Report Number: 5
Standards and Assessments

CBAS-R: Evaluating and Aligning Standards and Assessments

Allison McCarthy
Theodore Christ
University of Minnesota

Date Initiated: 07/09/09
Finalized Draft: 10/12/10
Reviewed by Advisory Board: TBA

Preparation of this technical report was supported in part by a grant from the Office of Special Education Programs, U.S. Department of Education (H327A060014). Opinions expressed herein do not necessarily reflect the position of the U.S. Department of Education.
Abstract
Building an assessment that is instructionally relevant and useful in classrooms requires its alignment with standards of education. CBAS-2 aims to inform and align with instruction by including five domains of reading. The representation of the domains concurs with the need for content balance on an assessment. The standards for evaluating alignment as well as methodology useful for obtaining the objective of a standard-aligned assessment are reviewed.

*Keywords: assessment, standards, instruction*
<table>
<thead>
<tr>
<th>Activity</th>
<th>Lead Person</th>
<th>Timeframe</th>
</tr>
</thead>
<tbody>
<tr>
<td>Recommendations to PI’s</td>
<td>Allison McCarthy</td>
<td>July 2009</td>
</tr>
<tr>
<td>Writing</td>
<td>Allison McCarthy</td>
<td>July 2009</td>
</tr>
</tbody>
</table>
## Table of Contents

Abstract .................................................................................................................................................. 2  
Introduction........................................................................................................................................ 5  
Methodology....................................................................................................................................... 5  
Item Development................................................................................................................................. 5  
  Universal Design.............................................................................................................................. 5  
Item Review......................................................................................................................................... 6  
  Expert Review under UDA ............................................................................................................... 6  
Statistical Analyses............................................................................................................................. 6  
  Think Aloud Techniques .................................................................................................................. 6  
Alignment Criteria.............................................................................................................................. 6  
References........................................................................................................................................... 8
CBAS-R: Evaluating and Aligning Standards and Assessments

Standards are developed so that all educational stakeholders agree upon what students should know and be able to do (Valencia & Wixson, 2001). Then, everything else that is valued in the system should follow, including assessments, professional development, and materials (Valencia & Wixson, 2001). State standards for primary grade language arts are either too broad or too specific. Those benchmarks that are too general read as broad goals rather than guidelines for what students should know (Wixson & Dutro, 1999). In contrast, specific benchmarks are more akin to lists of skills and activities that teachers should use for instruction, which does not allow for flexibility (Wixson & Dutro, 1999). A balance between specificity and generalization for the primary grades (K-3) would look like guidelines that show a progression from one grade to the next. Importantly, there needs to be a balance between guidance and allowing for flexibility that each teacher would need to make instruction effective for a specific classroom (Wixson & Dutro, 1999).

Methodology

There is a lot to be considered when determining if an assessment is appropriate to be used as a statewide test. Items must undergo review and should comprehensively represent standards developed in that core area. A few methodologies are used for this process and described below.

Item Development

Universal Design. Universal Design of Assessment (UDA) means that an assessment can appropriately assess a wide range of students and the results obtained can be used to make inferences about all students assessed (Johnstone, Thompson, Bottsford-Miller & Thurlow, 2008). Developing a test under UDA means that the test is inclusive of populations, measures
what its creators intended, bias is mostly eliminated, has clear instructions and procedures, and can be adapted to accommodations (Johnstone et al., 2008).

**Item Review**

**Expert Review under UDA.** Experts specifically focus on minimizing bias and ensure that items align with curriculum (Johnstone et al., 2008). To review items in light of a large-scale test developed in the UDA framework, a table of considerations was created (See page 4).

**Statistical Analyses.** Descriptive statistics are used to examine the distribution of scores on assessments (Johnston et al., 2008). Differential item Functioning (DIF) analysis is also used to determine if an item is not producing expected outcomes based on the individual student’s overall score. This might help with examining bias in an item or test (Johnstone et al., 2008).

**Think Aloud Techniques.** This is used instead of statistical analysis to determine why items are difficult or challenging in an unintended manner. This technique sounds exactly like what it is: students talk through the item while a reviewer listens to the problems they are having. This is very time consuming (Johnstone et al., 2008) and not feasible for CBAS.

**Alignment Criteria**

Accurate and appropriate alignment of standards and assessments provides information about how students have performed and what improvements need to be made for instruction and for individual students (Webb, Herman, & Webb, 2007). When standards and assessments are not appropriately aligned, teaching to the test becomes more common and test results provide no useful information. Thus, alignment criteria are a necessity when examining a statewide assessment and the use of these criteria has become more widespread. One of the most commonly used set of criteria are those developed by Webb (Webb model), which guides reviewers to rate the “depth-of-knowledge level” for each item (Webb et al., 2007). Reviewers
are content experts, content area supervisors at the district level, and content area teachers (Webb, 2007). Reviewers rate categorical concurrence, depth-of-knowledge consistency, and range of knowledge match between the standards and assessments (Webb et al., 2007; Webb, 2007). Categorical concurrence refers to whether the same content is found in the standards and the assessment. This is at an acceptable level if, across reviewers, an average of six items correspond to the content of one standard (Webb et al., 2007; Webb, 2007). Depth-of-knowledge consistency refers to the correspondence between the ‘cognitive demands’ of the standard and an assessment. This is at an acceptable level if at least 50% of items are at or above the depth-of-knowledge (difficulty) level of the standards (Webb et al., 2007). There are four levels of cognitive demand: (1) recall; (2) skill/concept; (3) strategic thinking; and (4) extended thinking (Webb, 2007). Range of knowledge refers to the levels of knowledge that students need to respond correctly to assessment items (Webb et al., 2007). The Webb model also suggests that there should be a balance of representation of knowledge when standards and assessments are aligned (Webb, 2007). If a rater determines that there is more emphasis on one standard than another, this criterion receives a low score (Webb, 2007).

Using this set of criteria, reviewers found that reliability of the alignment process would improve once clear standards are provided. Alignment can be improved once standards are improved (Webb et al., 2007). However, issues remain about when an assessment is well enough aligned with standards (Webb, 2007).
References


