: Based on the student's IEP, what level of instruction is the student receiving in general? Discuss which subjects are most impacted.

- a) Student is instructed from grade level materials.
- b) Student is instructed from specialized (not functional) materials that are reduced in depth, breadth, and complexity
- c) Student is instructed from a functional curriculum only
- d) Student is instructed from both specialized and functional materials

Question 3: Based on the student's IEP, what types of instructional supports does the student rely on to access instructional information?

- a) No supports, student relies on accommodations only
- b) Student relies on both accommodations and modifications
- c) Student relies on modifications only

Question 4: What communication tools does the student require to participate in classroom instruction/ assessments?

- a) Independent communication?
- b) Name of specific tool _____

Question 5: Does the student navigate use of a computer independently?

- a) Student navigates computer independently
- b) Student requires supervision and adult support to navigate computer

c) Student requires occasional assistance when navigating computers for use

Question 6: How has the student progressed in the current academic year based on progress monitoring tools of achievement and/or on classroom or district assessments?

Discuss data

Question 7: How does the student interact with OAKS sample assessments?

Discuss observations in terms of accommodations or assessment decisions

Question 8: What statewide assessment(s) did the student participate in last year?

- a) OAKS
- b) Extended Assessment
- c) Both
- d) None

Question 10: What was the student's performance level(s) on each of the statewide assessments s/he took? Indicate whether OAKS or ORExt by circling.

ELA (Smarter Balanced/ ORExt)	Mathematics (Smarter Balanced/ORExt)	Science (OAKS/ORExt)
Level 1 (Below Basic)	Level 1 (Below Basic)	Level 1 (Below Basic)
Level 2 (Basic)	Level 2 (Basic)	Level 2 (Basic)
Level 3 (Proficient)	Level 3 (Proficient)	Level 3 (Proficient)
Level 4 (Advanced)	Level 4 (Advanced)	Level 4 (Advanced)