
Validating Alternate Achievement Standards

Marianne Perie
Center for Assessment



Validating Alternate Achievement Standards

- Should fit into larger context of a validity argument for the full alternate assessment based on alternate achievement standards
- Achievement standards help us interpret results and determine next steps
- Therefore, we must determine whether they were implemented well and are being used appropriately

Validation

- Cut scores do not have validity but the interpretation and uses of the cut scores do
- Just as there are no absolute criteria against which specific cut scores can be evaluated, there are no perfect criteria for evaluating standard setting studies (Kane, 1994, 2001)
- But... even though there are no absolute criteria you still must provide evidence that the cut scores are reasonable and appropriate



Validation (continued)

■ Key questions

- ❑ Was the standard setting procedure internally valid?
- ❑ Do the cut scores divide students reasonably in terms of achievement?
- ❑ Do the effects of the performance standards match what was intended?
- ❑ Were there any unintended consequences for using the performance levels?

Timing of Collecting Validity Evidence

- Some evidence can be collected early – during the standard setting process
- Some can be collected within the first year as the achievement levels are being used
- Other evidence must be collected over the long term

Validating the Procedures

- Was a documented method used?
 - If method was adopted from another, were the modifications justified?
 - Were all the typical steps applied?
- Was the panel composition appropriate?
- How much variation existed among the panelists?
 - Was there evidence of convergence across rounds?
- What does the panelist feedback tell us about the clarity of the procedures, adequacy of training, completeness of the materials, and their confidence in the results?

Validating the Results

- Comparing results to external sources of information is a way of checking if the cut scores are at approximately the right level.
 - Grades
 - Teacher evaluations
- Comparing results to results using another methodology may provide useful information
 - Contrasting groups in same or later year may provide information
 - Different methodologies produce different results (Zieky, 2001)
 - If differences are too large, the validity of the process may be questioned (Kane, 2001)



Other Validity Evidence

- Another consideration is the reasonableness of the cut score
 - Do the impact data make sense?
 - Are the results what stakeholders expected?
 - How do they compare to results from the general assessment?
- Try to get multiple opinions on the reasonableness of the results using the adopted cut scores

Consequential Validity

- What effect is the use of the performance levels having on student opportunity to learn?
 - Teacher professional development
 - Pedagogical changes
 - Student exposure to richer curriculum
- Are there any unintended consequences from using the performance levels?
 - Certain content standards being de-emphasized
 - Certain students being neglected
 - Teacher retention decreasing



Validation Over Time

- Follow students over time
 - Is improvement captured by the performance levels?
 - Do the classifications over time make sense? (e.g., results that show a student to be proficient in grade 3, basic in grade 4, below basic in grade 5 and proficient in grade 6 would not be considered reasonable)
- Survey teachers/other stakeholders over time
 - Are performance levels capturing true achievement?
 - Do they see possibility for improvement for the students?
 - What suggestions do they have for modifying PLDs and/or cut scores?