# Appendix 6.2.1

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

## Science, Math, & ELA Standard Settings

## June 15 -17, 2015

Oregon Department of Education

Behavioral Research and Teaching - University of Oregon

OFFICE OF LEARNING
STUDENT SERVICES
UNIT

BRT
behavioral research & teaching

# Welcome

*9:00 – 9:45 AM*

- Welcome!!
- Tell us about yourself
  - Name
  - District/school
  - Role

# Handouts

- Housekeeping
  - Agendas ( light pink)
  - Background Information ( light green)
  - PowerPoint Notes pages (3-slides per page)
  - Confidentiality Form
- Standard Setting Documents
  - Essentializing standards decision tree
  - Content Standards (Essentialized Standards)
  - Achievement Level Descriptors
  - Bookmarking Documents
    - Rating Sheets (light blue)
    - Standard Setter Evaluation form ( light yellow)
  - Ordered-item Booklets (OIBs)

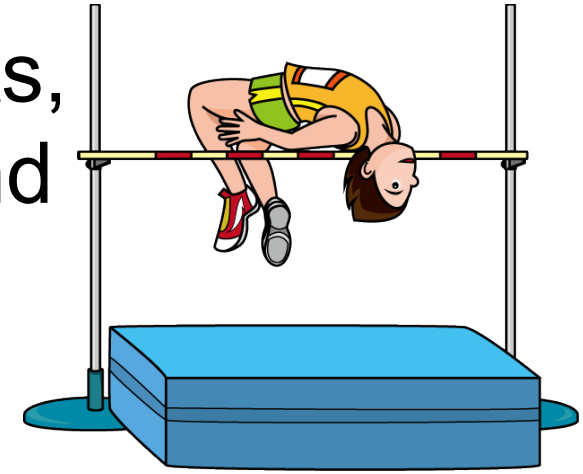
# Housekeeping

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

# Orientation to The Task of Setting Standards

# Purpose

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



# Outcomes

*Quantitative*

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

*Qualitative*

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

# Requirements for Your Role

- Minimum
  - Knowledge of the population
  - Knowledge of the assessment
  - Knowledge about accountability
- Ideal
  - Advocate for the population
  - Advocate for the assessment
  - Advocate for accountability

# What is Not Needed

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

# Table Discussion Rules

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

Adapted from *Brookfield & Preskill, 2005*

# Relevant Background Information

# How did we get here?

*9:45-10:30 AM*

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

# Students with Significant Cognitive Disabilities (SWSCDs)

- National Survey Results – Student Attention

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

# Students with Significant Cognitive Disabilities (SWSCDs)

- National Survey Results - Mathematics

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

# Students with Significant Cognitive Disabilities (SWSCDs)

- National Survey Results - ELA

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

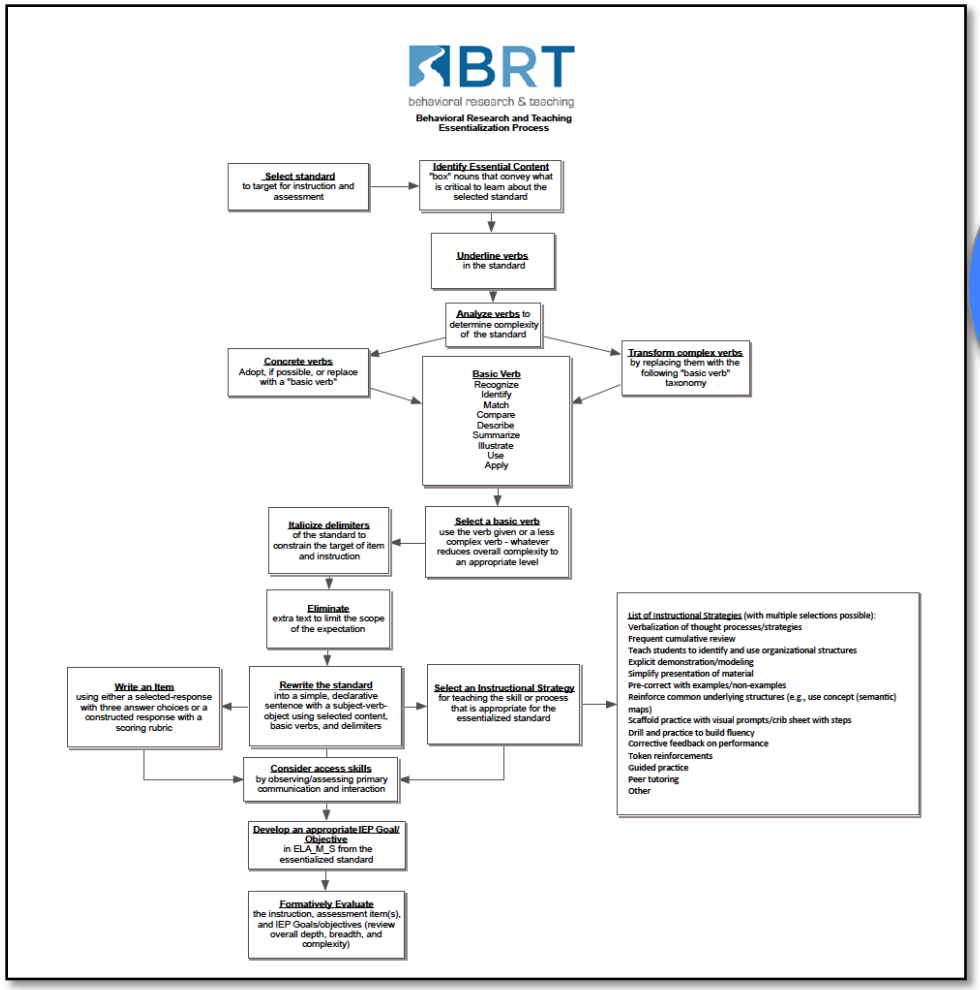
# Video of Student Population of Oregon Extended Assessments

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

# Essentializing Oregon's CCSS

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

# Essentialization Flowchart




Save your eyes & check your handout

# OExt Items

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

  (RDBC)

# Item Development

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

# What is the OR Extended Assessment?

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

# What are Alternate Achievement Standards?

Cut scores

Achievement level descriptors

# OR Statewide Assessment Options

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

# Administration and Format

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

# The ORExt Is

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

# What is the Purpose of the OR Extended Assessment?

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

# Four Performance Levels

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

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

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

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

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

# Standard Setting

# Educational Standard Setting

10:30-11:00 AM

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

# Standard Setting Outcomes

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

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

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

# How are alternate achievement standards set?

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

# Bookmarking Standard Setting

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

# ORExt Assessment Science Outcomes: Participation (2013-14)

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

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

# Extended Assessment Science Outcomes: Performance (2013-14)

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

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

# ORExt Assessment Math Outcomes: Participation (2013-14)

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

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

# Extended Assessment Math Outcomes: Performance (2013-14)

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

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

# ORExt Assessment Reading Outcomes: Participation (2013-14)

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

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

# Extended Assessment Reading Outcomes: Performance (2013-14)

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

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

# Setting Cut Scores

# General Process

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

# Process

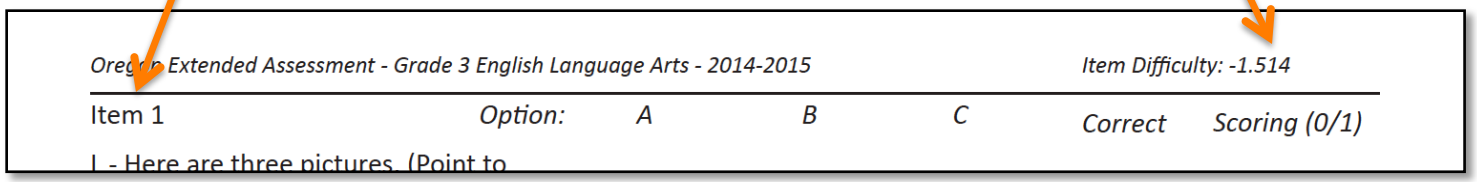
- All portions of the standard setting will be conducted in grade-level groups
- Each Group has a Table Facilitator
- Standard setters (Oregon Teachers)
  - Two special educators
  - One general educator
- BRT Table Facilitator
  - Manages time and materials
  - Keeps discussion focused
  - Takes notes

# Materials

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

# Structure of the OIBs

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



# Items in Booklet

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

# Bookmarking Expectations

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

# Round 1: Item level considerations

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

# Round 1: Process

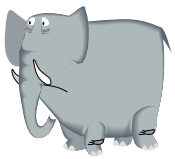
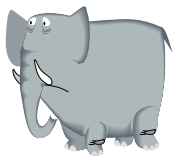
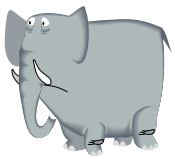
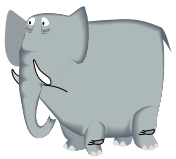
*11:00-12:00 PM*

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

- **Categories**
  - Level 4
  - Level 3
  - Level 2
  - Level 1

# Relationship Between Categories and Cut scores

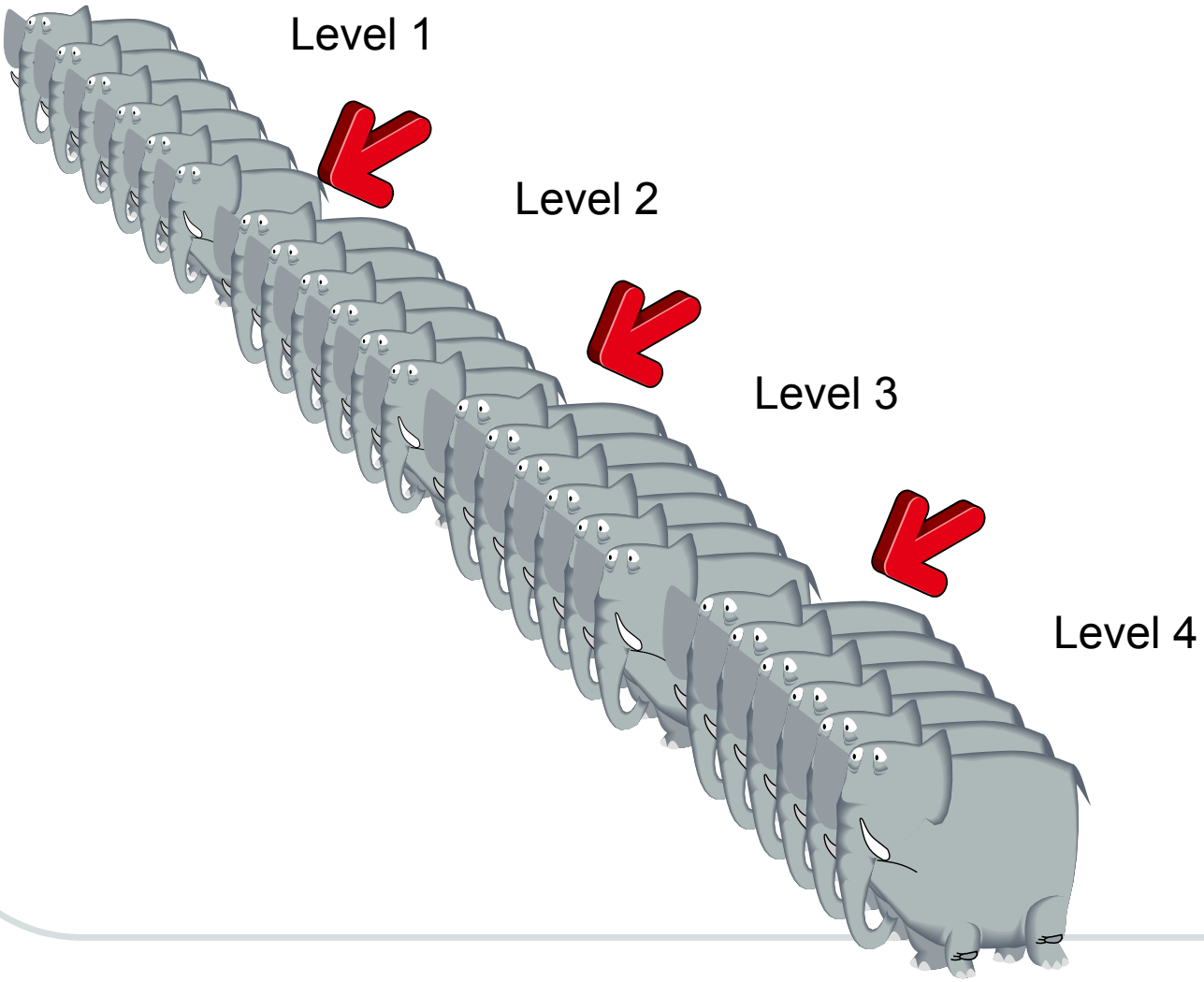
| Level 1 | Level 2 | Level 3 | Level 4 |
|---|---|---|---|
| Extremely limited to no performance of knowledge and skills | Inconsistent performance of knowledge and skills | Consistent performance of knowledge and skills | Superior performance of knowledge and skills |

# Meaning of a Cut Score

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

# Placing the Bookmarks



# Bookmarking Decision Rules

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

# Round 2: Group Consensus

*12:00-1:00 PM*

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

# The Scale of the test (hypothetical)

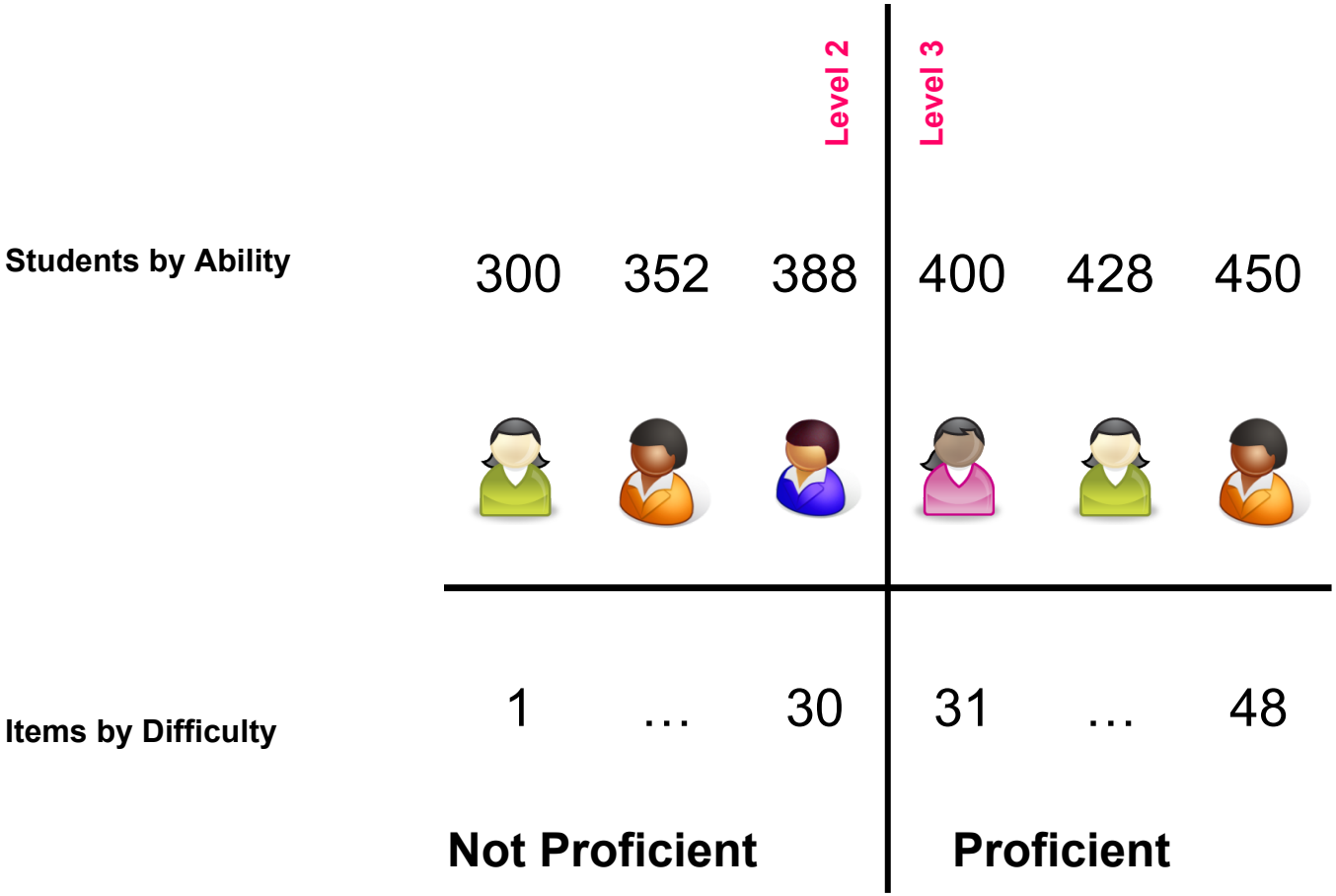
**Students by Ability**

300 304 309 . . . . . . . . . . . . 450

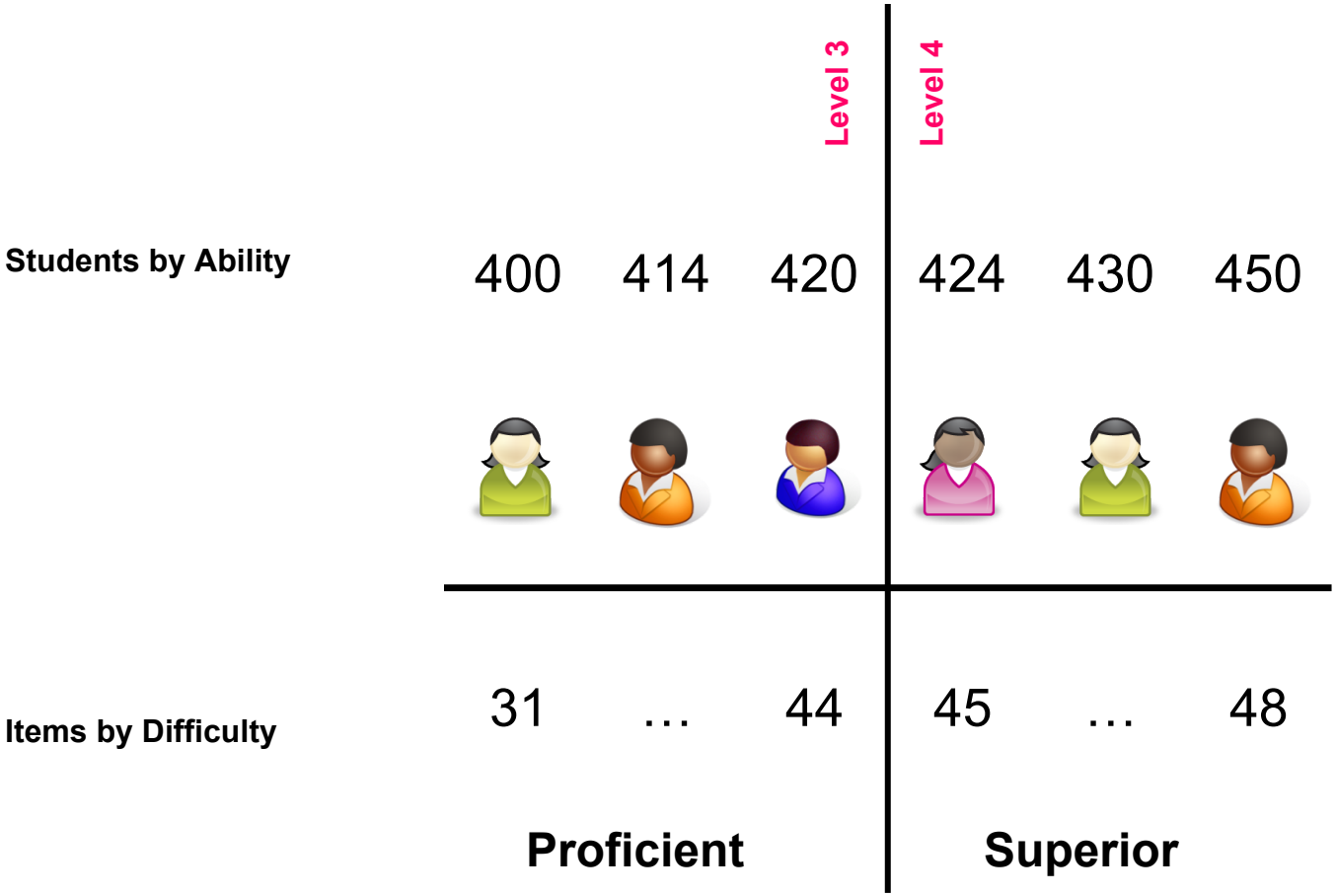
**Items by Difficulty**

1 2 3 … … … . . . 48

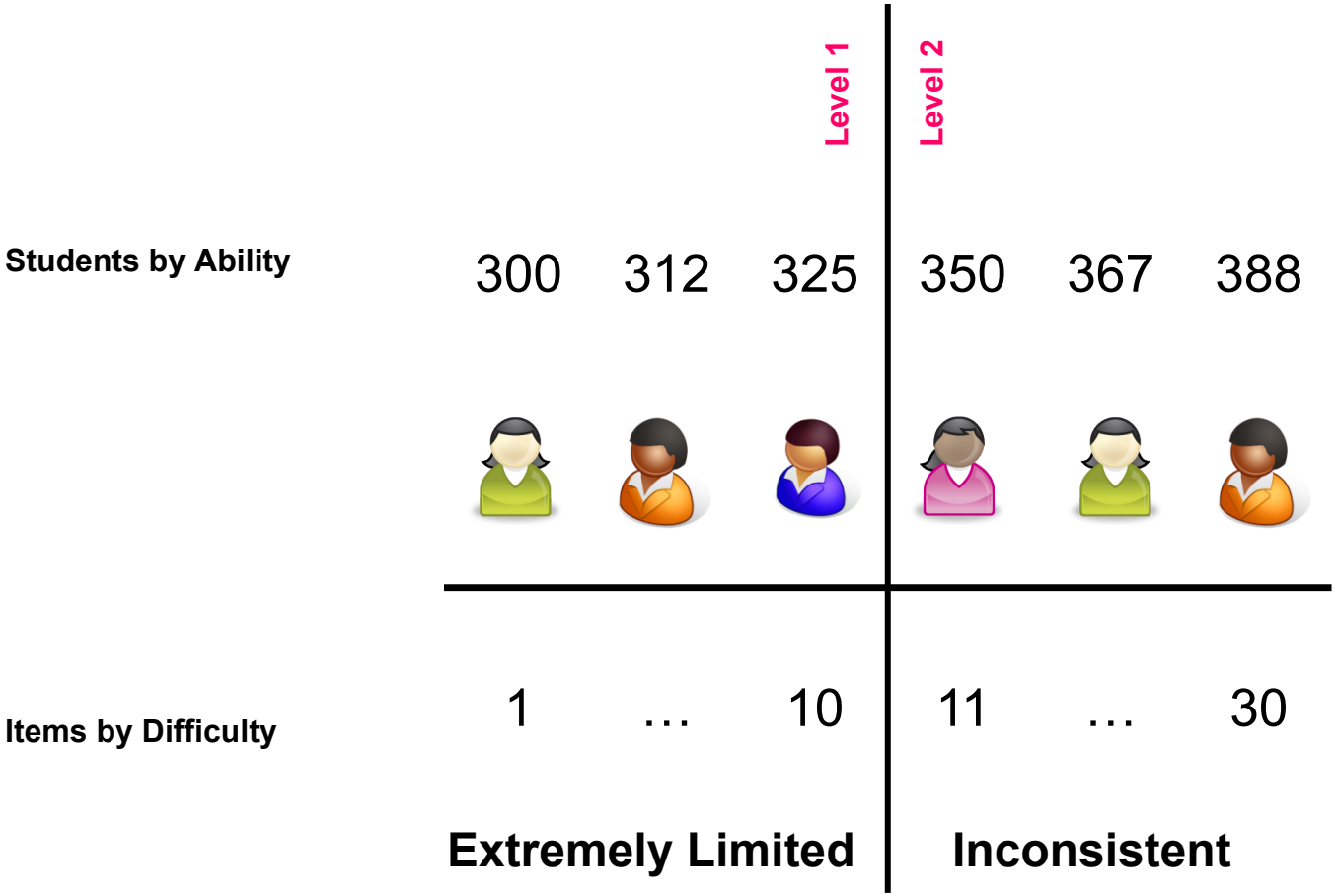
# Test Scale – First Decision



# Test Scale – Second Decision



# Test Scale – Third Decision



# Summary

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

# Break

*1:15 – 1:30 PM*

# Round 3: Data-based Decisions

*1:15 -2:30 PM*

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

# Impact Data Review

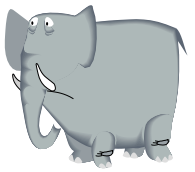
- Round 1 & 2 Complete
- Review percentages of students who would be placed in each performance level based on the selected cut scores
- Revise cut scores, if needed, for final determinations

# How do we know it's "right"?

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

# Sample Results (Hypothetical)

| | Nearly Meets | Meets | Exceeds |
|---|---|---|---|
| Hypothetical Cutscore (Median) | *10* | *35* | *45* |



Impact data

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

# Cut Score Rationale Statement

- After Round 3, please write the rationale used to generate your final cut scores; this can be recorded on the inside cover of the final page of the OIB
- If the group agrees about the rationale, only one statement needs to be recorded

# Establishing Achievement Level Descriptors

# Achievement Level Descriptors vs. Content Standards

*2:30 – 3:30 PM*

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

# Achievement Level Descriptor Overview

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

# ALD Categories

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

# Activity

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

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

# Summary

*3:30 – 4:00 PM*

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

# References


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

# Questions?

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

# Safe Travels & Happy Summer!