Selection, Links, Alignment: Establishing a Validity Position for Oregon's Extended Assessments

Submitted to Oregon Department of Education June 1, 2017

DCE Educational Communications LLC

Summary

Oregon's Extended Assessments (ORExt) in English Language Arts, Mathematics, and Science were evaluated in a low-complexity alignment study conducted in Spring of 2017. Averages of reviewer professional judgments over five separate evaluations were gathered, reviewed, and interpreted in the pages that follow. In the three evaluations that involved determining the relationship between standards and items, reviewers identified sufficient to strong relationships among assessment components in all grades and all subject areas. In the two evaluations involving Achievement Level Descriptors, reviewers identified thirty instances of sufficient to strong relationships out of thirty-four possible relationship opportunities resulting in an overall affirmed relationship with areas for refinements identified.

Selection, Links, Alignment: Establishing a Validity Position for Oregon's Extended Assessments

Submitted to Oregon Department of Education June 1, 2017

Objective: Use professional judgment of content to determine the strength of the relationships among the components of Oregon's Alternate Assessment System.

Goal: Validate the achievement inferences made by users of the alternate assessment.

Overview:

The Oregon Extended Assessments (ORExt) are the state's alternate assessments based on alternate achievement standards (AA-AAS) specifically developed to assess the population of students whose significant cognitive disabilities preclude them from meaningfully accessing Oregon's general assessments (the Smarter Balanced Assessments in English Language Arts (ELA) and Mathematics, and Oregon's Assessment of Knowledge and Skill in Science). Oregon's Smarter Balanced Assessments are based on the Common Core State Standards (CCSS), and Oregon's Assessment of Knowledge and Skill (OAKS) in Science is currently based on the Oregon Science Standards (ORSci), though transitioning toward the Next Generation Science Standards (NGSS) by the spring of 2018. (The ORExt in Science is dually-linked to both the ORSci and the NGSS.)

In keeping with alternate achievement standard allowances suggested and recommended by the Individuals with Disabilities Education Act (IDEA) and the Every Student Succeeds Act (ESEA, 2015),

(D) ALTERNATE ASSESSMENTS FOR STUDENTS WITH THE MOST SIGNIFICANT COGNITIVE DISABILITIES.—(i) ALTERNATE ASSESSMENTS ALIGNED WITH ALTERNATE ACADEMIC ACHIEVEMENT STANDARDS.—A State may provide for alternate assessments aligned with the challenging State academic standards and alternate academic achievement standards described in paragraph (1) (E) for students with the most significant cognitive disabilities, if the State— (I) consistent with clause (ii), ensures that, for each subject, the total number of students assessed in such subject using the alternate assessments does not exceed 1 percent of the total number of all students in the State who are assessed in such subject.

the ORExt were designed to serve students with significant cognitive disabilities via the use of an Essentialized Assessment Framework (EAF), in which each original grade level standard was reduced in depth, breadth, and complexity to provide access for this small, heterogeneous population of users, while still reflecting grade level content. As noted in the development of the EAF, the intent of the Essentialized Standards is to increase access for students with the most significant cognitive disabilities, while maintaining the intended link to grade level content.

The accountability assessment of students with the most significant cognitive disabilities is a relatively young area of development and a variety of approaches have been used to both develop and study the assessments nationally. Alignment studies conducted on this topic over the past 15 years, share the goals established by Kane (1992, 2006) of attempting to establish validity by posing a set of interpretive questions/arguments designed to link evidence to inference by eliminating assumptions (a la Flowers, Wakeman, Browder, & Karvonen, 2007). Typically, these studies provide suggestions for ways to

strengthen and improve the validity of the assessment in future development. In keeping with that framework, this study does both.

The study described in the following pages uses a convergence of evidence model to evaluate the validity of the Essentialized Standards, the items used to develop the assessment and to test this population of students, and the Achievement Level Descriptors (ALDs) used to describe the assessment outcomes for this population of students.

Procedure

The relationship between Oregon's Extended Assessments (Science, Math, ELA) and the CCSS and ORSci /NGSS was evaluated in five parts (Evaluations 1 – 5). Each evaluation examined the strength of the relationship between two related parts of the assessment. The strength of the argument at each evaluation stage can be used either individually (to affirm the subsequent product(s)) or cumulatively (to affirm the inferences made using the results of the test). An affirmation of the relationship (*link or alignment) at each of the five stages will serve to the validate the components of Oregon's Extended Assessments and to support the claim that they assess students in this population in manner that is comparable to their peers who take the general assessment (Oregon's Smarter Balanced Assessments or OAKS).

*Note: Establishing a *linked* relationship (such as the type of relationship anticipated between source standards and Essentialized Standards) affirms that the general intent of the original/source standards was maintained, though stripped of complexities that hinder access and interaction for students with the most significant cognitive disabilities (e.g., complex language, nuance). Alternatively, establishing an *aligned* relationship (such as the type anticipated between Essentialized Standards, alternate items, and alternate Achievement Level Descriptors) affirms that the original intent of the source product was maintained without compromise or any reduction to the expectation.

Research design:

Each grade and content area underwent the same process for review. The process is described here.

The process used for this study was "affirmational" which means that reviewers were presented with a final assessment and decisions that were previously made by test developers and reviewed in an initial review cycle in 2014 (See Appendix A). This process was chosen because, based on the results of the initial study conducted in 2014, the assessment items and Essentialized Standards have been in use in Oregon schools since that time (2014).

Reviewers representing the field of education as either assessment experts, experts of the population, or knowledgeable about the standards, were invited from educational fields around the state for their expertise in either special education, assessment, or Oregon's content standards. Individuals were invited to a training session in which they were provided with background information on the study, the assessment, and Oregon's need for an objective review, and were assigned several questions (entitled "evaluation questions") in which they were asked to apply their professional judgment to the materials they were provided, to evaluate each of the relationships between and among developed components.

The evaluated in this study were (a) the assessment items, (b) the Essentialized Standards, and (c) the Achievement Level Descriptors. For each of these critical materials in the assessment of Oregon's

students, reviewers were asked to evaluate the general / perceived relationship (alignment or link) to a verified source. In the case of the Essentialized Standards, the verified source was Oregon's content standards. In the case of the items, the verified source was the Essentialized Standard. In the case of the Achievement Level Descriptors, the verified source was the Essentialized Standard.

The study was designed to create a linear series of conclusions that combine to support the final statements that the assessment items and the decisions made using the assessment results are in line with the spirit of Oregon's source content standards, and meet the expectations of the *Every Student Succeeds Act* (ESSA, 2015), in which states are responsible for measuring academic achievement as measured by proficiency on annual assessments for all public schools in the state.

To establish this line of reasoning five evaluations were designed to answer five questions.

1. Were the decisions to select and deselect standards for this population's assessment sound?
2. Did the Essentialized Standards, as written, demonstrate an appropriate link to the original source standard (whether CCSS or ORSci/NGSS)?
3. Did the items written align directly with the Essentialized Standard for the grade, content area? Were they free of bias? Were they accessible to students in this population?
4. Did the Essentialized Standards, as written, demonstrate an appropriate alignment with the Achievement Level Descriptors as written?
5. Did the items align overall, to the Achievement Level Descriptors (a one-time confirmation)?

Participants

The review was conducted by expert reviewers with professional backgrounds in either Special Education (the population), Assessment, or in Oregon's adopted content standards. Reviewers were assigned to review grade-level items relative to their experience and expertise. In all, 39 reviewers participated. Thirty-four (34) participated in all 5 evaluations: thirteen (13), for the English Language Arts review, fifteen (15) for the Mathematics review, and six (6) for the Science review. All participants were assigned to at least one specific content area as shown in Table 1. Note: Four individuals were assigned to two areas of review.

The thirty-nine individuals who participated in the study had a robust legacy of experience in the field and in the state. Participants represented 25 unique school districts across the state representing both urban and rural perspectives. All 39 of the individuals participating in the study held current teaching licenses. Two individuals also held administrative licenses. Years of experience in their area ranged from 3 – 30 years of experience with an average of 17 years of experience. (Mode = 11 years, Median = 16 years). One individual indicated 50 years of experience in the field. Three of the 39 individuals held a Bachelor's degree only. Thirty-six held a Bachelor's degree and at least one Master's degree. Two held a Bachelor's degree, at least one Master's degree, and a doctoral degree. Fourteen (36%) of the individuals identified as experts in a specific Content area and 25 (64%) of the individuals identified Special education as their primary area of expertise. Participant experience and background is summarized in Appendix B.

Table 1: Alignment Study Participants

Grade	English Language Arts	Mathematics	Science
3	3[*1]	2	
4	3	2	
5	2[*1]	2	3
6	2	3	
7	3[*1]	3	
8	2	3[*1]	3[*2]
11	2[*1]	3[*1]	2
Total	17 (13)	17 (15)	8 (6)

Note: Asterisk indicates the number of individuals who conducted Evaluation 3, but who did not conduct evaluations 1,2,4 or 5.

Materials

Participants conducted the review using *two* primary source materials for the 5 evaluations. (1) An Excel spreadsheet with side-by-side columns that provided the non-secure information under review (for evaluations 1, 2, 4, and 5), and (2) the Distributed Item Review platform (DIR) that contained the secure information under review (for evaluation 3). Materials for evaluations 1, 2, 4, and 5 are described separately in this section though they were presented to reviewers as columns on a single Excel spreadsheet.

Evaluation 1 Materials: The Excel spreadsheet contained a generated list of all Source standards (CCSS and ORSci/NGSS). Standards omitted from the Source standards and not selected for the development of the OREXt because they were inaccessible for the population were identified by red coloring. Standards omitted from the source standards because their content was included in another Essentialized Standard that was selected for development of the OREXt were marked in green. These were reviewed and approved as appropriately selected based on reviewers' knowledge of the population and of the content area.

Evaluation 2 Materials: The Excel spreadsheet contained a generated list of all Source standards (CCSS and ORSci/NGSS) formatted in columns next to Oregon's Essentialized Standards. These were compared to the source standards evaluated in Evaluation 1 above.

Evaluation 3 Materials: Reviewers were provided with access to the Distributed Item Review (DIR) platform which is an electronic platform designed to allow participants to review and comment on developed items in comparison to other materials (in this case, the Essentialized Standards) in a secure

environment. Reviewers reviewed items for bias and accessibility, and compared them to the Essentialized Standards noted in Evaluation 2.

Evaluation 4 Materials: The Excel spreadsheet contained three levels of the Achievement Level Descriptors, that were developed by test developers. These were compared to the Essentialized Standard.

Evaluation 5 Materials: The Excel spreadsheet contained an empty column and row for reviewers to mark their agreement with the alignment between the Achievement Level Descriptors and the items.

A sample of a subject area Excel Spreadsheet is included in slide 8 of the presentation used to describe the 5-evaluation process is included in Appendix C.

EVALUATION 1

Evaluation 1: Evaluate the deselection of standards by grade.
Overarching Question:

- Were the "right" standards included in the development of the assessment?

Inference:

- The de-selection of standards for omission in Oregon's Extended Assessment were conducted rationally. The final scope of content standards is justifiable for the population for the subject area.

Measure: Average Reviewer Agreement of Content Inclusion.

- Agreement by reviewer and across reviewers to test developer's decisions on content inclusion: Low Agreement (.50 - .64), General Agreement (.65 - .84), and Strong Agreement (.85 – 1.00).

Test Development Process: Inclusion and Exclusion of Standards.

In the development of the Essentialized Standards, developers reviewed all standards and made one of three decisions regarding each standard:

- *Use* the standard (we will essentialize this standard. It is instructionally critical, instructionally prioritized, and accessible to students in this population);
- *Exclude* the standard (this standard is not instructionally critical, this standard is not instructionally prioritized, this standard is not accessible to students in this population, this standard appears in its entirety elsewhere among these standards) OR
- *Combine the* standards (this item is covered by another Essentialized Standard and to include it would result in redundancy).

Reviewer Process: Were the right standards included in the development of the assessment?

Evaluation 1 was conducted via review of the full selection of standards. Reviewers were provided with the wording of the source standards and the wording of the Essentialized Standards in a side-by-side

format. Reviewers were asked to use their professional judgment (expertise and knowledge) to confirm or refute the original decisions made (by test developers) to *use, exclude*, or *combine* standards when creating a list of Essentialized Standards for test development. Reviewers were asked to agree (yes) or disagree (no) with the inclusion, exclusion, or combination. In all evaluations, reviewers were encouraged to leave specific comments to elaborate on their response beyond the yes or no option. An overall average rate of agreement ("yes" responses) for each individual was calculated to indicate the overall strength of inclusion. An overall rate of agreement (by standard) across individuals was also conducted. Results of evaluation 1 (Inclusion) are shown in Table 2.

Evaluation 1: Conclusions.

Overall, reviewers agreed with the test developers' decisions to include and exclude standards from essentialization or inclusion in the assessment. Across all subject areas and grade levels, average reviewer agreement with the selection/inclusion the standards for the development of the test ranged from .82 to 1.00 agreement (in the range of general to strong agreement). ELA and Science both showed strong agreement (ranging from 0.96 – 1.00 and 0.98 – 1.00 respectively), and Mathematics ranged from general to strong agreement (0.82 – 1.00). No instances of low, or below low inclusion.

Table 2: Evaluation 1: Average Agreement on Inclusion of Standards

Grade (Participants)	ELA Evaluation 1: Average (SD)	Math Evaluation 1: Average (SD)	Science Evaluation 1: Average (SD)
3	0.99 (.08)	0.98 (.09)	
4	0.96 (.11)	0.82 (.24)	
5	1.00	0.99 (.09)	1.00
6	0.96 (.14)	0.98 (.16)	
7	0.98 (.08)	0.93 (.14)	
8	0.96 (.14)	0.89 (.21)	0.98 (.13)
11	1.00	1.00	0.99 (.08)

EVALUATION 2

Evaluation 2: Evaluate the (strength of the) link between the source content standards (CCSS, ORSci/NGSS) used for the general population and the Essentialized Standards developed for students with significant cognitive disabilities.

Overarching Question:

- What is the strength of link between the Essentialized Standard and the source standard?

Inference:

- The process of essentializing a given Source Standard did not fundamentally or critically alter the knowledge or skill set intended by the source standard.

Measure: Average Reviewer Agreement of Strength of Link between Source Standards and Essentialized Standards.

- Average strength of link identified by reviewer and across reviewers between source content standards and Essentialized Standards were calculated. Ranges across reviewers were evaluated as follows: Low Link on average in a subject area by grade was considered in the range of 1.00 – 1.29, sufficient Link on average was considered in the range of 1.30 – 1.69, and Strong Link on average was considered in the range of 1.70 – 2.0.

Test development process. Creating Essentialized Standards.

The Essentialized Standards were developed specifically to meet the needs of students with the most significant cognitive disabilities. To meet the expectations of the ESSA, test developers of the Extended Assessment spent over five years perfecting the process of essentializing standards, i.e. the process of identifying the critical components (the essence) of a given standard that all students should know, and removing complex expectations that hamper accessibility. This is achieved standard by standard by reducing breadth (the number of standards) and depth (the scope of each standard) while maintaining the basic integrity of the standard as appropriate for the population and grade. The full process of essentialization is explained in Appendix D, the Essentialized Assessment Framework User Guide 2015 – 2016.

Oregon's Essentialized Standards were developed to provide variety of implementation in each standard at three possible levels of complexity: Low, Medium, or High. The parameters of low, medium, and high vary by subject area but predominantly impact the depth of the standard, by varying elements within the item to manipulate the complexity of the standard (e.g., the number of words, letters, or sentences presented as part of an item; the magnitude of the values, or the complexity of the images used in a problem; or the types of examples used in an array).

Reviewer Process. What is the strength of the link between the source standard and the Essentialized Standard?

As with Evaluation 1, for Evaluation 2, reviewers were provided with the source standard (either CCSS or ORSci/NGSS) and were asked to use their professional judgment, accompanied by a close review of the standards, to determine whether the Essentialized Standards were linked to the source standard, and to what degree. Reviewers were asked to rate the strength of the link as either "no link", "sufficient link", or "strong link". (0) No link -- indicates that the reviewer found no defensible connection between the content in the Essentialized Standard and the content in the source standard. (1) Sufficient link – indicates that there is a connection between the content in the Essentialized Standard and some aspect of the source standards. (2) Strong link indicates that the connection between the Essentialized Standard and at least one aspect of the source standard is obvious and clear.

For the purpose of this evaluation both a sufficient and a strong link were considered adequate to establish a link for the population.

Evaluation 2: Conclusions.

Ratings were compiled for individual reviewers. The average link across all standards was calculated by individual reviewer and across all reviewers by grade. When averaged across reviewers, 1.00-1.29 was considered in the low range, 1.30 – 1.69 was sufficient, and 1.70 – 2.0 was strong. Table 3 shows the average of reviewers' Evaluation 2 by subject and grade. Overall, the average links between Essentialized Standards and source standards ranged from 1.5 (sufficient link) to 1.9 (strong link). English Language Arts linkages ranged from 1.5 (Grade 4) to 1.90 (Grade 5), Mathematics linkages ranged from 1.6 (Grade 5) to 2.0 (Grade 11), and Science linkages ranged from an average of 1.8 (Grade 8) to 1.9 (Grade 5). Average reviewer evaluations of the link between the Essentialized Standards and the source standards indicated that standards were considered sufficiently to strongly linked on average in all grades and subjects. No instances of low or below low links.

Table 3: Average Strength of Link between Essentialized Standard and source standard

Grade	ELA Evaluation 2	Math Evaluation 2	Science Evaluation 2
3	1.9 (.31)	1.8 (.25)	
4	1.5 (.37)	1.7 (.46)	
5	1.9 (.44)	1.6 (.37)	1.9 (.13)
6	1.8 (.28)	1.9 (.36)	
7	1.7 (.32)	1.7 (.25)	
8	1.9 (.33)	1.5 (.39)	1.8 (.41)
11	1.8 (.41)	2.00	1.8 (.50)

EVALUATION 3

Evaluation 3: Evaluate the alignment between the Essentialized Standards and the items. Review for bias in items and review for accessibility of items.

Overarching Questions:

- Is there strong alignment between the Essentialized Standards and the items that were developed based on those standards?
- Did the process of writing an item fundamentally or critically alter the integrity of the Essentialized Standard?

Inference:

- The items written for this grade and subject area will not unduly advantage or disadvantage one student over another based on life experiences that are exclusive or atypical, will be accessible to students with significant cognitive disabilities, and are adequately linked to the Essentialized Standards.

Measure 1: **Average Reviewer Agreement of Strength of Alignment between Essentialized Standards and Individual Items.**

- Strength of the alignment between each Essentialized Standard and each test Item as averaged by and across reviewers: No Alignment on average, Sufficient Alignment on average, Strong Alignment on average. Low Alignment on average in a subject area by grade was considered in the range of 1.00 – 1.29, sufficient Alignment on average was considered in the range of 1.30 – 1.69, and Strong Alignment on average was considered in the range of 1.70 – 2.0.

Measure 2: Average Reviewer Agreement of Item's Accessibility for students with significant cognitive disabilities.

- Agreement by reviewer and across reviewers that the item, as written, is accessible to students with the most significant cognitive disabilities: Low reviewer Agreement, General reviewer Agreement, Strong reviewer Agreement. Low Agreement (.50 - .64), General Agreement (.65 - .84), and Strong Agreement (.85 – 1.00).

Measure 3: Average Reviewer Agreement of Item's Freedom from Bias.

- Agreement by reviewer and across reviewers that the item, as written, is free from bias: Low reviewer Agreement, General reviewer Agreement, Strong reviewer Agreement. Low Agreement (.50 - .64), General Agreement (.65 - .84), and Strong Agreement (.85 – 1.00).

Test Development Process. Developing, aligned, bias-free items.

Items for Oregon's Extended Assessment were developed according to the specifications outlined during the development of the Essentialized Standards. Each item was written with either low, medium, or high complexity as defined by the needs of students in this population. For all items large font, clear white space, plain language, and simple line-graphics, as appropriate, were used to ensure that standards of

universal design were included in addition to the reduction in depth, breadth, and complexity of content expressed by the Essentialized Standards. Over 5,500 items were developed in English Language Arts, Mathematics, and Science. Once items were developed by test developers they were maintained in a secure item pool and reviewed by external educators with expertise in the content area.

Reviewer Process. What is the strength of the alignment between the item and the Essentialized Standard?

For the purpose of educator review, the 2016 operational test items were reduced to the number and proportion (by standard) necessary for each test, with 36 items per assessment. Items were placed in the Distributed Item Review platform (DIR) with the subject, and the item number, along with any associated graphics. In addition, reviewers were provided (in the DIR platform) with test development resources such as a video of the population, the item development description and specifications, Oregon's Accessibility manual that describes the allowable supports a student can have in the state while testing, and training slides. Reviewers were asked to use their professional judgment accompanied by a close review of the items to determine whether the items were aligned to the Essentialized Standards, and to what degree. Reviewers were asked to rate the strength of the link as either "no link", "sufficient link", or "strong link".

(0) No alignment -- indicates that the reviewer found no defensible connection between the content in the item and the content of the Essentialized Standard.

(1) Sufficient alignment – indicates that there is a connection between the content in the item and the content of and some aspect of the Essentialized Standard.

(2) Strong alignment indicates that the connection between the item and the Essentialized Standard is obvious and clear. For the purpose of this evaluation, both a sufficient and a strong link were considered adequate to establish a link for the population. Ratings were compiled for individual reviewers, and counts of 2s, 1s, and 0s were tracked.

Reviewers also used their professional judgment to answer two questions about the items.

1. Is the item free of bias (Yes or No)?
2. Is the item accessible to all students (Yes or No)?

Reviewers were encouraged to provide comments to the items as part of this review. All comments made by the reviewers were provided to the test developers for test improvements, corrections, and refinements. Reviewers submitted over 200 comments as part of Evaluation 3. Comments consisted predominantly of feedback on the items, sometimes in the form of helpful questions, and other times in the form of specific word edits or re-writes. Some examples are provided in Table 4.

Table 4: Evaluation 3 Sample Comments

Subject Grade	Reviewer Comment
ELA (Grade 3)	*Some children will not know what a pet is at this level. Why does it say, "A cat is a pet"? It could just say "which word is cat"*
ELA (Grade 4)	*Items should all be presented in one color*
ELA (Grade 6)	*The ALD for level 4 indicates 3 sentences of 5 or more words. The third sentence in this item only has 3 words.*
Mathematics (Grade 6)	*Item does not refer to number of observations. Could it be used with a different standard?*
Mathematics (Grade 7)	*Remove the word "another"*
Mathematics (Grade 8)	*Consider using scalene triangles as the incorrect choices.*

Evaluation 3: Conclusions.

Measure 1 Alignment: Table 5 shows the average reviewer agreement regarding each item's alignment to the Essentialized Standards. Averaged ratings across reviewers in each grade indicated sufficient to strong alignment between the test items and Essentialized Standards across all three subject areas. In ELA there were no ratings indicating findings of low alignment across reviewers. All but one grade of the ELA review indicated findings of strong alignment across reviewers (Grade 4 – sufficient alignment). In Mathematics there were no instances of low agreement across reviewers, four of the seven grades showed findings of sufficient alignment across reviewers (Grades 4, 5, 7, and 11), and the remainder showed findings of strong alignment. In Science two of the three grades showed findings of sufficient agreement (Grades 5 and 8), and the third showed findings of strong alignment. No instances of low or below low alignment.

Table 5 Evaluation 3: Average Strength of Alignment Rating

Grade	ELA Alignment (sd)	Math Alignment (sd)	Science Alignment (sd)
3	1.88 (.38)	1.85 (.50)	
4	1.62 (.51)	1.32 (.77)	
5	1.89 (.36)	1.64 (.56)	1.50 (.54)
6	1.85 (.36)	1.81 (.48)	
7	1.73 (.49)	1.58 (.74)	
8	1.86 (.42)	1.79 (.53)	1.33 (.56)
11	1.97 (.17)	1.52 (.50)	1.89 (.32)

Measure 2: Accessibility: Table 6 shows the average reviewer agreement regarding each item's accessibility for the population of students with the most significant cognitive disabilities. Averaged scores across reviewers in each grade indicated strong accessibility of items across all three subject areas with one exception indicated in Grade 3 ELA. This reflected the review of one individual whose concerns (predominantly regarding the inaccessibility of paper and pencil administration for Deaf students and blind students) swayed the average significantly. This individual rated all items as inaccessible and biased noting: "*This format does not provide a method for all SPED students to access the test. For a student who may be blind, deaf, have poor fine motor skills, and an inability to orally verbalize, this test does not appear to support these possible student needs. An improvement for including more students, may be to consider having a computer based test. A computer based test may offer accessibility to a wider range of students, and they are highly engaging. General Education students have this method of test taking available."* No instances of low or below low accessibility.

Table 6 Evaluation 3: Average Agreement Regarding Item Accessibility

Grade	ELA Accessible (sd)	Math Accessible (sd)	Science Accessible (sd)
3	0.67 (.47)	1.00 (0)	
4	0.98 (.14)	0.93 (.26)	
5	1.00 (0)	1.00 (0)	0.93 (.26)
6	1.00 (0)	0.99 (.10)	
7	0.98 (.14)	0.92 (.28)	

8	1.00 (0)	0.90 (.30)	0.93 (.26)
11	1.00 (0)	0.98 (.14)	0.93 (.26)

Freedom from Bias: Table 7 shows the average reviewer agreement regarding each item's freedom from bias for students with the most significant cognitive disabilities. Averaged scores across reviewers in each grade indicated strong freedom from bias across all three subject areas with only one exception indicated in Grade 3 ELA. This reflected the review of one individual whose concerns predominantly centered around the accessibility of test as a paper and pencil administration (as opposed to a computer based test) as noted previously. No instances of low or below low freedom from bias.

Table 7 Evaluation 3: Average Agreement Regarding Items Free from Bias

Grade	ELA Bias (sd)	Math Bias (sd)	Science Bias (sd)
3	0.65 (.48)	0.96 (.20)	
4	0.96 (.19)	0.89 (.32)	
5	1.00 (0)	0.93 (.25)	1.00 (0)
6	0.96 (.20)	0.98 (.14)	
7	0.97 (.17)	0.98 (.14)	
8	1.00 (0)	1.00 (0)	0.97 (.17)
11	1.00 (0)	0.96 (.19)	0.97 (.17)

EVALUATION 4

Evaluation 4: Evaluate the alignment between the Essentialized Standards and the (alternate) Achievement Level Descriptors.

Overarching Questions:

- What is the strength of the alignment between the Essentialized Standards and the Achievement Level Descriptors?
- Does the achievement inference/claim stated in the Achievement Level Descriptor adequately convey the skillset assessed by the Essentialized Standard?

Inference:

- The Achievement Level Descriptors (ALDs) for each subject and grade level are appropriately aligned with the content being measured and can be used to describe the standards tested.

Measure (for each ALD level): Average Reviewer Agreement of Alignment between the Essentialized Standard and the Achievement Level Descriptor (by Level).

- Agreement by reviewer and across reviewers that the Achievement Level Descriptor (ALD), as written by test developers, aligns with the Essentialized Standard as written by test developers: Low reviewer Agreement, General reviewer Agreement, Strong reviewer Agreement. Low reviewer Agreement (.50 - .64), General Agreement (.65 - .84), and Strong Agreement (.85 – 1.00).

Test Development Process: Developing Achievement Level Descriptors.

The Achievement Level Descriptors of a test system provide the qualitative description of the test's claim and the students' ultimate achievement on the test. An Achievement Level Descriptor is intended to accurately describe both what the test is measuring and whether, or to what extent the student has achieved that capacity. Similar to assessment items, Achievement Level Descriptors are typically developed via committee process, by experts in the field, in assessment and familiar with the population. The Achievement Level Descriptors for Oregon's Extended Assessment reflect four categories of achievement. The categories are unnamed and are simply numeric, but roughly reflect low (1) to high (4) achievement. A score of 1 = no achievement, 2 = inconsistent or partial achievement not quite meeting expectations, 3 = proficient, sufficient to meet expectations, and 4 = exceptionally proficient: achievement that exceeds the requirements in the standards. Students achieving at or around a level 2 are students with lower consistency in their responses, or students who predominantly are successful on the low difficulty items. Students achieving at a four or above, are likely students who are able to respond to more of the items (including those items with high difficulty).

Is there alignment between the Essentialized Standard and the Achievement Level Descriptor at each level?

Reviewers were provided with (1) the Essentialized Standard, (2) the Essentialized Standard's low, medium, and high parameter guidance for item development, and (3) the Achievement Level Descriptors that were developed based on those two elements. Reviewers were asked to review the Achievement Level Descriptor at each level, by comparing it to the Essentialized Standard (using the item-development considerations of the low, medium, high parameters) and indicate by stating yes or no, whether the Achievement Level Descriptor at that level, could be said to align with the Essentialized Standard. Yes = the Achievement Level Descriptor (at this level) accurately describes the achievement of a student who has (inconsistent, proficient, exceptionally proficient) understanding of the standard. For the purpose of the alignment study an overall average agreement for each individual was calculated. Results of Evaluation 4 are shown in Table 8.

Evaluation 4: Conclusions.

Table 8 shows reviewers' determinations of Evaluation 4 by subject and grade level. Reviewer responses were averaged across all reviewers. ELA reviewer agreement was in the strong range overall, ranging from .85 – 1.0. Math reviewer agreement was in the general to strong range overall with the exception of one grade reflecting low agreement at .52 (Grade 4) .68 - 1.00. Science reviewer agreement was in

the general to strong range .83 (Grade 11) - .98 (Grade 5). One instance of low alignment (Grade 4 Math).

Table 8: Achievement Level Descriptor Alignment to Essentialized Standard by Level

Grade	ELA Evaluation 4 (ALD Level 2)	ELA Evaluation 4 (ALD Level 3)	ELA Evaluation 4 (ALD Level 4)	Math Evaluation 4 (ALD Level 2)	Math Evaluation 4 (ALD Level 3)	Math Evaluation 4 (ALD Level 4)	Science Evaluation 4 (ALD Level 2)	Science Evaluation 4 (ALD Level 3)	Science Evaluation 4 (ALD Level 4)
3	0.98 (.09)	0.98 (.09)	0.98 (.09)	0.80 (.25)	0.80 (.25)	0.82 (.25)			
4	0.84 (.17)	0.83 (.17)	0.80 (.17)	0.52 (.10)	0.52 (.10)	0.52 (.10)			
5	0.90 (0.30)	0.95 (0.22)	0.81 (0.40)	0.78 (.25)	0.76 (.26)	0.76 (.26)	0.98 (.09)	0.98 (.09)	0.93 (.14)
6	1.00	1.00	1.00	0.96 (.19)	0.96 (.19)	0.95 (.20)			
7	0.88 (0.16)	0.85 (0.17)	0.85 (0.17)	0.68 (.13)	0.68 (.13)	0.68 (.13)			
8	0.86 (.23)	0.86 (.23)	0.86 (.23)	0.84 (.24)	0.84 (.24)	0.82 (.30)	0.92 (.28)	0.92 (.28)	0.92 (.28)
11	1.00	1.00	1.00	0.96 (.20)	1.00	1.00	0.83 (.24)	0.83 (.24)	0.83 (.24)

EVALUATION 5

Evaluation 5: Evaluate the alignment between the Achievement Level Descriptors and the Items.

Process:

Overarching Questions:

- What is the strength of the alignment between the Essentialized Standards and the Achievement Level Descriptors?
- Does the achievement inference/claim stated in the Achievement Level Descriptor adequately convey the skillset assessed by the Essentialized Standard?

Inference:

- The Achievement Level Descriptors (ALDs) for each subject and grade level are appropriately aligned with the items as written and can be used to describe the skillset achieved by the student who is successful at the items.

Measure: Overall Reviewer Agreement of Alignment between the all items and all Achievement Level Descriptors.

- Agreement by reviewer and across reviewers that the Achievement Level Descriptor (ALD), as written by test developers, aligns with the Item as written by test developers: Low reviewer Agreement, General reviewer Agreement, Strong reviewer Agreement.
- Low agreement with test developers will range from .5 - .64, agreement with test developers is considered in ranges from .65 - .84 and high agreement with test developers is considered in ranges .85 – 1.0.

Test Development Process: Developing Achievement Level Descriptors linked to item parameters and standards.

During the development of the Essentialized Standards, test developers included a summary statement aligned to each individual standard, which indicated the characteristics that would be present in a low, medium, or high item. These summary statements were subsequently adopted to inform the development of the Achievement Level Descriptors described in Evaluation 4 of this study. This explicit alignment between item development and Achievement Level Descriptor development, served as justification for the method followed in Evaluation 5 described below.

Is there alignment between the Extended Assessment items and the Achievement Level Descriptors?

Reviewers were asked to conduct each of the five evaluations in sequence -- the information and decisions made in each preceding evaluation thus informing subsequent decisions. In the case of Evaluation 5, by this point in the study, reviewers had spent time: evaluating the content of the Essentialized Standards (in Evaluations 1 and 2), the items (in Evaluation 3), and the Achievement Level Descriptors (in Evaluation 4). Based on this familiarity with the development, the standards, the items, and the Achievement Level Descriptors, reviewers were asked to then finally provide a single "summary affirmation" of their perceived alignment between the Achievement Level Descriptors and the Items. This one-time affirmation was different from first 4 Evaluations as it did not require the reviewers to provide line item data (by item or standard, as was the case with the prior evaluations). A single response of "yes" or "no" at the conclusion of the review was solicited to affirm that "The ALDs created for this subject and grade level align reliably and consistently with the corresponding Oregon Extended assessment items."

Evaluation 5: Conclusions

Table 9 shows reviewer average agreement by grade and subject. Responses of "yes" (scored as 1), were averaged across reviewers by grade in each subject area. Reviewers were able to consistently affirm the overall alignment of the items to the Achievement Level Descriptors. Given the range of reviewers per review category, low average agreement was .5. ELA average responses ranged from .5 (one instance in Grade 8) to 1. Mathematics average responses ranged from .5 (two instances: Grades 3 and 4) to 1. Science average responses were consistent at an average of 1 across all grades. Three instances of low alignment (ELA Grade 8, Mathematics Grades 3 and 4).

Table 9: Evaluation 5: Average Alignment between Items and Achievement Level Descriptors

Grade	ELA Evaluation 5 (participants)	Mathematics Evaluation 5 (participants)	Science Evaluation 5 (participants)
3	1 (2)	.5 (2)	
4	.67 (3)	.5 (2)	
5	1 (1)	1 (2)	1 (3)
6	1 (2)	1 (3)	
7	1 (3)	1 (3)	
8	.5 (2)	1 (2)	1 (1)
11	1 (1)	1 (2)	1 (2)

OVERALL

Overall: Evaluate overall reviewer confidence based on the reviews they conducted on each of the 5 evaluations.

Overarching Questions:

- What is the strength of the assessment as suggested by a summary of all of the reviews conducted?

Inference:

- Reviews that consistently result in reviewer agreement/affirmation with development decisions (i.e scores of "yes, 1, or 2"), indicate a high amount of reviewer confidence in the decisions made regarding development and use of the assessment.
- This summary of professional confidence in the decisions can be considered an indicator of validity of the assessment.

Measure: Overall average of reviewer agreement/affirmation of each source standard and associated decisions by grade and content.

- Average agreement/affirmation by reviewer and across reviewers across all evaluations. Highest possible score is 1.
- Low agreement/low confidence with test developers' decisions will range from .5 - .64, agreement/general confidence with test developers' decisions will range from .65 - .84 and high agreement/strong confidence with test developers' decisions will range from .85 – 1.0.

Combined Score Conclusion:

The evaluation conducted at each stage of the review required a simultaneous knowledge of and focus on several critical pieces of background information:

- The student with significant cognitive disabilities and the potential range of needs: What is the range of needs that could be exhibited by a student in this population?
- The source standard: What is the source standard?
- The inference or intended expectation of that standard: What is the fundamental educational goal of the source standard?
- The classroom presentation of the standard during instruction: In what ways might the most experienced teachers present this standard during instruction?
- The assessment presentation of the standard during testing (as an item): In what ways can an item manifest/embody the standard during an assessment – is there better way? and
- The intended claim that would be made based on a successful score on the item: What will an educator claim if a student in this population is successful at this item?

Reviewers were selected for their expertise in a variety of these areas and were relied upon for their ability to consider each component objectively. During the review, reviewers made each decision independently so that conclusions from each evaluation could be assembled to form an overall impression or suggestion regarding the validity of the ORExt.

In addition to the individual conclusions drawn following each of the 5 evaluations noted in this report, a summary score based on evaluations 1, 2, and 4 was calculated to provide a quantitative description of the overall assessment by grade and by subject. To accomplish this, a simple average score was calculated to capture the general average sentiment expressed across standards, across reviewers within a content area by grade. See Table 10 for a summary of the overall confidence as calculated by average of reviewer decisions across the assessment. These scores, which range from .71 – 1.0 will be included in a validity argument to suggest reviewers' overall confidence in the decisions made by the developers of the test at the various decision-points of the assessment. Using the same criteria used throughout the study, ELA was in the general confidence to strong confidence range from .71 - .97, Mathematics similarly (.77 – 1.0) and Science consistently fell into a strong confidence range (.90 - .97). No instances of overall low confidence.

Table 10: Overall Confidence as Calculated by Average Review

	ELA	Mathematics	Science
3	.97	.85	
4	.82	.78	
5	.71	.80	.97
6	.93	.85	
7	.84	.81	
8	.94	.77	.92
11	.97	1.0	.90

Findings and possible inferences:

First, reviewers were asked to conduct an affirmational review of the rationale used by test developers to omit certain content standards. This finding was used to infer that the final standards selected for inclusion or omission in Oregon's Extended Assessment were chosen rationally and that the final scope of content standards can be considered justifiable for the population for the subject area.

Conclusion: This review, with a lowest average rate of .82 (on a scale of 1), permits the inference: **the scope of the standards selected for translation to Essentialized Standards were rationally selected**. None of the standards de-selected (for inaccessibility or for being covered elsewhere) were strongly identified for re-inclusion, nor were identified as a critical hole for this population of students.

Second, reviewers were asked to identify the strength of the link between the source standard and the Essentialized Standard. This finding was used to infer that the process undertaken to essentialize a given Source Standard did not fundamentally or critically alter the knowledge or skill set intended by the source standard for this population of students (further confirming that the content selected for assessment is comparable).

Conclusion: This review, with a range of 1.5 – 1.9 (on a scale of 2) permits the inference: **the Essentialized Standards were found to link sufficiently to the source standards** on average beyond the "sufficient" average of 1.0.

Third, reviewers were asked to identify the strength of the alignment between the Essentialized Standards and the items and to review the items developed using the Essentialized Standards for bias, and accessibility. The finding from this review was used to infer that the items written for this grade and subject area (using these Essentialized Standards) were adequately linked to the Essentialized Standards were free from bias, and were accessible to students with significant cognitive disabilities.

Conclusion: The alignment review (1.32 – 1.89), accessibility review (.67** – 1.0), and freedom from bias review (.65** – 1.0) all permit the inference that **the test items indicate a relationship with the source standards, the test items are not overly biased towards or against any**

particular group of individuals, and the test items are written such that the content and intent can be accessed by students with the most significant cognitive disabilities. (**Note: this range was skewed by feedback from one reviewer --ELA-Grade 3 – whose comments were noted in this study. Removing that individual's comments would result in a range of .90 – 1.0 accessibility range and .89 – 1.0 freedom from bias range respectively.)

Fourth, reviewers were asked to review the statements used to describe student achievement on the test (the Achievement Level Descriptors) and their alignment to the Essentialized Standards that the students were tested on. The finding from this review was used to infer that the skills and achievements described by the Achievement Level Descriptors for each subject and grade level are aligned with the content standard being measured.

Conclusion: The reviews ranging from .68* – 1.0 permit the inference that **the descriptions made regarding student skillset are an accurate reflection of the standards from which the assessment was developed** at all three levels evaluated. (*One outlier for ELA-Grade 4 provided a review of a .52 average).

Fifth, and finally, reviewers were asked to review the alignment of the Achievement Level Descriptors to the items. The finding from this review was used to infer that each item in the developed assessment(s) was appropriately aligned to its associated Achievement Level Descriptor (further confirming that decisions made using this test were aligned with the intent of the source standard).

Conclusion: Fourteen of the seventeen grade-level reviews resulted in an average reviewer range of .67 – 1.0 indicating an appropriate alignment between ALDs and the items as written. This review permits the inference that, overall, the Achievement Level Descriptors are accurate reflections of the items. In three instances (Mathematics-Grades 3 and 4, and ELA-Grade 8) the average alignment by reviewer was .5 (indicating that one of the two individuals in that category did not agree that the items and ALDs were aligned).

Assumptions

The strength of the affirmations made in this study presupposes several critical elements that were not part of this study but that are supported by other studies or reviews that have been or will be conducted over the course of the assessment's existence.

1.) **That participants (reviewers) were truly experts in their field.** The training provided was not sufficient to guarantee expertise in each area necessary for a solid review. Pre-test or screening of reviewers with an in-depth survey that identifies experience, training, and a general knowledge base in some of the critical areas is recommended for future studies.

2) That the assessment is administered with fidelity and integrity: This study assumes that items are administered with fidelity. Though alternate assessment administrators are trained annually by state trainers in the assessment, a fidelity of implementation study would add critical objective dimension to this data and would add further strength to the claims made by the Achievement Level Descriptors.

3) **That the students are being instructed with curriculum that is also aligned to the same content standards that are assessed:** When making any claims about Achievement Level Descriptors a critical link between item and achievement is instruction. This was not part of the study.

4) **That the source standards (CCSS and ORSci and NGSS) are assessing content that is meaningful, and sufficient for all students:** All assumptions, inferences, and arguments are linked to the national content standards. To make any claims of validity about this study, require the ultimate presumption that the content standards selected nationally for all students, are the appropriate basis for this small, heterogeneous group of students as well. This study does not delve into the many other needs that would encompass the valid assessment of this population of students.

5) **That the outcomes of the test are as expected:** All relationships examined for this study are internal to the test and to its development. No external criteria were included to verify the relationship of the achievement level descriptors to actual student outcomes (see assumptions above).

Discussion and Conclusion

As Kane has been quoted by alternate assessment experts and psychometricians over the decades, "the interpretation [of test scores] involves an argument leading from the scores to score based statements or decisions, and the validity of the interpretation depends on the plausibility of this interpretive argument" (1992). The argument therefore must be sound, consistent, plausible, and defensible.

In test design and development, particularly those activities associated with alternate assessments, the network of inferences extends well prior to the generation of assessment scores, and is also a significant component of the validity of the development process. In addition to evaluating the network of arguments between test score and test use, we must investigate the network of assumptions that exist across the full continuum of development. A good argument will typically examine each potential weak point or counter-argument prior to generating a conclusion. The decision to build and implement an assessment for students with the most significant cognitive disabilities that is based on standards that were created for the general population requires a series of inferences and interpretive arguments many of which were undertaken in this study. The ratings for each evaluation were based on reviewer averages in which anything over .65 (on a 1.0 scale) or 1.3 (on a 2.0 scale) were considered in the acceptable range of average ratings across reviewers. In all cases, as reviewers considered their responses, they provided specific feedback and guidance that will be provided to the test developers to inform their future iterations and refinements of the OExt Assessments. As has been demonstrated in the previous pages, the reviewers were able to affirm the proposed inferences across all 5 evaluative arguments posed in this study in OExt English Language Arts, Mathematics, and Science. This is further emphasized by similar findings in a small-scale study conducted in 2014.

No instances of emphatic shared concerns surfaced across reviewers for any one standard or item. In other words, reviewers agreed frequently and consistently on the relationships that were working between the components. While areas of weakness were identified in all reviews by individual reviewers, there were no consistent review responses that identified areas that were not working overall. The areas of weakness that were identified during this study were typically unique to a reviewer. In most cases, feedback on relationships (links, alignment, and other observations) was explicit enough to inform ongoing development of OExt.

Note: The initial group size was impacted by timing and weather. There were two areas that had only one individual as a reviewer. Typically, average reviewer ratings identify areas of concern under .65 as "low". Because of this small n, however, the findings of this study suggest that evaluations that fell

below an average of .75 for scores out of 1 (1.5 out of 2) should undergo additional review (See Table 11). Two areas in particular are identified for targeted additional review Mathematics Grade 4 and ELA Grade 3.

Table 11: Evaluation Areas scoring under 65% (.65) average

	ELA	Mathematics	Science
3	Accessibility (.67) Bias (.65)	ALD/Item (.5)	
4	ALD/Item (.67)	EAF item Alignment (1.32) ALD/Standard (.52) ALD/Item (.5)	
5	Overall (.71)		
6			
7		ALD/Standard (.68)	
8	ALD/Item (.5)		EAF Item Alignment (1.33)
11			

References

Designing Content Targets for Alternate Assessments in Science: Reducing Depth, Breadth, and/or Complexity. Gong, B. (2007). Presentation at the web seminar Best Practice and Policy Consideration in Science Teaching and Testing for Students with Significant Cognitive Disabilities. Retrieved May 27, 2017, from http://www.nciea.org/publications/FINALGongSciAA_BG07.pdf

Every Student Succeeds Act (2015). S.1177. Retrieved May 27, 2017 from https://www.congress.gov/114/plaws/publ95/PLAW-114publ95.pdf

Works Referenced

Davis-Becker, S., L., & Buckendahl, C. (2013). A Proposed Framework for Evaluating alignment Studies. Educational Measurement: Issues and Practice, 32(1), 23-33.

Flowers, C., Wakeman, S., Browder, D. & Karvonen, M. (2007). *Links for academic learning: An alignment protocol for alternate assessments based on alternate achievement standards*. Charlotte, North Carolina: University of North Carolina at Charlotte.

Kane, M. T. (1992). An argument-based approach to validity. *Psychological Bulletin*, 112, 527-535.

Validity Arguments for Alternate Assessment Systems. Marion, S. (2008). Presentation for the Center for Assessment Reidy Interactive Lecture Series. Retrieved May 27, 2017, from http://www.nciea.org/publication_PDFs/RILS08_SM2_092508.pdf

Selection, Links, Alignment: Establishing a Validity Position for Oregon's Extended Assessments

APPENDIX A: PRIOR STUDY

Oregon Extended Assessment Linkage Study
2014-15

After initially developing the Essentialized Assessment Frameworks (EAF), the research team comprised of three experts in alternate assessment test development and validation conducted a study to examine the linkage between the Essentialized Standards (ES) and parent academic content standards. The purpose of this linkage study was twofold. The first purpose was to determine the appropriateness of the academic content standards (Common Core State Standards [CCSS] for *English Language Arts* and *Mathematics*, and the Next Generation Science Standards [NGSS] and Science Academic Content Standards for Oregon for *Science*) selected for inclusion and exclusion in the ES. The second purpose was to determine whether ES were appropriately linked to their parent academic content standard. In other words, the first purpose examined whether an appropriate breadth of parent academic content standards were represented in the ES, while the second purpose examined the linkage between a given ES and their parent content standard through a one-to-one comparison. Overall, under both of these purposes we gleaned feedback to guide further refinement of the EAF/ES.

Participants. Across the three content areas (English Language Arts, Mathematics, and Science) 11 researchers and educators participated in the linkage study. All participants had at least a Masters degree in education/education research or related field, and had experience teaching and assessing students in K-12 general and special education settings. The participants included six former K-12 educators who either earned or were earning doctorate degrees in Educational Leadership or Educational Psychology, with particular expertise in test development and validation. Prior to reviewing the linkage of the EAF and ES, participants attended a one-hour training session on September 16, 2014, either in-person or through an online webinar. Training slides for the linkage study are shown below.

Research and Study Design. Over September and October 2014, 11 individual participants examined the linkage between the EAF/ES and the respective academic content standards (Common Core State Standards [CCSS] for *English Language Arts* and *Mathematics*, and the Next Generation Science Standards [NGSS] and Oregon Science Standards for Oregon), six in English Language Arts, three in Math, and four in Science, with two reviewers serving in both math and Science. Each reviewer within a given content area reviewed all the ES and academic content standards. Prior to reviewing linkage, reviewers were asked to familiarize themselves with both the EAF/ES and respective academic content standards. Essentialized standards were then displayed in separate grade-level Excel spreadsheets by content area. Reviewers were first asked to whether they agreed that the parent academic content standard should/should not have been included in the EAF/ES (yes/no). Reviewers were then asked to rate the linkage between the ES and parent academic content standard using a 3-point scale (0 = no link, 1 = sufficient link, 2 = strong link). Lastly, reviewers provided comments if they provided either a 'no' or rating of '0', respectively.

Results and ES Refinement. Summary descriptive statistics are displayed in section 2.5 of the 2014-2015 Technical Report narrative. Overall, agreement with standards selected for essentialization and average linkage ratings were very high across all content areas and grades. Representative reviewer comments from each content area are shown below—including comments that the research team used to guide ES refinement.

"Identifying a topic or main idea is not linked with the standard's purpose of measuring opinion expression. This essentialized standard could be better linked by having students identify the opinion expressed within a text (e.g., Bill refused to eat his peas. Does Bill: love peas, hate peas, forgot peas)." – Reviewer 2, ELA

"There's no link to the analyze dialog/incidents piece, perhaps [the ES] could have students identify a feeling or event that happened to a character?" – Reviewer 4, ELA

"The graph for the first choice needs to changed to be more obviously wrong." – Reviewer 2, Math

"The link is sufficient to test the concept. It would be a stronger link if divisors other than 2 were used." – Reviewer 3, Math

"VERY good example of [a content standard] addressed completely in another essentialized standard. Good essentialization - very accessible." – Reviewer 1, Science

"If the interaction is the important part here, maybe the L [parameter] could be 'How do animals use air? How do plants use water?'" – Reviewer 3, Science

Based on reviewer feedback (for example, comments like those displayed above) in the linkage study, the research team, made up of three content area specialists with expertise in alternate assessment test development and validation, edited and refined the ES, including the low, medium and high (L/M/H) difficulty parameters designed to guide later test item development. Editing and refinement of the EAF/ES based on results from the linkage study were completed in December 2015 in preparation for item development in Winter 2015.

Linkage Study Training Slides

Oregon Extended Assessment

Linking Study – Fall 2014

Tuesday, September 16, 2014
4:00 – 5:00 PM
Behavioral Research and Teaching (BRT) – University of Oregon
Oregon Department of Education

BRT
behavioral research & teaching

1

Agenda

- As you are all experienced with the population of Students with Significant Cognitive Disabilities (SWSCDs) and are also familiar with the Essentialized Standards, we do not need to provide training on those topics
- We will focus on the direct tasks at hand
  - We need documentation of the validity of our decision making surrounding standard selection
  - We need documentation that the Essentialized Standards that we developed strongly link to the target standards
- Your judgments help us make the entire process better (instruction, curriculum, & assessment)

BRT
behavioral research & teaching

2

Ultimate Goal

Improving achievement for students with significant cognitive disabilities by linking

- Academic standards,
- Instruction, and
- Assessment

3

behavioral research & teaching

Big Picture

- All students in Oregon are required to demonstrate proficiency on grade-level content standards
- Students with significant cognitive disabilities need to demonstrate progress toward reaching proficiency on grade-level content standards
- Oregon's Extended Assessment is designed to assess the progress of students with significant disabilities toward meeting these content standards

4

behavioral research & teaching

Linking Study Spreadsheets

A	B	C	D	E	F	G
Standard	Common Core Standard	Essentialized Standard	Low Medium High Parameters	Linkage Rating (0 = no link; 1 = sufficient link; 2 = strong link)	Agree with determination that this standard should/should not have been included	Comments

**Math has an extra column between B & C, as there are several sub-standards (a-g). Math folks will need to add a letter to the identifiers in this PPT, but they should be good at adding!*

5

BRT
behavioral research & teaching

Content Standard Selection

Content standards were selected based upon three criteria:

1. This standard is a critical standard to learn in order to be able to access subsequent grade level standards
2. This standard is given more weight instructionally by teachers
3. This standard is accessible for SWSCDs, both in terms of performance match (i.e., cognitive complexity, depth of knowledge, breadth of knowledge) and in terms of sensory requirements (e.g., will be accessible to students with sensory impairments)

Read the Common Core Standard (Column B) and the Essentialized Standard (Columns C & D – including the L/M/H Parameters)

6

BRT
behavioral research & teaching

Task #1

Let us know whether we have included the appropriate grade level standards (and excluded the appropriate grade level standards) based upon these criteria

- If a standard was not included, it is highlighted in red

  (In ELA, we did not include the Speaking & Listening standards, nor the Literacy in History/SS, Science, and Technical Subjects standards due to implementation of the same criteria. Please include a statement at the bottom of the Comment column stating whether you agree with this exclusion or not; if not, please explain why)

- If a standard is highlighted in green, it means that we feel the content was covered by a different Essentialized Standard (and the standard that we feel it links to it is identified)

The fields in each of the spreadsheets you will be given have a column that is pre-populated with "Yes" – you only need to change those that you disagree with to "No" in this column, and then provide us with a rationale in the *Comments* column

7

BRT
behavioral research & teaching

Target Standard: Example

Grade 3 Reading Common Core Standard RL1

- Ask and answer questions to demonstrate understanding of a text, referring explicitly to the text as the basis for the answers.

Standard Selection Criteria

1. Is this standard is a critical standard to learn in order to be able to access subsequent grade level standards

   YES

2. This standard is given more weight instructionally by life skills teachers

   YES

3. This standard is accessible for SWSCDs, both in terms of performance match (i.e., cognitive complexity, depth of knowledge, breadth of knowledge) and in terms of sensory requirements (e.g., will be accessible to students with sensory impairments)

   YES

8

BRT
behavioral research & teaching

Target Standard: Non-example

Grade 11 Math

- M.9_12.A.A_SSE.2.3.b: Complete the square in a quadratic expression to reveal the maximum or minimum value of the function it defines.

Standard Selection Criteria

1. Is this standard is a critical standard to learn in order to be able to access subsequent grade level standards

   NO

2. This standard is given more weight instructionally by life skills teachers

   NO

3. This standard is accessible for SWSCDs, both in terms of performance match (i.e., cognitive complexity, depth of knowledge, breadth of knowledge) and in terms of sensory requirements (e.g., will be accessible to students with sensory impairments)

   NO

BRT
behavioral research & teaching

9

Task #2

- Your second task is to determine the level of linkage between the Essentialized Standard (ES) and the target standard(s) (CCSS for ELA and Math; OR Science and NGSS for Science)
- This is the scale you will use for these determinations (it is also found in row 1 on each spreadsheet)
  - **0 = No Link:** there is no connection between the content in the ES and the content in the target standard(s)
  - **1 = Sufficient Link:** there is a connection between the content in the ES and some aspect of the content in the target standard(s) that is easily recognizable, but not as strong as it could be
  - **2 = Strong Link:** the connection between the content in the ES and the content in at least one aspect of the target standard(s) is obvious and clear.

[*Remember, is it one strand of the standard rope? Then, how strong is the strand*]

BRT
behavioral research & teaching

10

Strong Link

2 = Strong Link: the connection between the content in the ES and at least one aspect of the content in the target standard(s) is obvious and clear.

Target Standard

Essentialized Standard

Strong link, with a few degrees of separation between the standards (not aligned, nor does it include all aspects of the standard – not our goal!)

11

BRT
behavioral research & teaching

Strong Link Example

Grade 5 Science Target Standard

- 5-ESS2-2 Describe and graph the amounts and percentages of water and fresh water in various reservoirs to provide evidence about the distribution of water on Earth. [Assessment Boundary: Assessment is limited to oceans, lakes, rivers, glaciers, ground water, and polar ice caps, and does not include the atmosphere.]

Content: Amounts of water in various reservoirs, not including the atmosphere

12

BRT
behavioral research & teaching

Strong Link Example, cont.

Essentialized Standard, with L/M/H Parameters

- Compare the amount of water in different reservoirs on Earth. L - Restricted to questions about what Earth features that are made of water (i.e., oceans, lakes, rivers, streams) as compared to common objects that aren't (i.e., rock, brick, toy, ball); M - Restricted to questions about what Earth features that are made of water (i.e., oceans, lakes, rivers, streams) as compared to other natural features that aren't (mountains, volcanoes, forest, etc.); H - Restricted to comparing the relative amounts of water in various features of the hydrosphere (i.e., oceans, lakes, rivers, streams, ponds, etc.) using, for example, bar graphs that reflect the relative %s of water in the ocean vs. lakes vs. rivers; or Pacific Ocean vs. other oceans.

- Discussion: This is clearly content that links to the grade level standard. The student is identifying water at the low difficulty range, but then comparing the relative amounts of water in different reservoirs at the high difficulty range.

13

BRT
behavioral research & teaching

Sufficient Link

1 = Sufficient Link: there is a connection between the content in the ES and the content in at least one aspect of the target standard(s) that is easily recognizable, but not as strong as it could be

Target Standard

Essentialized Standard

Easily recognizable connection, but more degrees of separation between the standards

14

BRT
behavioral research & teaching

Sufficient Link Example

Grade 8 Math Target Standard

- M.8.F.2.5. Describe qualitatively the functional relationship between two quantities by analyzing a graph (e.g., where the function is increasing or decreasing, linear or nonlinear). Sketch a graph that exhibits the qualitative features of a function that has been described verbally.

Content: Analyze a graph to determine change (increasing/ decreasing, linear or nonlinear)

15

BRT
behavioral research & teaching

Sufficient Link, cont.

Essentialized Standard, with L/M/H Parameters

- Identify slope as positive, negative, zero, or undefined. L - identify positive slopes 1-3; M - identify negative slopes 4-10; H - identify zero or undefined slopes

- *Discussion:* It can be argued that this is a 2, but it is at the very least a strong 1. The student is indeed comparing functions. They are only linear and they are only in four formats, but it gets at the standard's focus on comparing a relationship between two variables.

16

BRT
behavioral research & teaching

No Link

0 = No Link: there is no connection between the content in the ES and the content in the target standard(s)

Essentialized Standard

Target Standard

These two standards are not the same stuff

17

BRT behavioral research & teaching

No Link Example

Grade 7 Writing

- 7.W3 - 3. Write narratives to develop real or imagined experiences or events using effective technique, relevant descriptive details, and well-structured event sequences. a. Engage and orient the reader by establishing a context and point of view and introducing a narrator and/or characters; organize an event sequence that unfolds naturally and logically.

- Content: Expressing a real or imagined story that engages and orients the reader

18

BRT behavioral research & teaching

No Link Example, cont.

Essential Standard, with L/M/H Parameters

- Identify a number in writing. L – Identify numbers 1-10; M – Identify numbers 11-20; H – Identify numbers 21-40

- *Discussion:* Though this standard is indeed related to writing (so the situation could certainly be worse), it is not about expressing a story in any way. It has to do with identifying the written form of numbers.

19

BRT
behavioral research & teaching

Process Review

1. Read the standards
2. Determine standard selection agreement/disagreement (If you disagree with the standard selection or exclusion, change the "Yes" to a "No" in Column G, and then explain why in the *Comments* column)
3. Rate the Linkage of the ES to the Target Standard/s with a 0, 1, or 2 in Column E (if you rate a linkage as "0" please explain why in the *Comments* column)
4. E-mail your completed spreadsheets to Dan at dfarley@uoregon.edu using the "_DF" filename extension (with your initials instead of mine)
5. We would love to have all of the spreadsheets back by October 10, 2014

20

BRT
behavioral research & teaching

Next Steps/Follow-up

- Questions/Comments Contact: Dan Farley at dfarley@uoregon.edu or 541-525-5780
- Thank you for your time!

21

BRT
behavioral research & teaching

Preliminary Item Alignment, Bias and Accessibility

Subsequent to the finalization of the EAF/ES and item writing, the research team conducted a preliminary alignment and item quality study for the full *English Language Arts*, *Mathematics*, and *Science* item bank. The purpose of the alignment study was threefold. The first purpose was to determine the *degree of alignment* between items and their targeted ES. The second was to determine whether items were free of *bias*. The third purpose was to determine whether items were *accessible* to students with significant cognitive disabilities.

Participants. In total, 53 reviewers participated in the alignment study: 21 in *English Language Arts*, 22 in *Mathematics*, and 10 in *Science*. Reviewers represented 44 school districts from across Oregon, and two were from the Oregon Department of Education. All reviewers had a Master's degree in an education-related field, with 35 reviewers specializing in special education. Participating reviewers took part in a training session, either in-person or online by webinar on November 4, 2014. A member of the research team trained those reviewers who were unable to make the main training through a one-on-one online webinar. Training slides for the alignment study are shown below.

Research and Study Design. Reviewers were assigned to review grade-level items relative to their expertise and experience. For *English Language Arts* and *Math*, three unique individuals reviewed each item in each of Grades 3-8 and 11, with a fourth reviewer serving in Grade 6 for Math. For *Science*, three unique individuals reviewed each item in each of Grades 5, 8, and 11, with a representative from the Oregon Department of Education serving as a fourth reviewer at each grade level. The table below displays the number of items reviewed by each reviewer, at each grade level, in each of the three content areas.

Total Number of Items Reviewed by Content Area and Grade

Grade	Items Reviewed		
	ELA	Math	Science
3	356	286	-
4	368	338	-
5	354	299	495
6	314	351	-
7	315	247	-
8	328	260	504
11	313	299	504

The alignment study was conducted using a secure web-based platform called the Distributed Item Review (DIR) designed to distribute test items to experts across broad geographic regions for the purpose of analyzing them for quality dimensions of alignment, bias, and accessibility/sensitivity. Reviewers rated items in batches of 25-50 items, with the research team reviewing results and concurrently providing feedback. Notably, resources helpful to rating items (i.e., training slides, a video of a representative student

population being administered the Oregon Extended Assessment, the First Contact Census study, 2014-15 Oregon Accessibility Manual) were uploaded to the DIR and available to reviewers throughout the study. Reviewers were able to stop/restart their review at any point during the study, and change previously submitted item responses (i.e., based on research team feedback). Screenshots of the alignment review within the DIR are found in the training slides shown below (slides 36-48).

Beginning November 4th, we asked reviewers within the DIR to respond to the following statements, with completed responses due for all grade-level items on December 1, 2014.

1. Rate the strength of alignment between the test item and standard. (0 = insufficient alignment, 1 = sufficient alignment, 2 = strong alignment)
2. Item is Free of Bias. (yes/no)
3. Item is Accessible to SPED Students. (yes/no)

Additionally, for any "0" or "No" rating, the reviewer was asked to provide a rationale and recommendation(s) for improving the item. The research team used this feedback to edit and improve items during and after the alignment study.

Results and Item Editing. The research team edited or removed items with changes documented based on reviewers' ratings and the following criteria:

1. Deemed *insufficiently* aligned (average rating of <1.0)
2. Deemed *biased* (majority rating of "No")
3. Deemed *inaccessible* to SPED students (majority rating of "No").

The research team completed edits of items based on the criteria above, as well as the discretion of the research team. For example, although an item might have been rated as aligned (average rating of ≥ 1.0 across all raters), the research team used feedback from reviewers to edit/improve items in many cases. Similarly, because a majority of reviewers rated an item as *bias free* or *accessible* did not necessarily mean that they did not also give appropriate suggestions to improve a given item. Likewise, at least one reviewer in this example would have rated the item as having bias or as inaccessible and also provided feedback. The research team carefully considered reviewer ratings and feedback in all cases that it was given, and used the information as a basis to improve items in combination with the criteria enumerated above.

The research team made edits and improvements to item stems, answer options, and graphics, and typically included:

- Typos (e.g., misspellings, missing words, poor grammar),
- Bias and sensitivity (e.g., use of varied names to represent diverse populations, removal of references to religious or politically-charged topics),
- Accessibility (e.g., removal or limiting of construct irrelevant details, Universal Design for Assessment features),
- Alignment to Essentialized Standards, and
- Overall item improvement (e.g., clarifying graphics, diversifying examples of content relative to other items).

We calculated the following descriptive statistics from the alignment study using the entire item bank, including those items that were not used in 2014-2015 operational test forms (see table above for the number of items reviewed in each grade and content area). **It should be noted that the following statistics include reviewer ratings collected *prior to* or *concurrent with* item editing by the research team. Consequently, alignment, bias and accessibility ratings of the current item bank are likely quite higher over what is presented in this technical report. A formal alignment, bias, and accessibility study of the full (and edited) item bank is planned for 2015-2016.**

In the following table, average alignment, bias, and accessibility ratings are based on the total number of reviewer responses (see parenthetical in column headings), and are displayed by grade for the three content areas. Average *alignment* ratings are on a scale of 0-2, where a value of 2 indicates a perfect average alignment rating across all reviewers. We calculated average *bias free* and *accessibility* ratings, by converting the yes/no responses to 1/0, respectively, and then computing the average on a scale of 0-1, where a value of 1 indicates a perfect bias (free) or accessibility rating across all reviewers. For *English Language Arts*, average grade-level alignment ratings ranged from 1.66 to 1.93 (*M* = 1.83), bias ratings ranged from 0.96 to 0.99 (*M* = .98), and accessibility ratings ranged from 0.98 to 0.99 (*M* = .99). For *Math*, average grade-level alignment ratings ranged from 1.33 to 1.89 (*M* = 1.69), average grade-level bias ratings ranged from 0.67 to 0.99 (*M* = .94), and average grade-level accessibility ratings ranged from 0.65 to 0.99 (*M* = .88). For *Science*, average grade-level alignment ratings ranged from 1.80 to 1.88 (*M* = 1.83), average grade-level bias ratings were 0.99 across all grades, and average grade-level accessibility ratings ranged from 0.98 to 0.99 (*M* = .99).

Average Aligned, Bias Free, and Accessible Ratings by Content Area and Grade (# reviewers)

Grade	ELA (3)			Math (3*)			Science (4)		
	Align	Bias	Access	Align	Bias	Access	Align	Bias	Access
3	1.93	0.98	0.99	1.59	0.99	0.91	-	-	-
4	1.92	0.99	0.99	1.89	0.99	0.88	-	-	-
5	1.66	0.99	0.99	1.78	0.98	0.81	1.80	0.99	0.98
6	1.78	0.96	0.98	1.84	0.99	0.99	-	-	-
7	1.90	0.96	0.99	1.33	0.96	0.93	-	-	-
8	1.72	0.98	0.99	1.78	0.94	0.65	1.86	0.99	0.98
11	1.88	0.96	0.99	1.51	0.67	0.90	1.88	0.99	0.99
Total	1.83	0.98	0.99	1.69	0.94	0.88	1.83	0.99	0.98

Note. *4 reviewers rated items in Grade 6 Math. Align = average *alignment* rating (0-2 scale); Bias = average *bias free* rating (0-1 scale); Access = average *accessibility* rating (0-1 scale); Total = across grade average.

Two representative reviewer comments from each content area are shown below—selected to demonstrate typical responses that the research team used to guide item editing and refinement.

Rather than "I said", what about using an actual name? That would test whether [the student] could identify the speaker. – Reviewer 3, Grade 5, ELA

The restaurant picture is unclear…tables with diners would help. – Reviewer 1, Grade 11, ELA

The graphic shows a rope about to be cut in half, not cut in half. [Display] the rope in two pieces, rather than one piece, with scissors about ready to cut. Many students with special needs are VERY literal. – Reviewer 1, Grade 3, Math

The [Essentialized] Standard refers to expressions. To fit the [Essentialized] Standard the answer would be "4 nickels". Converting to 20 cents is an additional step not covered by this Standard. – Reviewer 2, Grade 7, Math

The correct response doesn't show the actual mixture; instead it shows the two components of the mixture [as] separate (not mixed). – Reviewer 4, Grade 5, Science

In science class the students will not see the measurement in Fahrenheit. Water boils at 100 degrees Celsius, so you may want to take that out as a distractor, since the water [in the stem graphic] looks like it is boiling. – Reviewer 1, Grade 11, Science

Alignment, Bias and Accessibility Training Slides

Oregon Extended Assessment

Item Reviews – Fall 2014

Behavioral Research and Teaching (BRT)
University of Oregon
November 5, 2014

BRT
behavioral research & teaching

1

Objectives

- To understand the development process of Oregon Extended Assessment (OExt) items.
- To become familiar with the steps associated with the reduction of depth, breadth, and complexity of state content standards as it relates to English Language Arts, Math, and Science.
- To ensure understanding of the standard essentialization process.
- To ensure the OExt test items align to Oregon's essentialized standards, while maintaining fidelity to the reduction process.
- To make judgments about SPED and Bias of the field test items.

BRT
behavioral research & teaching

2

Why Are We Doing this Work?

- Conduct a professional review by experts (you)
- Use a structured and credible process to evaluate the alignment to essentialized standards, SPED student accessibility, and Bias of the items on the OExt
- To make changes to the field test items as suggested by you

3

BRT
behavioral research & teaching

Ultimate Goal

Improving academic achievement for students with significant cognitive disabilities by linking

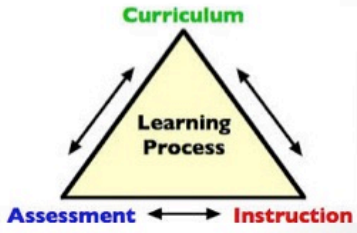
- Academic standards
- Instruction, and
- Assessment

4

BRT
behavioral research & teaching

Big Picture

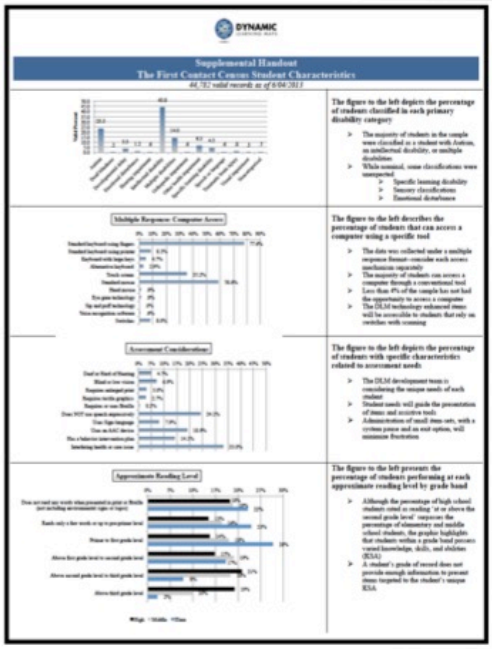
- All students in Oregon are required to demonstrate proficiency on grade-level content standards.
- Students with significant cognitive disabilities need to demonstrate progress toward reaching proficiency on grade-level content standards.
- The ORExt is designed to assess the progress of students with significant disabilities toward meeting these (essentialized) content standards.



BRT
behavioral research & teaching

5

Students with Significant Cognitive Disabilities (SWSCDs)



BRT
behavioral research & teaching

6

Video of Student Population of Oregon Extended Assessments

BRT
behavioral research & teaching

7

Guidelines for the ORExt

- Each test item focuses on a specific essentialized standard or standards.
- Items are written to cover a range of difficulty (low – medium – high).
- Items are written at different cognitive levels (think Bloom's Taxonomy)

  Remember, Understand, Apply
- Items should be free of age, gender, religious, ethnic or disability stereotypes, and bias.

BRT
behavioral research & teaching

8

General Guidelines for ORExt Multiple Choice Items:

- Are comparable in length and parallel in structure.
- Have only one correct answer.
- Have one near and one far distractor.
- Are clearly-worded and are appropriate for students in the assigned grade and population in terms of reading level, interests, and experience.
- Answer choices will be arranged with sufficient white space on the page to ensure that there is no opportunity for distraction or confusion of responses.

BRT behavioral research & teaching

9

General Guidelines for ORExt Multiple Choice Items, cont.

- Test items should not be worded in the negative ("Which of these is NOT . . . "), except in rare instances when it offers substantial advantages for the item construction or representation of the targeted construct.
- Do not be overly-concerned with scaling of the item graphics/text or additional text in answer options (e.g., ; A, ; B, ; C), as the graphics will be scaled/ edited during the test form development process, and font will be 18-pt or larger for all text.

BRT behavioral research & teaching

10

Items which must be edited or rejected:

- Do not match any essentialized standard
- Have more than one possible correct answer
- Are not accessible to the student population
- Are "tricky" or confusing (text and/or graphics)
- Are potentially biased in terms of age, gender, religious, ethnic, or cultural issues
- Are grammatically incorrect (unless purposefully so)
- Address English Language Arts, Math, and Science content or vocabulary incorrectly

11

Item Review: Content Review

Critical Content Related Questions:

Is this English Language Arts, Math, or Science?

Is this item an adequate representation of the essentialized standard?

Is the item appropriately reduced in terms of breadth, depth, and complexity?

Is this item accurate with the suggested response?

Is the intent of the item clear?

Any suggested rewording (for clarity of wording, accuracy, or intent)?

12

Item Review: Bias/Sensitivity

- Gender
- Race/Ethnicity/Culture
- Age
- Disability
- Religion
- Socioeconomic Status
- Region (e.g., local vocabulary)
- General

13

BRT behavioral research & teaching

Item Review: Special Education

- Will students with significant cognitive disabilities understand this language?
- Will students with significant cognitive disabilities understand the intent?
- Do you have suggested rewording/graphics edits (for clarity and universal access)?

14

BRT behavioral research & teaching

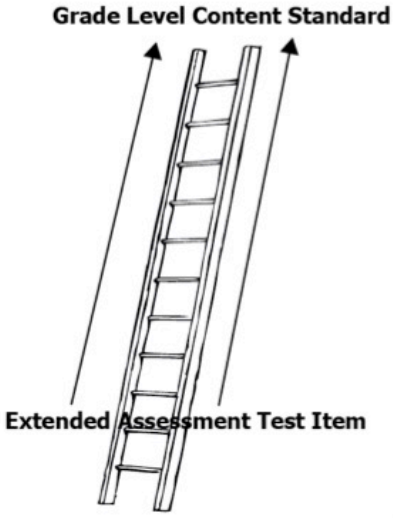
ORExt Items

- Are linked to OR Content Standards by the Essentialized Standards
- Essentialized Standards are systematically reduced in:
  - Depth
  - Breadth
  - Complexity

BRT
behavioral research & teaching

15

Linking Content Standards with ORExt Test Items



BRT
behavioral research & teaching

16

Essentializing Oregon Content Standards

BRT
behavioral research & teaching

17

Essentializing Coding System

- (a) Essential content (nouns) is boxed
- (b) Essential intellectual operations (verbs) are underlined (with complex verbs also **bolded**), and
- (c) Delimiters (of content or intellectual operations) are *italicized*.

BRT
behavioral research & teaching

18

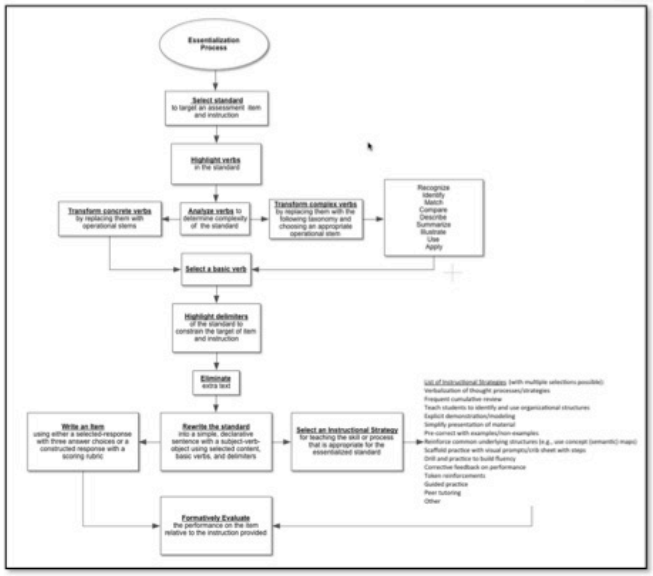
Essentialization Process

- 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

BRT
behavioral research & teaching

19

Essentialization Flow Chart



BRT
behavioral research & teaching

20

Example 1: How to Essentialize a Standard (Reading)

- 4.RF4 - Read with sufficient accuracy and fluency to support comprehension.
- **Read** text *with sufficient accuracy and fluency to support comprehension*.
- Essentialized standard: **Read** *appropriate* text *with accuracy*.

21

BRT
behavioral research & teaching

Example 2: How to Essentialize a Standard (Writing)

- 11-12W.4 Produce clear and coherent writing in which the development, organization, and style are appropriate to task, purpose, and audience.
- **Produce** *clear and coherent* writing *in which the* development, organization, *and* style *are appropriate to task, purpose, and audience*.
- Essentialized standard: **Write** *relevant* text *with accuracy*.

22

BRT
behavioral research & teaching

Example 3: How to Essentialize a Standard (Language)

- 11-12L1 - Demonstrate command of the conventions of standard English grammar and usage when speaking or writing.
- **Demonstrate** *command of the* conventions *of standard English grammar and usage when speaking or writing.*
- Essentialized standard: *Accurately* identify icons *when using expressive language.*

23

BRT behavioral research & teaching

Example 4: How to Essentialize a Standard (Math)

- 4.NBT4 - Fluently add and subtract multi-digit whole numbers using the standard algorithm.
- *Fluently* add *and* subtract *multi-digit* whole numbers *using the standard algorithm.*
- Essentialized standard: Add *two-digit* whole numbers *with fluency.*

24

BRT behavioral research & teaching

Example 5: How to Essentialize a Standard (Science)

- 5-PS1-3 - Conduct an investigation to determine whether the mixing of two or more substances results in new substances.
- **Conduct** *an* investigation *to determine whether the mixing of two or more* substances *results in new substances.*
- Essentialized standard: Recognize *when* substances *are mixed together.*

BRT behavioral research & teaching

25

The OrExt Item Review Process

BRT behavioral research & teaching

26

Item Review: The Process

1. Each item contains the Scoring Protocol (question that an Assessor reads to the student), Student Materials (what is placed before the student), the associated Essentialized Standard/s, and three questions for you to answer.
2. Rate the Alignment of the Item to the Essentialized Standard/s – **0**, **1**, *or* **2** (see the following slides for example ratings).
3. Answer the question about whether the item is Accessible for a student with significant cognitive disabilities – **Yes** *or* **No**.
4. Answer the question about whether the item is free of Bias (i.e., gender, race, ethnicity, culture, age, disability, religion, region) – **Yes** *or* **No**.
5. Write in the comment box any comments or suggestions you have for each item. Your comments are critical improving item alignment to the standard/s when you rate an item alignment as a "0". Also, if you have any suggestions for making the item more accessible to students with significant cognitive disabilities or less biased when you answer "no" to either question, please include those in the comments box.

27

BRT
behavioral research & teaching

Three Independent Judgments

1. **Item Alignment**
   Identify the level of alignment b/t items and Essentialized Standards – **0**, **1**, *or* **2**
2. **Item is Accessible to SPED Students**
   Yes *or* **No**
3. **Item is Free of Bias**
   Yes *or* **No**

28

BRT
behavioral research & teaching

Item Alignment Scale

2 = Item is ***strongly aligned*** to the Essentialized Standard

1 = Item is ***sufficiently aligned*** to the Essentialized Standard

0 = Item is ***not sufficiently aligned*** to any part of the Essentialized Standard

**Make note of difficulty level based on item code, L – M – H, when gauging alignment.

29

BRT behavioral research & teaching

Example of Items and Alignment Ratings - ELA

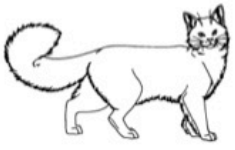
Example of a 0 = insufficient alignment:

ELA - Grade 11 Essentialized Standard RL1 - H item

Answer questions about a text that is read to student.	L - Paragraph of 4 sentences read to student M - Paragraph of 5 sentences read to student H - 2 paragraphs read to student

Here is a picture. (Point to picture.) Which word is a description of the picture: dog, cat, or bird?
[0 = incorrect / 1 = cat]

Which word is a description of the picture?



dog **cat** **bird**

30

BRT behavioral research & teaching

Example of Items and Alignment Ratings - Math

Example of a 1 = sufficient alignment:

Math - Grade 7 Essentialized Standard TNS - M item

Match rational numbers to their corresponding decimal (.5, .25, .75 with multiples of 1/2, 1/4, 3/4).	L - match .5 with 1/2, 2/4, 3/6, and 4/8; M match .25 with 1/4, 2/8, 3/12, and 4/16; H-match .75 with 3/4, 6/8, 9/12, and 12/16.

Here are three answer choices. (Point to each.) Which fraction is the same as .25: 6/32, 7/32, or 8/32?
[0 = incorrect / 1 = 8/32]

Which fraction is the same as .25?

$\frac{6}{32}$ $\frac{7}{32}$ $\frac{8}{32}$

BRT behavioral research & teaching

31

Example of Items and Alignment Ratings - Science

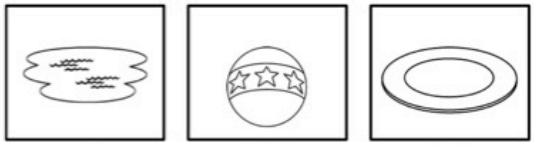
Example of a 2 = strong alignment:

Science - Grade 5 Essentialized Standard S05ESS2.2 - L item

Identify features made of water, and compare the amount of water in different reservoirs on Earth.	L - Restricted to questions about what Earth features that are made of water (i.e., oceans, lakes, rivers, streams) as compared to common objects that aren't (i.e., rock, brick, toy, ball

Here are three pictures. (Point to each.) Which one is a lake: A, B, or C?
[0 = incorrect / 1 = A]

Which one is a lake?



A B C

BRT behavioral research & teaching

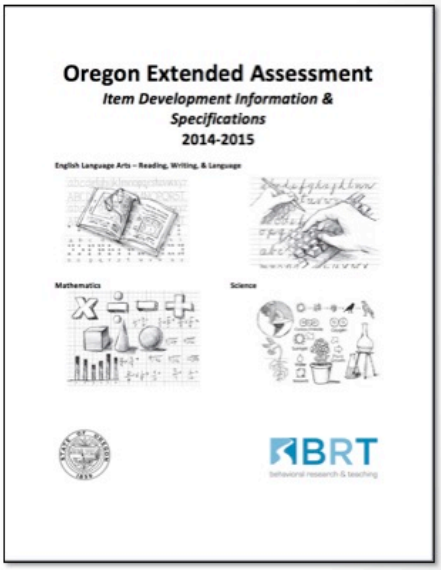
32

Resources to Support Your Review

- The following documents will support accurate and consistent decisions during your review
  - Dynamic Learning Maps SWSCD Survey Results
  - Item Development Information & Specifications (please read this in its entirety, so you can see exactly what we asked item writers to do)
  - Accessibility Manual (for your reference, p. 35-41)
  - Training Slides
- Videos of this webinar and student population

33

Item Development Information & Specifications



34

ORExt Item Development Information & Specifications

- Background (p. 2)
- RDBC (p. 2)
- EAFs (pp. 3-4)
- ORExt Test Design (pp. 4-6)
- Test Development Considerations (pp. 6-9)
- Item Specifications (pp. 10-11)
- Anticipated Accommodations (pp. 12-14)

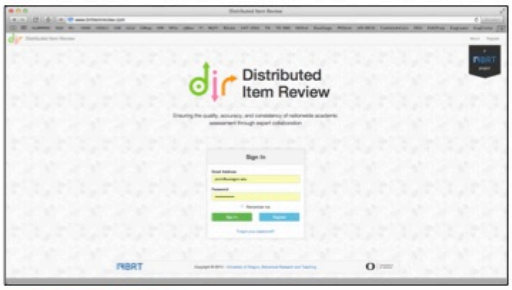
BRT behavioral research & teaching

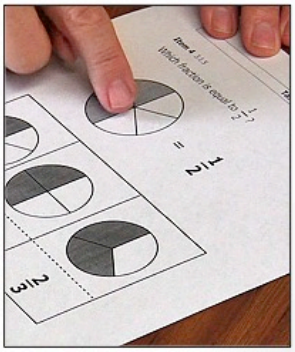
35

Distributed Item Review (DIR)

http://www.brtitemreview.com

A web-based system for presenting ***test items*** to ***experts*** across a ***broad geographic region*** so they can ***review*** them for important dimensions of ***bias, sensitivity, and alignment with standards***.





BRT behavioral research & teaching

36

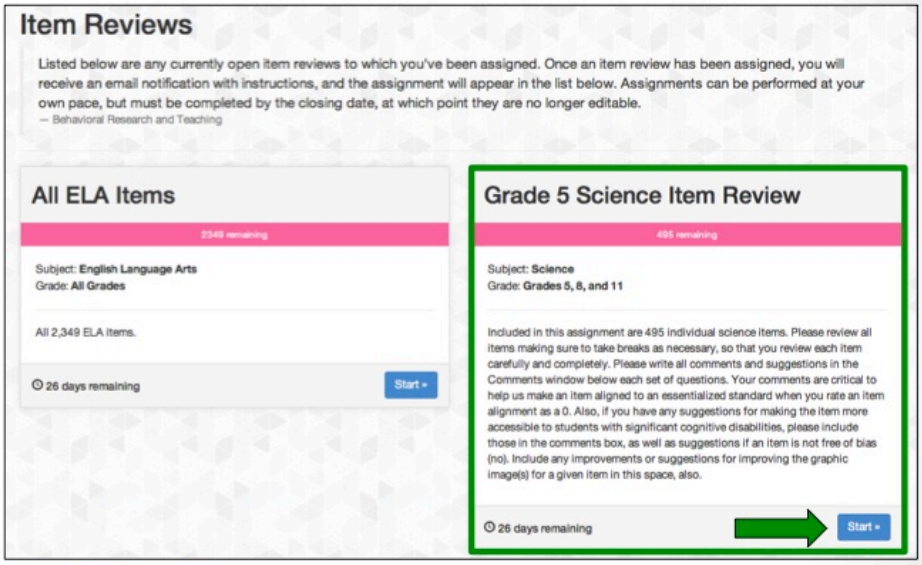
Accessing Your Grade-level Assignment in the DIR

1. Carefully look over the review (i.e., subject, grade, details).
2. Get going by clicking on "Start" button.

BRT behavioral research & teaching

37

Grade 5 Science Item Review Example



BRT behavioral research & teaching

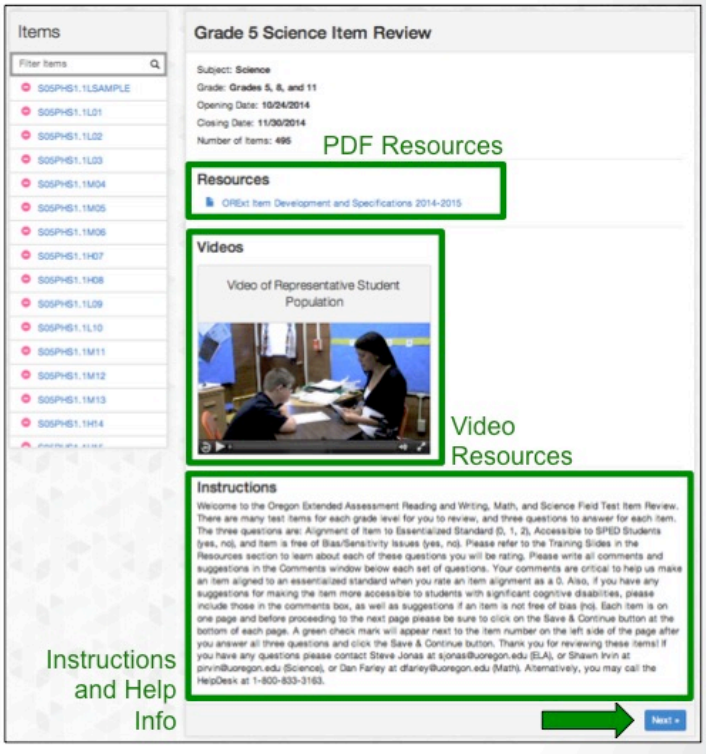
38

Accessing Your Grade-level Assignment in the DIR, cont.

3. Carefully review specific details (i.e., dates, # items, PDF resources, video resources, instructions).
4. Begin reviewing items by clicking on "Next" button.

BRT behavioral research & teaching

39



BRT behavioral research & teaching

40

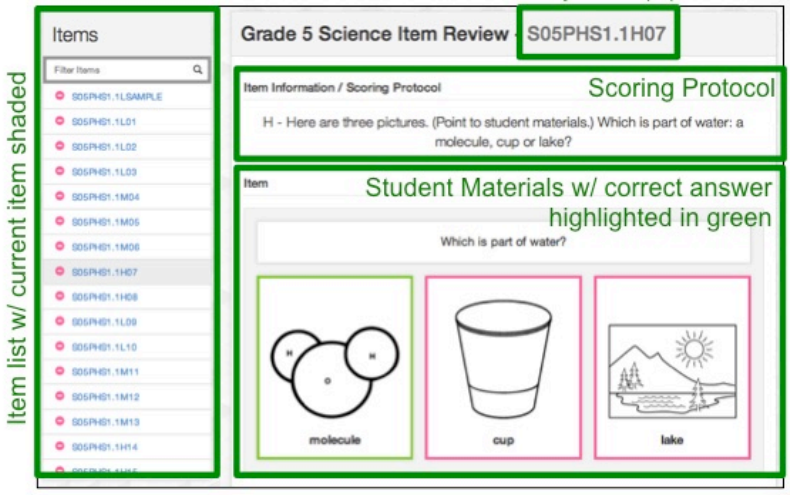
Reviewing Items in the DIR

5. Carefully review the item code, scoring protocol, student materials, and three answer options – one correct, near distractor, and far distractor

BRT
behavioral research & teaching

41

Item code w/ difficulty level (H)



BRT
behavioral research & teaching

42

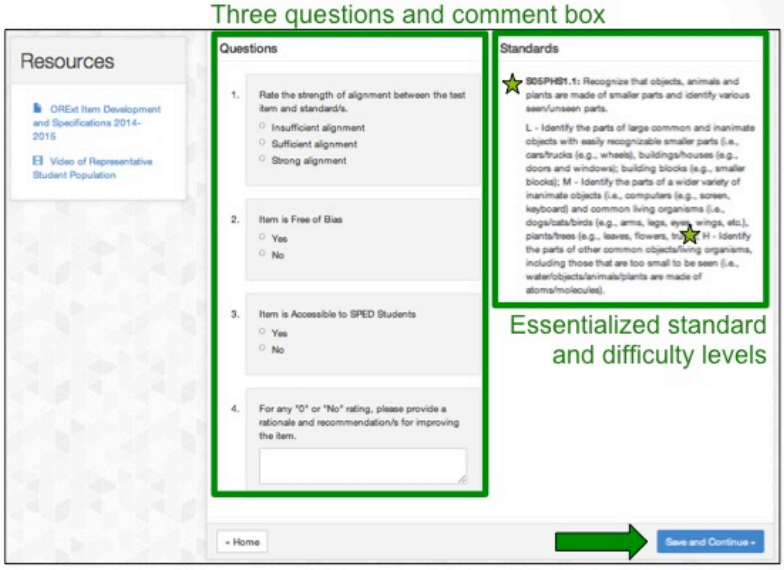
Reviewing Items in the DIR, cont.

6. Carefully answer all three questions below the test item (i.e., alignment rating, SPED, Bias).
7. Provide information in comment box for any rating of '0' or any response of 'No'.
8. Click "Save and Continue" to move to next item.

BRT behavioral research & teaching

43

Three questions and comment box



BRT behavioral research & teaching

44

Important Things to DO

- You may (and should) stop and start your review at anytime and any number of times – the DIR keeps track of your spot by giving you a green dot (•) next to the item ID code.
- You may go back and edit an item by scrolling/clicking on it in the Item List – click "Save and Continue" to keep changes.
- Please budget your time such that you can meet all required deadlines, as compensation is dependent upon completion of item assignments and meeting item review deadlines (Nov 17 and Dec 1)

BRT
behavioral research & teaching

45

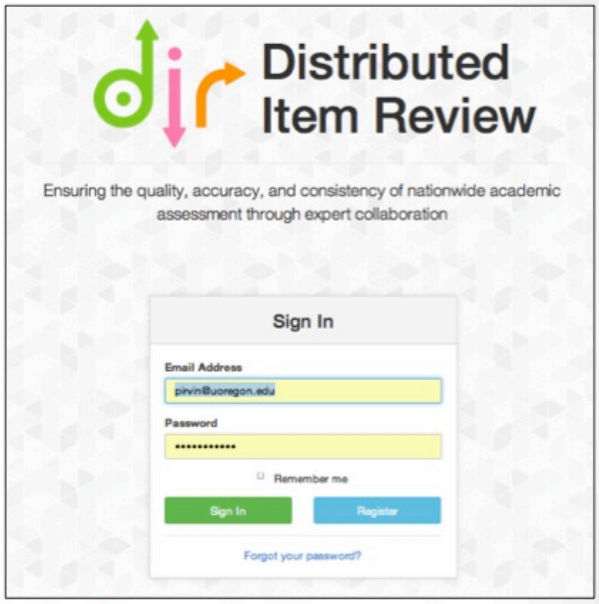
Important Things to AVOID

- This process is aimed exclusively at determining the alignment of items to essentialized standards, potential bias/sensitivity issues in the items, and access to these items for SWSCDs. You are not being asked to review the essentialized standards, nor make judgments regarding the appropriateness of statewide assessment, etc. Policy issues are not our concern here.
- Do not be overly-concerned with scaling of the item graphics/ text or additional text in answer options (e.g., ; A, ; B, ; C), as these will be adjusted during the test form development process, with font at 18-pt or larger.
- Do not be overly-concerned with math coding systems, as we will ensure that the items will be appropriate when presented to students (e.g., using 8x^2 instead of the appropriate exponent of $8x^{2)}$

BRT
behavioral research & teaching

46

DIR Walk-Thru



BRT behavioral research & teaching

http://www.brtitemreview.com

47

Next Steps

- Register and login to the Distributed Item Review (DIR) website: http://www.brtitemreview.com/
- Rate the item alignment to the essentialized standard with a 0, 1, *or* 2 for each of the items.
- Answer the Yes/No questions about SPED accessibility and Bias.
- Write comments or suggestions for improving an item, as needed, in the comments box.
- For any item rated as '0'/'No', provide suggestions for making the item a 1 or 2/more accessible/free of bias.
- Complete half of your judgments and comments by **November 17, 2014 – we will be monitoring progress.**
- Complete all of your judgments and comments by **December 1, 2014 – we will be monitoring progress.**

BRT behavioral research & teaching

48

Next Steps, cont.

- Make sure that you send Dan Farley a copy of the following two documents
  - Scope of Work
  - W-9
  - dfarley@uoregon.edu OR FAX at 541-346-5689
- Invoices will be distributed by e-mail in late November, to give you a chance to complete the work and submit the invoice for payment near the final due date of December 1, 2014.

49

BRT behavioral research & teaching

Follow-up

Questions/Comments Contact:

- Item Review Project overall or Math
  - Dan Farley at dfarley@uoregon.edu, or 541.346.3133
- Science
  - Shawn Irvin at pirvin@uoregon.edu
- English Language Arts:
  - Steve Jonas at sjonas@uoregon.edu
- If you have questions for ODE, contact:
  - Bradley J. Lenhardt at Brad.Lenhardt@state.or.us or 503.947.5755

50

BRT behavioral research & teaching

Selection, Links, Alignment: Establishing a Validity Position for Oregon's Extended Assessments

APPENDIX B: Participant Data

ORExt Participant Background

Subject	Grade Level	Background	Years of Experience	Highest Level of Education	Assigned
ELA	K to 5	Content Specialist	27	MA	CS-ELA-Gr 3
ELA	4	Content Specialist	8	MA	CS-ELA-Gr 4
ELA	K to 5	Content Specialist	24	MS	CS-ELA-Gr 5
ELA	6 to 8	Content Specialist	21	MA	CS-ELA-Gr 7
ELA	9 to 12	Content Specialist	6	MAT	CS-ELA-HS
Math	K to 5	Content Specialist	30	MS	CS-Math-Gr 3
Math	K to 8	Content Specialist	50	MS	CS-Math-Gr 4
Math	6	Content Specialist	33	MBA	CS-Math-Gr 6
Math	K to 12	Content Specialist	11	MA	CS-Math-Gr 7 & Gr 8
Math	6 to 12	Content Specialist	12	MAT, MS	CS-Math-HS
Math	6 to 12	Content Specialist	-	-	CS-Math-HS
Math/Science	4 to 5	Content Specialist	19	MS	CS-Sci-Gr 5
Science	8	Content Specialist	25	MS	CS-Sci-Gr 8
All Sciences	9 to 12	Content Specialist	5	MAT, PhD	CS-Sci-HS
Any or all	6 to 8	SPED	9	MS	SPED-ELA & Math-Gr 8
ELA	K to 8	SPED	10	MS	SPED-ELA-Gr 3
Any or all	K to 8	SPED	12	MS	SPED-ELA-Gr 3
Any or all	1 to 5	SPED	30	BA	SPED-ELA-Gr 4

ORExt Participant Background (Continued)

Subject	Grade Level	Background	Years of Experience	Highest Level of Education	Assigned
ELA	K to 12	SPED	25	MA	SPED-ELA-Gr 4
Any or all	5	SPED	11	BA	SPED-ELA-Gr 5
Any or all	K to 12	SPED	18	MS	SPED-ELA-Gr 6
Any or all	6 to 8	SPED	35	MS	SPED-ELA-Gr 6
ELA/Science	4 to 8	SPED	3	MA	SPED-ELA-Gr 7
ELA/Math	K to 12	SPED	-	MS	SPED-ELA-Gr 7
ELA/Math	6 to 8	SPED	20	MA	SPED-ELA-Gr 8
ELA/Math	K to 8	SPED	5	MS	SPED-Math-Gr 3
ELA/Math	K to 5	SPED	30	MA	SPED-Math-Gr 4
Math	K to 5	SPED	13	MS	SPED-Math-Gr 5
Any or all	3 to 6	SPED	17	BS	SPED-Math-Gr 6
Any or all	6 to 8	SPED	11	MA	SPED-Math-Gr 6
Any or all	K to 7	SPED	10	MS	SPED-Math-Gr 7
Math/Science	K to 8	SPED	18	MS	SPED-Math-Gr 7
Math/Science	K to 8	SPED	17	MS	SPED-Math-Gr 8
ELA/Math	9 to 12	SPED	3	MA	SPED-Math-HS
Math/Science	K to 5	SPED	5	MS	SPED-Science-Gr 5

OORExt Participant Background (Continued)

Subject	Grade Level	Background	Years of Experience	Highest Level of Education	Assigned
ELA/Math/Science	Any	SPED	16	MS	SPED-Science-Gr 8
Math/Science	6 to 8	SPED	27	MS	SPED-Science-Gr 8
ELA/Science	6 to 12	SPED	16	Doctorate J.D., MA	SPED-Science-HS
Any or all	K to 5	SPED	10	MA	SPED-Sci-Gr 5

Selection, Links, Alignment: Establishing a Validity Position for Oregon's Extended Assessments

APPENDIX C: Training Materials

Oregon Extended Assessment

Linking and Alignment Study – Winter 2017

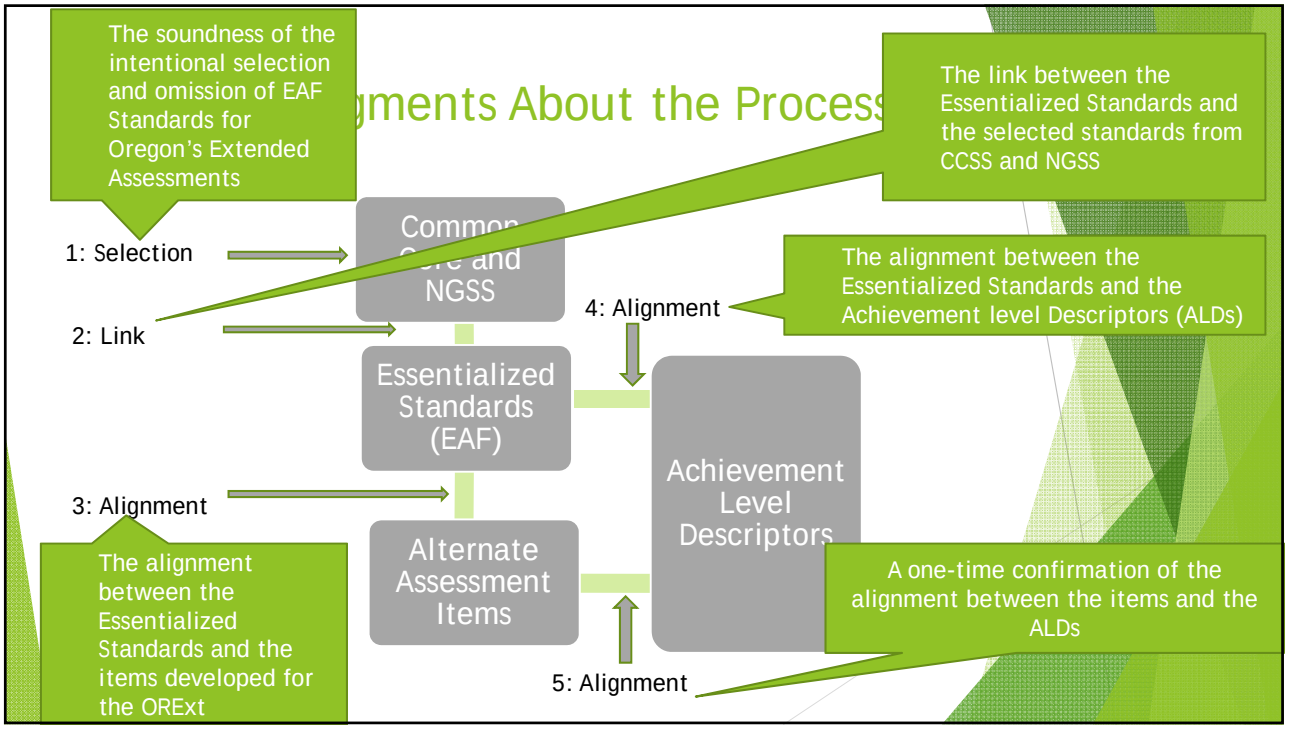
Wednesday January 11, 2017
3:00 – 4:30 PM
Dianna Carrizales-Engelmann
DCE Educational Communication LLC
for the Oregon Department of Education

Objectives

- To affirm the judgments made by developers of Oregon's Extended Assessments -- specifically regarding the relationship and connections between and among:
  - Assessment Items
  - Source Standards (CCSS, NGSS)
  - Essentialized Standards (ES)
  - Achievement Level Descriptors (ALDs)
  - Instructional decisions
  - Population (SWSCD)

Who is here

- Experts:
  - Educators who are experienced with the population of Students with Significant Cognitive Disabilities (SWSCDs)
  - Educators familiar with the Essentialized Standards and Oregon's CCSS and NGSS
  - Staff at ODE and test developers from BRT invested in the distribution of a valid assessment for Oregon's students



Your Judgments About the Process

- Assist in documenting the validity of the decision-making process for the development of Oregon's Extended Assessments, around:
- The soundness of the intentional selection and omission of EAF Standards for Oregon's Extended Assessments
- The link between the Essentialized Standards and the selected standards from CCSS and NGSS
- The alignment between the Essentialized Standards and the items developed for the ORExt
- The alignment between the Essentialized Standards and the Achievement level Descriptors
- The alignment between the items developed for the ORExt and the Achievement level Descriptors

Ultimate Goal

Improving achievement for students with significant cognitive disabilities by linking

- Academic standards
- Instruction
- Assessment
- Instructional/Achievement inferences

Big Picture

- All students in Oregon are required to demonstrate proficiency on grade-level content standards
- Students with significant cognitive disabilities need to demonstrate progress toward reaching proficiency on grade-level content standards
- Oregon's Extended Assessments were developed to assess the progress of students with significant disabilities toward meeting these content standards

Linking Study Spreadsheets

A	B	C	D	E	F	G	H	I	J	K	L	M
Standard	Common Core Standard	Essentialized Standard	Low Medium High Parameters	Level 2 ALD Statement	Level 3 ALD Statement	Level 4 ALD Statement	**EVALUATION 1:** Agree with determination that this standard should/should not have been included (YES/NO)	**EVALUATION 2:** Linkage Rating (0 = no link; 1 = sufficient link; 2 = strong link)	**EVALUATION 4**: Level 2 ALD Match? (YES/NO)	**EVALUATION 4**: Level 3 ALD Match? (YES/NO)	**EVALUATION 4**: Level 4 ALD Match? (YES/NO)	**EVALUATION 5:** The ALDs created for this subject and grade level align reliably and consistently with the corresponding ORExt assessment items. (YES/NO))

Note 1: *Math has an extra column between B & C, as there are several sub-standards (a-g). Math individuals will need to add a letter to the identifiers in this PPT, but they should be good at adding!*

Note 2: *Evaluation 3 will occur within the DIR system*

Content Standard Selection

Content standards were selected based upon three criteria:

1. This standard is a critical standard to learn in order to be able to access subsequent grade level standards
2. This standard is given more weight instructionally by teachers
3. This standard is accessible for SWSCDs, both in terms of performance match (i.e., cognitive complexity, depth of knowledge, breadth of knowledge) and in terms of sensory requirements (e.g., will be accessible to students with sensory impairments)

Read the Common Core Standard (Column B) and the Essentialized Standard (Columns C & D – including the L/M/H Parameters)

EVALUATION #1: Were the "right" standards included in the assessment?

In column H*, let us know whether we have included the appropriate grade level standards (and excluded the appropriate grade level standards) based upon these criteria

- If a standard was not included, it is highlighted in red

  (In ELA, the developers did not include the Speaking & Listening standards, nor the Literacy in History/SS, Science, and Technical Subjects standards due to implementation of the same criteria. Please include a statement at the bottom of any comments you may make in the Comment column stating whether you agree with this exclusion or not; if not, please explain why)

- If a standard is highlighted in green, it means that the developers believed the content to be covered by a different Essentialized Standard (and the standard that they believe it links to it is identified)

The fields in each of the spreadsheets you will be given have a column that is pre-populated with "Yes" – you only need to change those that you disagree with to "No" in this column, and then provide your rationale in the *Comments* column

Target Standard: Example

Grade 3 Reading Common Core Standard RL1

- Ask and answer questions to demonstrate understanding of a text, referring explicitly to the text as the basis for the answers.

Standard Selection Criteria

1. Instructionally critical: Is this standard a critical standard to learn in order to be able to access subsequent grade level standards?

   YES

2. Instructionally prioritized: Is this standard given more weight instructionally by life skills teachers?

   YES

3. Accessible: Is this standard accessible for SWSCDs, both in terms of match (i.e., cognitive complexity, depth of knowledge, breadth of knowledge) and in terms of sensory requirements (e.g., will be accessible to students with sensory impairments)?

   YES

Target Standard: Non-example

Grade 11 Math

- M.9_12.A.A_SSE.2.3.b: Complete the square in a quadratic expression to reveal the maximum or minimum value of the function it defines.

Standard Selection Criteria

1. Instructionally critical: Is this standard is a critical standard to learn in order to be able to access subsequent grade level standards

   NO

2. Instructionally prioritized: This standard is given more weight instructionally by life skills teachers

   NO

3. Accessible: This standard is accessible for SWSCDs, both in terms of performance match (i.e., cognitive complexity, depth of knowledge, breadth of knowledge) and in terms of sensory requirements (e.g., will be accessible to students with sensory impairments)

   NO

EVALUATION #2: Link between the Essentialized Standards and the Source Standard?

- Your second task is to determine the level of the link between the Essentialized Standard (ES) and the target standard(s) (CCSS for ELA and Math; OR Science and NGSS for Science)
- This is the scale you will use for these determinations (it is also found in row 1 on each spreadsheet)
  - **0 = No Link:** there is no connection between the content in the ES and the content in the target standard(s)
  - **1 = Sufficient Link:** there is a connection between the content in the ES and some aspect of the content in the target standard(s) that is easily recognizable, but not as strong as it could be
  - **2 = Strong Link:** the connection between the content in the ES and the content in at least one aspect of the target standard(s) is obvious and clear.

[*How strong is the connection?*]

Strong Link

2 = Strong Link: the connection between the content in the ES and at least one aspect of the content in the target standard(s) is obvious and clear.

Target Standard

Essentialized Standard

Strong link, with a few degrees of separation between the standards (Note: full alignment, i.e., reference to all aspects of the standard, is not the goal)

Strong Link Example

Grade 5 Science Target Standard

- 5-ESS2-2 Describe and graph the amounts and percentages of water and fresh water in various reservoirs to provide evidence about the distribution of water on Earth. [Assessment Boundary: Assessment is limited to oceans, lakes, rivers, glaciers, ground water, and polar ice caps, and does not include the atmosphere.]

Content: Amounts of water in various reservoirs, not including the atmosphere

Strong Link Example, cont.

Essentialized Standard, with L/M/H Parameters

- Compare the amount of water in different reservoirs on Earth.
- Discussion: The content in the essentialized standard links to the content in the source standard the variation is only introduced when we determine the boundaries within which the student may respond to the standard. In this example, in the low difficulty range the student is identifying water, while at the high difficulty range she is comparing the relative amounts of water in different reservoirs.
  - **Level 2 (Low)** - Restricted to questions about Earth features that are made of water (i.e., oceans, lakes, rivers, streams) as compared to common objects that aren't (i.e., rock, brick, toy, ball);
  - **Level 3 (Medium)** - Restricted to questions about Earth features that are made of water (i.e., oceans, lakes, rivers, streams) as compared to other natural features that aren't (mountains, volcanoes, forest, etc.);
  - **Level 4 (High)** - Restricted to comparing the relative amounts of water in various features of the hydrosphere (i.e., oceans, lakes, rivers, streams, ponds, etc.) using, for example, bar graphs that reflect the relative %s of water in the ocean vs. lakes vs. rivers; or Pacific Ocean vs. other oceans.

Sufficient Link

1 = Sufficient Link: there is a connection between the content in the Essentialized Standard and the content in at least one aspect of the target standard(s) that is easily recognizable, but not as strong as it could be

Target Standard

Essentialized Standard

Easily recognizable connection, but more degrees of separation between the standards

Sufficient Link Example

Grade 8 Math Target Standard

▶M.8.F.2.5. Describe qualitatively the functional relationship between two quantities by analyzing a graph (e.g., where the function is increasing or decreasing, linear or nonlinear). Sketch a graph that exhibits the qualitative features of a function that has been described verbally.

Content: Analyze a graph to determine change (increasing/decreasing, linear or nonlinear)

Sufficient Link, cont.

Essentialized Standard, with L/M/H Parameters

- Identify slope as positive, negative, zero, or undefined.
- *Discussion:* It can be argued that this is a 2, but it is at the very least a strong 1. The student is indeed comparing functions. They are only linear and they are only in four formats, but it gets at the standard's focus on comparing a relationship between two variables.
- Level 2 (Low) - identify positive slopes 1-3;
- Level 3 (Medium) - identify negative slopes 4-10;
- Level 4 (High) - identify zero or undefined slopes

No Link

0 = No Link: there is no connection between the content in the Essentialized Standard and the content in the target standard(s)

Essentialized Standard

Target Standard

These two standards are not the same

No Link Example

Grade 7 Writing

- 7.W3 - 3. Write narratives to develop real or imagined experiences or events using effective technique, relevant descriptive details, and well-structured event sequences. a. Engage and orient the reader by establishing a context and point of view and introducing a narrator and/or characters; organize an event sequence that unfolds naturally and logically.

- Content: Expressing a real or imagined story that engages and orients the reader

No Link Example, cont.

Essential Standard, with L/M/H Parameters

- Identify a number in writing.
- *Discussion:* Though this standard is indeed related to writing (so the situation could certainly be worse), it is not about expressing a story in any way. It has to do with identifying the written form of numbers.
- Level 2 (Low) – Identify numbers 1-10;
- Level 3 (Medium) – Identify numbers 11-20;
- Level 4 (High) – Identify numbers 21-40

EVALUATION #3: Alignment between the Essentialized Standard and the Item?

- Your third task is to determine the relationship between the Essentialized Standard (ES) and the secure assessment items as written.
- There is no column in the spreadsheet for this judgement. This review will occur within the Distributed Item Review (DIR platform).
  - **0** = Item shows *no alignment* to any part of the Essentialized Standard
  - **1** = Item shows *sufficient alignment* to the Essentialized Standard
  - **2** = Item shows *strong alignment* to the Essentialized Standard

**Make note of difficulty level based on item code, L – M – H, when gauging alignment.

Three Independent Judgments within this decision

1. **Item Alignment**
   Identify the level of alignment between items and Essentialized Standards – **0**, **1**, *or* **2**
2. **Item is Accessible to SPED Students**
   Yes *or* **No**
3. **Item is Free of Bias**
   Yes *or* **No**

- Does the performance demand conveyed in each item match that described in the associated performance descriptor.

Example of Items and Alignment Ratings - ELA

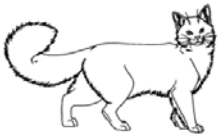
Example of a 0 = insufficient alignment:

ELA - Grade 11 Essentialized Standard RL1 - H item

Answer questions about a text that is read to student.	L - Paragraph of 4 sentences read to student M - Paragraph of 5 sentences read to student H - 2 paragraphs read to student

Here is a picture. (Point to picture.) Which word is a description of the picture: dog, cat, or bird?
[0 = incorrect / 1 = cat]

Which word is a description of the picture?



dog **cat** **bird**

Example of Items and Alignment Ratings - Math

Example of a 1 = sufficient alignment:

Math - Grade 7 Essentialized Standard TNS - M item

Match rational numbers to their corresponding decimal (.5, .25, .75 with multiples of 1/2, 1/4, 3/4).	L - match .5 with 1/2, 2/4, 3/6, and 4/8; M match .25 with 1/4, 2/8, 3/12, and 4/16; H- match .75 with 3/4, 6/8, 9/12, and 12/16.

Here are three answer choices. (Point to each.) Which fraction is the same as .25: 6/32, 7/32, or 8/32?
[0 = incorrect / 1 = 8/32]

Which fraction is the same as .25?

$\frac{6}{32}$ $\frac{7}{32}$ $\frac{8}{32}$

Example of Items and Alignment Ratings - Science

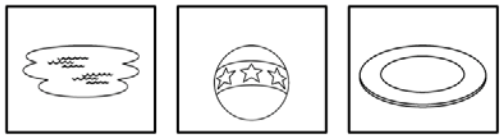
Example of a 2 = strong alignment:

Science - Grade 5 Essentialized Standard S05ESS2.2 - L item

Identify features made of water, and compare the amount of water in different reservoirs on Earth.	L - Restricted to questions about what Earth features that are made of water (i.e., oceans, lakes, rivers, streams) as compared to common objects that aren't (i.e., rock, brick, toy, ball

Here are three pictures. (Point to each.) Which one is a lake: A, B, or C?
[0 = incorrect / 1 = A]

Which one is a lake?



A B C

Resources to Support Your Review

- The following documents will support accurate and consistent decisions during your review
  - ORExt Item Development Information & Specifications
  - Accessibility Manual
  - Alignment and Linking Study Slides

Evaluation #4: Alignment between the Essentialized Standard and the Achievement Level Descriptors (Levels 2, 3, and 4)?

- Your fourth task is to determine the relationship between the Essentialized Standard (ES) and the Achievement Level Descriptors.
  - **Yes: ALD Level 2 (ALD Level 3, ALD Level 4)** Achievement Level Descriptor accurately describes the achievement of a student who has a partial or inconsistent (proficient, exceptionally proficient) understanding of the Essentialized Standard.
  - **No: ALD Level 2 (ALD Level 3, ALD Level 4)** Achievement Level Descriptor does not accurately describe the achievement of a student who has a partial or inconsistent (proficient, exceptionally proficient) understanding of the Essentialized Standard.

[*Given that the Achievement Level Descriptors (ALDs) were developed directly from each of the Essentialized Standards, the developers anticipate a strong alignment between the ALDs and the Essentialized Standards*]

Strong Alignment

2 = Strong Alignment: the connection between the content in the ES and the language of the content in the ALD is obvious and clear.

Achievement Level Descriptor

Essentialized Standard

Sufficient Alignment

1 = Sufficient Alignment: there is a connection between the content in the ES and the language in the ALD that is easily recognizable, but not as strong as it could be

Achievement Level Descriptor

Essentialized Standard

No Alignment

0 = No Alignment: there is no connection between the content in the ES and the language of the ALD

Essentialized Standard

Achievement Level Descriptor

EVALUATION 5: Alignment between the ORExt Item and the ALD (level 2, 3, or 4)

- At the conclusion of your spreadsheet there will be a final statement that is different from all of the previous statements.
- The statement asks you for a one-time affirmation of the alignment between the ORExt Items you have been reviewing, and the Achievement Level Descriptors you have been working with for this evaluation.
- Affirm that:
  - The ALDs created for this subject and grade level align reliably and consistently with the corresponding ORExt assessment items.

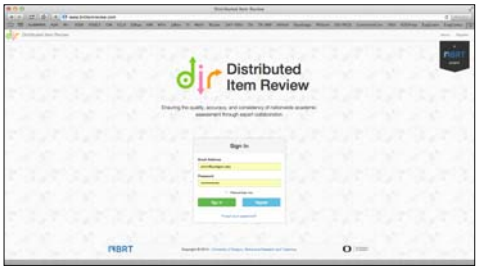
Process Review

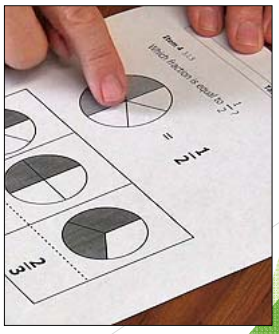
- First: Read the standards
- Conduct the 5 evaluations:
  - Evaluation 1: Determine standard selection agreement/disagreement (If you disagree with the standard selection or exclusion, change the "Yes" to a "No" in Column H, and then use the *Comments* column to explain why) [*Your Excel Spreadsheet*]
  - Evaluation 2: Rate the Linkage of the ES to the Target Standard/s with a 0, 1, or 2 in Column I (if you rate a linkage as "0" please explain why in the *Comments* column) [*Your Excel Spreadsheet*]
  - Evaluation 3: Rate the alignment between the Item and the Essentialized Standard [*DR Platform – see following slides for procedure*]
  - Evaluation 4: Rate the alignment of the ES to the ALDs (Level 2, 3, 4) [*Your Excel Spreadsheet*]
  - Evaluation 5: Rate the link between the Item and the ALDs [*Your Excel Spreadsheet*]
- E-mail your completed spreadsheets to carrizad@gmail.com by January 29th, 2017 using the "_DCE" filename extension (with your initials instead of mine)

Evaluation 3: Distributed Item Review (DIR)

http://www.brtitemreview.com

A web-based system for presenting ***test items*** to ***experts*** across a ***broad geographic region*** so they can ***review*** them for important dimensions of ***bias, sensitivity, and alignment with standards***.

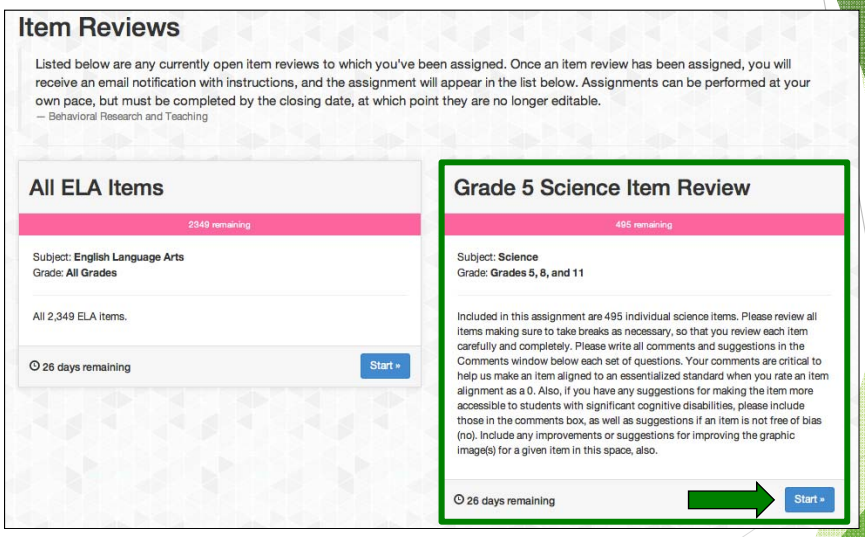




Accessing Your Grade-level Assignment in the DIR

1. Carefully look over the review (i.e., subject, grade, details).
2. Get going by clicking on "Start" button.

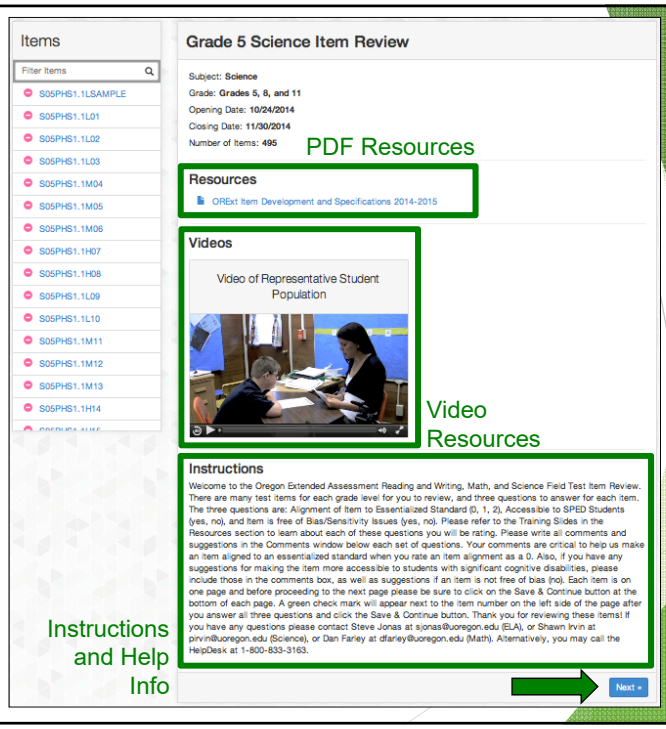
Grade 5 Science Item Review Example



37

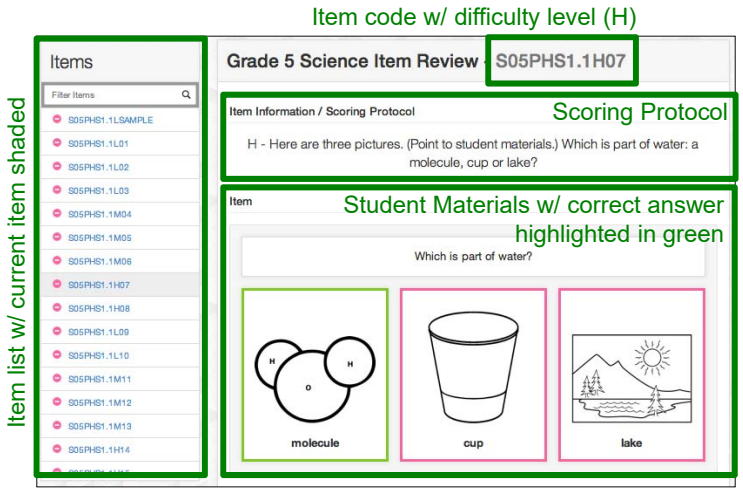
Accessing Your Grade-level Assignment in the DIR, cont.

3. Carefully review specific details (i.e., dates, # items, PDF resources, video resources, instructions).
4. Begin reviewing items by clicking on "Next" button.



Reviewing Items in the DIR

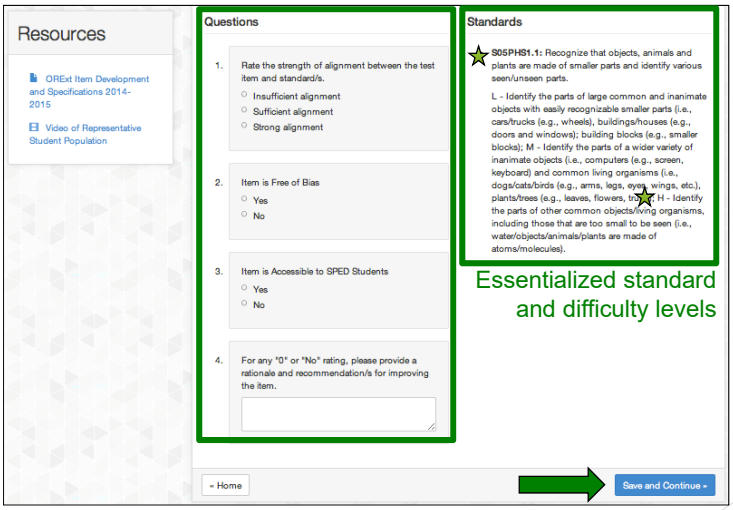
5. Carefully review the item code, scoring protocol, student materials, and three answer options – one correct, near distractor, and far distractor



Reviewing Items in the DIR, cont.

6. Carefully answer all three questions below the test item (i.e., alignment rating, SPED, Bias).
7. Provide information in comment box for any rating of '0' or any response of 'No'.
8. Click "Save and Continue" to move to next item.

Questions and comment box



Essentialized standard and difficulty levels

Important Things to DO

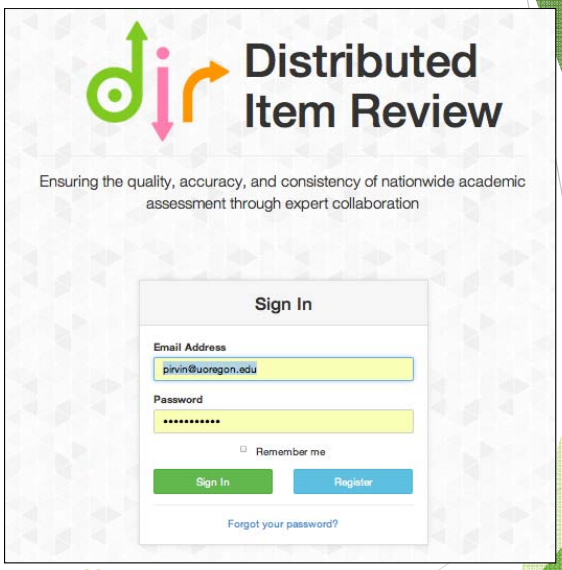
- You may (and should) stop and start your review at anytime and any number of times – the DIR keeps track of your spot by giving you a green dot (•) next to the item ID code.
- You may go back and edit an item by scrolling/clicking on it in the Item List – click “Save and Continue” to keep changes.
- Please budget your time such that you can meet all required deadlines, as compensation is dependent upon completion of item assignments and meeting item review deadlines (January 29th 2017)

Important Things to AVOID

- This process is aimed exclusively at determining the alignment of items to essentialized standards, potential bias/sensitivity issues in the items, and access to these items for SWSCDs. You are not being asked to review the essentialized standards, nor make judgments regarding the appropriateness of statewide assessment, etc.
- Do not be overly-concerned with scaling of the item graphics/text or additional text in answer options (e.g., ; A, ; B, ; C), as these will be adjusted during the test form development process, with font at 18-pt or larger.
- Do not be overly-concerned with math coding systems, as we will ensure that the items will be appropriate when presented to students (e.g., using 8x^2 instead of the appropriate exponent of $8x^{2)}$

DIR Platform Known Issues:

- ELA: The DIR review system does not recognize traced font.
  - All grades: Traced test items (where the letters appear as dashed lines) appear as regular font.
  - Grade 3 only: Some traced items show blank answer choices
- Math: The following symbols appear as written text rather than the math symbol:
  - Grade 6: Division, less than/equal to, greater than/equal to
  - Grade 8: Square root symbol.
- Science – No known issues



http://www.brtitemreview.com

Next Steps

- Register and login to the Distributed Item Review (DIR) website: http://www.brtitemreview.com/
- Rate the item alignment to the essentialized standard with a 0, 1, *or* 2 for each of the items.
- Answer the Yes/No questions about SPED accessibility and Bias.
- Write comments or suggestions for improving an item, as needed, in the comments box.
- For any item rated as '0'/'No', provide suggestions for making the item a 1 or 2/more accessible/free of bias.
- Complete all of your judgments and comments including those captured in the Excel Spreadsheet by January 29th 2017 at midnight **– we will be monitoring progress and may check in part way through the process.**

- **Email completed spreadsheets to me at carrizad@gmail.com**

Next Steps/Follow-up

- Questions/Comments Contact: Dianna Carrizales-Engelmann carrizad@gmail.com
- Thank you for your time!

Selection, Links, Alignment: Establishing a Validity Position for Oregon's Extended Assessments

APPENDIX D: Essentialized Standards User Guide

Essentialized Assessment Frameworks (EAFs)

2015-16 User Guide

Dan Farley
P. Shawn Irvin
Steve Jonas
Gerald Tindal
University of Oregon

Brad Lenhardt
Oregon Department of Education

Published by

Behavioral Research and Teaching
University of Oregon • 175 Education
5262 University of Oregon • Eugene, OR 97403-5262
Phone: 541-346-3535 • Fax: 541-346-5689
http://brt.uoregon.edu

&

Oregon Department of Education
255 Capitol Street NE • Eugene, OR 97310-0203
Phone: 503-947-5755 • Fax: 503-378-5316
http://ode.state.or.us

Copyright © 2015, Behavioral Research and Teaching & Oregon Department of Education. All rights reserved. This publication, or parts there of, may not be used or reproduced in any manner without written permission.

The University of Oregon and Oregon Department of Education are committed to the policy that all persons shall have equal access to its programs, facilities, and employment without regard to race, color, creed, religion, national origin, sex, age, marital status, disability, public assistance status, veteran status, or sexual orientation. This document is available in alternate formats upon request.

Background

Essentialized Assessment Frameworks (EAFs) were developed by Behavioral Research & Teaching (BRT) at the University of Oregon in consultation with the Oregon Department of Education (ODE). The development process included review and feedback from Oregon teachers, both general and special education, in three steps. First, the frameworks were linked to grade level content in the Common Core State Standards (English language arts & Math) and dually-linked to the Next Generation Science Standards (NGSS)/OR Science Standards. Second, they were designed to reflect grade level content that was reduced in terms of depth, breadth, and complexity (RDBC) in order to increase accessibility, as well as overall academic expectations, for students with significant cognitive disabilities (SWSCDs) in Oregon. Third, the EAFs formed the basis for developing new Oregon Extended Assessment items (ORExt) using a scaling technique that allows for modeling growth over grades.

Intended Uses

Educators in the field should use the EAFs as examples of the essentialization process. However, essentialization is an individualized process that should be conducted for each student based upon the student's present levels of functioning. Our intent is not to have teachers use the EAFs in a copy/paste fashion. Rather, the EAFs provide the field with additional information on the process for essentializing standards and multiple examples. The EAFs also are clearly related to the content of the ORExt and therefore provide a link between instruction and assessment. Ideally, educators can also use the essentialization process to develop Present Levels of Academic and Functional Performance (PLAAFPs), as well as Individualized Education Program (IEP) goals and objectives for SWSCDs. In fact, the essentialization process can generalize to all students to target instruction based on individual student needs.

Essentialization Process

The end result of the essentialization is a three-part statement that is based on targeted content, intellectual operations, and key delimiters to the content. Nouns are used to identify key content, verbs reflect the intellectual operation, and critical delimiters are conditional phrases or the object of the sentence. We have used the following conventions during the essentialization process: (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*. Additional reductions in depth, breadth, and complexity are made by limiting the scope of the content and/or changing the process (abstract) verb to be more accessible by using a product (concrete) verb.

Example of Essentialization with a Fraction Problem

4.NF.2.3.a (Grade 4, Number and Operations – Fractions, Build fractions from unit fractions by applying and extending previous understandings of operations on whole numbers, Standard 3a)

> "**Understand** addition *and* subtraction *of fractions as joining and separating parts referring to the same whole*."
>
> Essentialized Standard: "Add *two same-unit* fractions."

N.B. The original grade level standard has been reduced in terms of depth, breadth, and overall complexity. The essentialized standard remains reflective of grade level content, however. It is still focused on performing an operation with fractions, though the performance is limited to adding same unit fractions. This approach is critical, as the goal of essentialization is to maintain a strong link to grade level content while increasing accessibility for each student.

Caveat to EAF Structure: Each EAF document (ELA, Math, & Science) conveys the Essentialized Standards used to develop the new ORExt. However, not all CCSS and NGSS/ORSci standards were essentialized. Rather, standards were identified that were either (a) the most important to learn or (b) given the most opportunity to learn. Standards that were not essentialized have been highlighted in red. In the end, all EAFs have been vetted and approved by Oregon teachers in terms of their selection as well as their adaptation (content and structure). In some cases, this process resulted in very close relations among the grade level standards reflecting essentially the same core content across multiple standards (highlighted in green and a Essentialized Standard code to which they link).

Essentialized Standard Exemplars:. The spreadsheets demonstrate the determined linkages with grade level content of Essentialized Standards mapped out into three levels of difficulty: Low (L), Medium (M), and High (H).

The EAF documents are available at the following link, copyrighted © by Behavioral Research & Teaching (BRT) and Oregon Department of Education (ODE):

http://www.brtprojects.org/publications/training-modules

For questions or comments regarding the EAFs, please contact Dan Farley (dfarley@uoregon.edu)– BRT.