



## Assessments Based on Modified Academic Achievement Standards: Critical Considerations and Implications for Implementation

*Assessment and Accountability Comprehensive Center  
Special Populations Strand  
November 2007*

The AMAAS are intended to help address the requirement of the No Child Left Behind Act of 2001 (NCLB) that all students be included in state assessment and accountability systems.

This paper, prepared by the Special Populations Strand of the Assessment and Accountability Comprehensive Center (AACC), presents information about assessments based on modified academic achievement standards (AMAAS). The AMAAS are intended to help address the requirement of the No Child Left Behind Act of 2001 (NCLB) that all students be included in state assessment and accountability systems.

The AMAAS are intended for the population of students for whom the state's regular grade-level assessment, even with accommodations, does not provide appropriate access<sup>1</sup> to the assessed

<sup>1</sup> Access refers to the minimization or removal of conditions within a testing situation (sources of construct-irrelevant variance such as aspects of presentation/format of test information, aspects of response requirements, etc.) that may interfere with students' ability to demonstrate their content knowledge and skills (construct-relevant information), without significantly changing the assessed construct. Strategies that facilitate access are tailored to the particular needs of students (e.g., cognitive, linguistic, physical). When access is constrained, it could result in the measurement of sources of variance that are not related to the intended test constructs (construct irrelevance). Limited access could allow construct-irrelevant abilities to interfere with a student's ability to fully demonstrate what he or she knows and can do; consequently, the test results

content. Inappropriate access results when factors not related to the tested content (construct-irrelevant) interfere with a student's ability to fully demonstrate what he or she knows and can do. Consequently, the test results can underestimate the student's achievement. On the other hand, the state's alternate assessment based on alternate achievement standards (for the most severely cognitively disabled students) is too restrictive in breadth, depth, and range of complexity for the population of students targeted by the AMAAS. Thus, students could be excluded from a fuller range of grade-level skills and concepts, constraining their potential level of achievement, and underestimating their full potential. The expectation for the AMAAS is that with appropriate accommodations, specific and purposeful modifications to the test items and format, and careful attention to the range of grade-level content appropriate for students vis-à-vis their abilities/disabilities, the access needs of these students will be appropriately addressed and their academic achievement will be more accurately measured.

With the recently released regulations and non-regulatory guidance for the AMAAS (Federal Register, April 9, 2007), this paper is intended to

underestimate that student's construct-relevant achievement. Inappropriate access could affect the construct such that the assessment no longer sufficiently represents the assessed domain (underrepresentation); that is, a student's proficiency on the assessment cannot be generalized to proficiency across the domain (Johnstone, 2003; Messick, 1993; Sato, Rabinowitz, & Gallagher, forthcoming).

help Regional Comprehensive Centers (RCCs) and states develop, implement, and evaluate their AMAAS by:

- Highlighting key information about the AMAAS (from the regulations, non-regulatory guidance, and other relevant resources);
- Describing critical issues and consider-

ations related to the development and implementation of the AMAAS; and

- Presenting examples of general approaches states have taken in the development and implementation of their AMAAS.

Information presented about the AMAAS will be updated as new, relevant

research, guidance, and strategies become available for consideration and evaluation by the AACC. Additionally, information will be updated to meet the evolving needs of RCCs and states vis-à-vis AMAAS development and implementation.

### Assessments Based on Modified Academic Achievement Standards: Overview<sup>2</sup>

Key Information	Implications/Considerations <sup>3</sup>
What are alternate assessments based on modified academic achievement standards (AMAAS)?	
AMAAS are assessments that cover grade-level content.	In order for students to have an opportunity to achieve grade-level content, they must have access to and instruction in grade-level content.  Note: The <i>achievement</i> standards are modified, not the <i>content</i> standards. (See relevant information below.)
AMAAS differ from alternate assessments that are based on <i>alternate</i> academic achievement standards because (a) AMAAS are aligned to grade-level content standards and (b) proficiency must occur on grade-level content.	The AMAAS must <i>align</i> <sup>4</sup> to grade-level content standards. An alternate assessment based on alternate academic achievement standards typically <i>links</i> to grade-level content standards.
A state may modify an existing assessment or develop a new one. Examples of modifying an existing assessment to create AMAAS include: <ul style="list-style-type: none"> <li>• Replacing the most difficult items on the state's general education assessment with less complex items that still test grade-level content;</li> <li>• Simplifying the language of items;</li> <li>• Removing a distracter in multiple choice items (i.e., presenting the student with fewer answer choices);</li> <li>• Using technology; and</li> <li>• Allowing the student to dictate his/her responses.</li> </ul>	States have options in designing the AMAAS.  Modifications to AMAAS (items/test) should be specific and purposeful relative to the particular access needs of the full range of students eligible for this assessment.  Out-of-level assessments cannot be used.  Similar to any other state assessment under Title I, AMAAS must satisfy expected and reasonable technical requirements—validity, reliability, freedom from bias, etc.
States are not required to develop AMAAS.	States must have at least one alternate assessment in which students with disabilities can participate. With these new regulations, there is an assumption that this will provide more flexibility regarding state, local educational agency (LEA), and school accountability for the achievement of a small group of students with disabilities whose progress is such that, even after receiving appropriate instruction, including special education and related services designed to address the students' individual needs, the students' individualized education program (IEP) teams are reasonably certain that the students will not achieve grade-level proficiency within the year covered by the students' IEPs.
States must determine what is needed to meet the requirements for a regular high school diploma.	States cannot automatically preclude students from earning a regular high school diploma because they are eligible for and are administered the AMAAS.

<sup>2</sup> The information listed is based on the following sources:

- Title I—Improving the Academic Achievement of the Disadvantaged; Individuals with Disabilities Education Act (IDEA); Final Rule, 72 Fed. Reg. 17,748 (April 9, 2007) (to be codified at 34 C.F.R. pt. 200 and 300).
- U.S. Department of Education. (2007). *Measuring the Achievement of Students with Disabilities*. Washington, DC: Author.
- U.S. Department of Education. (2007). *Modified Academic Achievement Standards: Non-Regulatory Guidance*. Washington, DC: Author.

<sup>3</sup> Implications/Considerations are not inclusive and are discussed further in a subsequent section of this document.

<sup>4</sup> **Alignment** refers to the degree to which the content (e.g., skills, concepts) in two sets of standards or in an assessment and a set of standards match in terms of breadth, depth, and range of complexity. Alignment relationships tend to be direct relationships (matches between skills, content). **Linkage** refers to relationships that tend to be developmental, foundational, or proximal between standards and/or assessments (WestEd, 2004; Sato, Lagunoff, Worth, Bailey, & Butler, 2005; Bailey, Butler, & Sato, 2007).

Key Information	Implications/Considerations
<b>What are modified academic achievement standards?</b>	
<p>Modified academic achievement standards (i.e., “performance standards”) are challenging for eligible students and reflect students’ mastery of grade-level content; however, they are less difficult than grade-level achievement standards.</p> <p>Modified academic achievement standards cannot be clustered by grade ranges (e.g., 3-5, 6-9, 10-12); however, alternate assessment achievement standards can be clustered by grade ranges.</p>	<p>Note: Academic <i>achievement</i> standards (i.e., “performance standards”) are not the same as academic <i>content</i> standards.</p> <p>Modified academic achievement standards must:</p> <ul style="list-style-type: none"> <li>• be aligned to the state’s academic content standards;</li> <li>• be challenging for eligible students, but reflect a less rigorous expectation of mastery of grade-level academic content;</li> <li>• describe at least three levels of achievement;</li> <li>• include descriptions of the competencies associated with each achievement level; and</li> <li>• include assessment scores (cut scores) that differentiate among the achievement levels.</li> </ul> <p>States should have clear descriptions and rationale for their modified academic achievement standards.</p>
<b>Which students should be assessed with the AMAAS?</b>	
<p>Students with disabilities under section 602(3) of the IDEA should be assessed with AMAAS. These students may be in any of the disability categories and may represent a wide spectrum of abilities.</p> <p>IEP teams decide how a student participates in the state’s assessment system.</p>	<p>The state must establish clear and appropriate criteria for IEP teams to apply in order to determine whether a student is eligible for the AMAAS. Criteria must include:</p> <ul style="list-style-type: none"> <li>• Objective evidence demonstrating that a student’s disability has precluded her/him from reaching grade-level proficiency in the content area assessed (e.g., student performance on the state assessment or other assessments that can validly document academic achievement);</li> <li>• The student’s progress in response to appropriate instruction, including special education and related services designed to address the student’s individual needs. This progress (as determined by performance on multiple valid measures over time) is such that, even if significant growth occurs, the IEP team is reasonably certain that the student will not reach grade-level proficiency within the year covered by the IEP; and</li> <li>• IEP goals for the subject assessed are based on academic content standards for the grade in which the student is enrolled.</li> </ul> <p>States should develop a process to ensure that students are being appropriately identified for the AMAAS.</p> <p>Note: A student may take the AMAAS in one content area and take the general assessment in another content area, based on the student’s IEP and determination of the IEP team.</p>
<b>To what does the “2%” refer?</b>	
<p>“2%” refers to the percentage of <i>proficient</i> and <i>advanced</i> scores of students taking the AMAAS that can be counted toward state AYP calculations for NCLB.</p> <p>If <i>proficient</i> and <i>advanced</i> scores exceed 2% of all the students tested on the AMAAS, then those scores exceeding the 2% cap will be counted as <i>non-proficient</i> for calculating AYP.</p> <p>“2%” does <i>not</i> mean that only 2% of the student population can be administered the AMAAS. Nor does “2%” necessarily refer to the 2% of students performing just above the most significantly, cognitively disabled student population.</p>	<p>States should clearly define and disseminate the guidelines for determining how to count proficient scores.</p> <p>States also should establish procedures for monitoring implementation/practice.</p>

## *Critical Issues Related to Assessment Development and Implementation<sup>5</sup>*

States are at varying levels of development and implementation of their AMAAS. As states develop and implement their AMAAS, they should be aware of a number of critical technical and logistical issues.<sup>6</sup>

- Issue 1: Determining Who Should Participate in the AMAAS
- Issue 2: Ensuring the Appropriate Content is Measured
- Issue 3: Providing Appropriate and Sufficient Access for AMAAS Students
- Issue 4: Defining Proficient Performance on the AMAAS
- Issue 5: Ensuring Technical Adequacy

Each issue is described in greater detail below.

### *Issue 1: Determining Who Should Participate in the AMAAS*

Fundamental to designing a test that is valid for this population is having a clear understanding of the students who will be assessed. Without a clear understanding of *who* the test is designed to assess, it is difficult to know *how* to validly assess them.

<sup>5</sup> This discussion of critical issues is based on Sato, E. & Rabinowitz, S. (forthcoming). *Assessing Students in the Gap: Critical Issues and Considerations Impacting the Validity of Assessments Based on Modified Academic Achievement Standards* [working title]. San Francisco: WestEd.

<sup>6</sup> The issues are similar to those faced by states while developing alternate assessments based on alternate achievement standards because there is overlap on some of the characteristics of the significantly cognitively disabled students (i.e., 1% population) and the AMAAS population. However, the AMAAS population has unique characteristics and needs that necessitate differences in how these challenges ought to be addressed.

Without a clear understanding of *who* the test is designed to assess, it is difficult to know *how* to validly assess them.

The articulation of eligibility criteria for the AMAAS should be a result of carefully planned work with committees consisting of experts in assessment (both traditional and alternate/performance-based), standards, instruction and curriculum, and the broad special education student population. Ancillary materials and professional development also should be provided by the state to help local education agency (LEA) staff understand and apply these criteria.

Even with thoughtful development and dissemination of materials and professional development to ensure the proper identification of students, states still need to monitor the accuracy and consistency with which students are identified for the assessment.

Even with thoughtful development and dissemination of materials and professional development to ensure the proper identification of students for the AMAAS, states still need to monitor the accuracy and consistency with which LEAs are identifying the intended students for the assessment. Relatively little is known about the effectiveness and validity of the guidelines established by states and whether states' policies and procedures result in consistent and accurate decisions about participation (Almond & Bechard,

2005; Yovanoff & Tindal, 2007). For example, the labels used to determine student eligibility for AMAAS (e.g., those included in the U.S. Department of Education Code of Federal Regulations, 300.7) can be highly subjective due to potential variations in how these labels are understood and applied across local sites. Different levels of experience and familiarity of IEP teams with properly identifying students, especially students who have multiple disabilities or are also English Language Learners (ELLs), also can confound the proper identification of students and subsequently threaten the validity of the assessment. It is essential that LEAs correctly identify students for participation in the AMAAS; however, little is known at this time about whether the intended students are actually being identified.

## *Issue 2: Ensuring the Appropriate Content is Measured*

The recent AMAAS regulations require that students have access to the general education curriculum and grade-level academic content standards. Alignment among content standards, curriculum, instruction, and assessment is critical to the validity of assessments for accountability (National Research Council, 2001). States

**Alignment among content standards, curriculum, instruction, and assessment is critical to the validity of assessments for accountability.**

should work with committees (described previously) to review state grade-level content standards, verify that the targeted student population has the capacity to learn

the content *when properly instructed*, and determine whether it is reasonable to expect the student population to have adequate opportunities to learn the content.

Towles-Reeves, Muhomba, and Kleinert (2006) reviewed 18 studies conducted between 2002 and 2006 examining the readiness of special education teachers to implement a standards-based curriculum. Their study suggests a disconnect between the intent to provide instruction based on academic content to students with disabilities and teachers' ability to implement such practices. Although these studies focused on the 1% student population, it brings to question whether teachers of the AMAAS population are providing these students with sufficient opportunity to learn grade-level content. Since there is overlap on some of the characteristics of the 1% and AMAAS student populations, these students may not have adequate opportunity to learn such that they can reasonably and fairly be held accountable for their knowledge of the assessed content.

Providing students with the opportunity to learn is critical to the validity of an assessment. Therefore, states should examine the degree to which appropriate grade-level content is being taught and how this aspect of implementation impacts the validity of the states' AMAAS. States also should examine the degree to which the tests themselves properly align with the targeted grade-level content standards. As part of these alignment studies, states should document how AMAAS are similar to and different from existing non-modified assessments. This process will be an important step in validating the "modified" assessment and will help evaluate the results of standard setting for the AMAAS (see Issue 4).

### *Issue 3: Providing Appropriate and Sufficient Access for AMAAS Students*

Consistent with the provisions of the Individuals with Disabilities Education Act (IDEA, 2004), the AMAAS regulations are intended to help ensure that all students with disabilities have access to the general curriculum. Although appropriate access to test content is an issue for all students, it is a critical concern for students with disabilities. Historically, the population of students for which the AMAAS is intended

Although appropriate access to test content is an issue for all students, it is a critical concern for students with disabilities.

has been largely overlooked and underserved. Their disabilities typically preclude their reaching grade-level achievement standards in the same time frame as other students, but their disabilities are not severe enough to make the alternate assessment based on alternate achievement standards (i.e., 1% assessment) appropriate. Therefore, as a measure of their progress and achievement, they have had to take either a grade-level assessment with the same format and achievement levels as their general education counterparts or a 1% assessment based on content below their cognitive capacity; neither option accurately reflects what these particular students know and can do. The grade-level assessment (even with accommodations) does not provide the AMAAS student population with appropriate access to the assessed content. Inappropriate access results in the measurement of

sources of variance that are not related to the intended test content (*construct irrelevance*). Inappropriate access also allows construct-irrelevant factors to interfere with a student's ability to fully demonstrate what he or she knows and can do, and subsequently the test results underestimate the student's achievement. The 1% assessment is too restrictive (breadth, depth, range of complexity) for the AMAAS student population, thereby excluding them from a fuller range of grade-level skills and concepts, constraining their potential level of achievement, and underestimating their full potential.

States should work with committees (described previously) to specify the condi-

States should work with committees to specify the conditions within a testing situation that might interfere with students' abilities to demonstrate their content knowledge and skills.

tions within a testing situation (sources of construct-irrelevant variance such as aspects of presentation/format of test information, aspects of response requirements, etc.) that might interfere with students' abilities to demonstrate their content knowledge and skills (construct-relevant information). These committees should be involved in key aspects of the assessments' development and implementation, including the creation of item specifications used to develop items for the assessed indicators/grade-level expectations and the review of assessment items to ensure students' access needs (cognitive, linguistic, physical) were appropriately addressed and the intended grade-level constructs were not significantly changed.

States should work with committees to ensure students' access needs (cognitive, linguistic, physical) are appropriately addressed and the intended grade-level constructs are not significantly changed.

Once content decisions have been made, states then need to review the full range of accommodations that may be made available to students taking the AMAAS. Accommodations may be based on presentation, response, setting, and timing considerations (Thompson, Blount, & Thurlow, 2002). As with all testing programs, allowable AMAAS accommodations must be IEP-based and consistent with the mode of instruction for the student in question.

Lack of appropriate access has consistently created the greatest threat to validity because a lack of appropriate access can contribute to construct-irrelevant variance and under-representation (Abedi, Courtney, & Leon, 2003; Abedi, Hofstetter, & Lord, 2004; Bielinski, Sheinker, & Ysseldyke, 2003; Elliot, Kratochwill, McKeivitt, Schulte, Marquart, & Mroch, 1999; Helwig, Rozek-Tedesco, Tindal, Heath, & Almond, 1999; Kopriva, Samuelson, Wiley, & Winter, 2003; Sireci, Li, & Scarpato, 2003; Thurlow & Wiener, 2000). Despite expert judgments of accessibility and maintenance of the integrity of the content assessed, the *actual* strategies students use to access the content of each assessment remains unknown. Since the effectiveness of current practices for making high-stakes assessments accessible is unclear, the appropriateness of inferences about student test performance may be compromised when assessment conditions

are changed to increase access to students with special needs (Hafner, 2001; Tindal & Fuchs, 2000). States should examine the *actual* effects on students of the access strategies that result from their item/test design, development, and implementation (see relevant discussion in *General Approaches: Examples*).

#### *Issue 4: Defining Proficient Performance on the AMAAS*

The AMAAS regulations allow states to establish modified academic achievement standards for the AMAAS student population. While the assessment itself must be aligned with the state's grade-level academic content, the achievement standards (i.e., performance levels) may be set to reflect the performance expectations of these students who are not expected to reach grade-level achievement within the year covered by their IEP because of their disabilities. To implement these regulations, states need to both clearly define the content (breadth, depth, complexity) to include on the assessment and decide how mastery of this content will be defined vis-à-vis the student population (disabilities, capacities). States should work with their assessment and special education experts to develop performance level descriptors that reflect expectations of grade-level content proficiency for the AMAAS student population.

There are a number of challenges specific to setting performance standards on alternate assessments for students with disabilities (Almond & Bechard, 2005; Thompson, Johnstone, Thurlow, & Altman, 2005). Similar to the 1% student population, the AMAAS population has

States should work with their assessment and special education experts to develop performance level descriptors that reflect expectations of grade-level content proficiency for the AMAAS student population.

There are a number of challenges specific to setting performance standards on alternate assessments for students with disabilities.

great diversity; thus the applicability of one standard for all is problematic. Equally challenging is the relatively small number of students taking the assessment, which makes obtaining sufficient data sets difficult for complex standard-setting procedures. Therefore, states should have external validation of where the performance standards have been placed, such as determining:

- Whether students who are proficient in grade X are more likely to be proficient in grade X+1; and
- Whether achievement standards across grades and content areas are coherent—identifying expected percentages of students across grades and content areas.

As with all standard-setting procedures, states should develop and validate achievement level descriptors (ALD) across the full range of committees involved in the development of AMAAS. ALDs are statements that indicate what students are expected to know and do at the state's various achievement levels (e.g., basic, proficient). States should then incorporate these validated ALDs throughout the standard setting process.

As with other assessments, various methods are available to states to actually determine the modified achievement standards. Some involve committee reviews (e.g., Bookmark [Lewis, Mitzel, & Green, 1996]); others are more judgment-based (e.g., Contrasting Groups [Livingston & Ziekey, 1982]). States should consult with their Technical Advisory Committee to determine the most appropriate method or combination of approaches, keeping in mind the difficulties inherent in setting standards for this population (e.g., diversity,

limited population) and assessment purpose. Finally, states should examine the accuracy with which the modified academic achievement standards reflect both the expectations for and the performance capabilities of the AMAAS student population.

States should examine the accuracy with which the modified academic achievement standards reflect both the expectations for and the performance capabilities of the AMAAS student population.

### *Issue 5: Ensuring Technical Adequacy*

NCLB assessment requirements for students with disabilities are unprecedented for state assessment programs. Assessments with accountability consequences must meet rigorous technical quality criteria and provide valid information for their intended purposes. Results from federal peer review reports on alternate assessments based on alternate achievement standards have identified serious technical shortcomings of these assessments (U.S. Department of Education, 2006) and suggest that states may need extra guidance on and support for establishing the technical adequacy of their AMAAS assessments.

States must recognize the need to accommodate the wide range of disabilities and associated needs of this heterogeneous AMAAS student population in order to ensure valid measures of what these students know and can do. States should develop, implement, and monitor procedures to ensure that their AMAAS assessments are representa-

tive of what students have been taught, yield consistent, accurate scores, and yield outcomes that are congruent with

States should develop, implement, and monitor procedures to ensure that their AMAAS assessments are representative of what students have been taught, yield consistent, accurate scores, and yield outcomes that are congruent with the uses of the results.

the uses of the results. States also should consider adapting aspects of their general education assessments and procedures to build their AMAAS assessments, so that to the degree appropriate and possible, much of the demonstrated technical qualities of their general education assessments transfer.

According to Yovanoff and Tindal (2007), few models exist for determining the technical adequacy (validity, reliability, freedom from bias) of alternate assessments, and few research studies have been published on the technical adequacy of alternate or modified assessments. There also is little consensus on exactly what evidence should be collected and how much of this evidence would be sufficient to ensure technical adequacy. While criteria for evaluating the validity of general education assessments have been written about extensively (AERA, APA, & NCME, 1999; Green, 1998; Messick, 1993; Webb, Horton, & O'Neal, 2002), research suggests that technical criteria appropriate for assessments of the general education population do not fully transfer to assessments of special student populations (Rabinowitz & Sato, 2005,

2006). The technical evidence supporting state AMAAS should be evaluated for appropriateness and adequacy.

### **Key Considerations**

Therefore, key considerations that states face as they develop and implement the AMAAS include:

- Ensuring that the population of students eligible for the AMAAS is distinct enough from students who take the alternate assessment (i.e., the 1% population) and those who take the regular assessment with accommodations—there are substantive characteristics that distinguish the AMAAS student population;
- Determining what percentage of students will be eligible for and administered the AMAAS, even if only 2% can be used for accountability purposes;
- Ensuring that the content considerations and the design of the items and assessments effectively address the access needs of the students such that the assessment is a valid and reliable measure of what these students know and can do, inferences from results are valid, and the content remains at grade level;
- Ensuring that students have had adequate opportunity to learn the content assessed in the manner it is assessed;
- Ensuring that the AMAAS is aligned to grade-level content standards, comparable to the regular assessment, and that the accommodations and changes made to the items/test have not significantly changed the grade-level content/constructs; and
- Determining whether the cut scores for proficiency and other achievement levels should be set using a population-specific scale or the general education scale.

### General Approaches: Examples

The following are examples of how some states have begun to address (or plan to address) key issues related to the development and implementation of their AMAAS.<sup>7, 8</sup>

<sup>7</sup> State procedures are most effective when they rely on multiple methods and multiple data sources to address key issues. The examples listed are not inclusive.

<sup>8</sup> This discussion of general approaches and examples is based on Sato, E. & Rabinowitz, S. (forthcoming), *Assessing Students in the Gap: Critical Issues and Considerations Impacting the Validity of Assessments Based on Modified Academic Achievement Standards* [working title]. San Francisco: WestEd.

### Population Definition

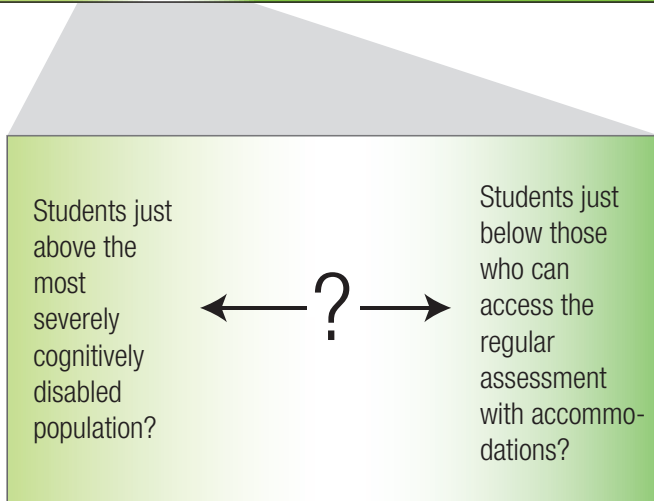
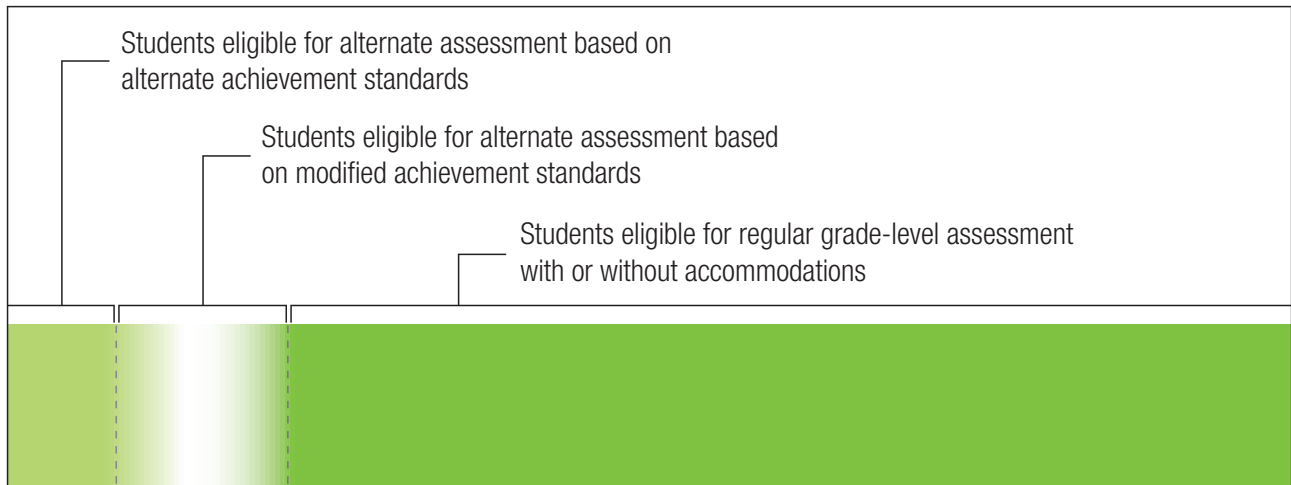
The AMAAS population of students tends to fall into one of three general categories:

- Students just above the most severely cognitively disabled population, for whom the alternate assessment based on alternate achievement standards is not appropriate. These students have IEPs, but their IEPs do not designate these students as eligible for the state’s alternate assessment based on alternate achievement standards.
- Students just below those who can access the regular assessment with accommodations. These students may have

additional access needs that may not be addressed through the state-approved accommodations. With some additional accommodations, these students could be expected to access and demonstrate proficiency on grade-level content.

- Students who fall between the two groups mentioned previously—students who cannot access the regular assessment, even with accommodations, yet exceed the expectations of students with severe cognitive disabilities. These students need an assessment with modifications to the achievement standards that reflect their particular abilities vis-à-vis grade-level academic content expectations.

### Who are the AMAAS students?



The AMAAS student population is diverse in its capacities and access needs. Methods used to inform a state's definition of the AMAAS student population include expert judgment, surveys, and observations.

No matter how the population is defined, the AMAAS student population is diverse in its capacities and access needs. Methods used to inform a state's definition of the AMAAS student population include the following.

**Expert judgment** has been used to define the AMAAS student population and establish eligibility criteria, item/test design vis-à-vis student abilities/disabilities and access needs, standard setting, etc. Experts need to have in-depth knowledge of assessment design and validation (for both traditional and alternate/performance-based), measurement, standards development and implementation, instruction and curriculum development, IEPs, and the special education student population. These experts are convened in committees so that cross-disciplinary discussions can occur to address the complex issues related to the heterogeneity of the population, the population's range of access needs, and issues of validity, reliability, and fairness.

**Surveys** have been developed as a means for soliciting information from the field regarding the AMAAS student population. Surveys also are used as an indicator of how accurately and consistently criteria and procedures for student identification are being implemented.

**Observations** also have been used to

gather information about the AMAAS students and the degree to which these students are accurately identified.

Using methods like the ones listed above can help validate the state's definition of the AMAAS student population and inform the assessment of these students.

### *Content Parameters*

For the AMAAS, students' capacities and access needs should be considered in relation to the state's grade-level academic content. States are expected to maintain the breadth and depth of grade-level content; however, the range of complexity can be modified. For example, many states use Webb's Depth of Knowledge (DOK) levels to designate levels of complexity: Recall, Basic Application, Strategic Thinking, and Extended Thinking (Webb, 1999). On the state's regular assessment, the full range of DOK may be reflected; however, on the state's AMAAS, a more restricted range of DOK (e.g., primarily Recall and Basic Application with limited Strategic Thinking and no Extended Thinking) may be represented. Methods used to inform the content parameters of the AMAAS include the following.

**Expert judgment** has been used to define the general content parameters for the AMAAS (breadth, depth, range of complexity), develop item/test specifications, verify content appropriateness, etc. Experts need to have in-depth knowledge of assessment design and validation (for both traditional and alternate/performance-based), measurement, standards development and implementation, instruction and curriculum development, IEPs, and the special education student population.

For the AMAAS, students' capacities and access needs should be considered in relation to the state's grade-level academic content. Methods used to inform the content parameters of the AMAAS include expert judgment, alignment studies, and statistical analyses.

These experts are convened in committees so that cross-disciplinary discussions can occur to address the complex issues related to the population's range of abilities/disabilities, the grade-level content accessible to and appropriate for the students, and issues of validity, reliability, and fairness.

**Alignment studies** have been conducted to determine the degree to which the AMAAS items align to the grade-level academic content standards. Additionally, the breadth and depth of grade-level content coverage have been examined. It is critical that an independent alignment study be conducted at some point in the development of the AMAAS by analysts who collectively possess expertise in assessment (both traditional and alternate/performance-based), measurement, standards, instruction and curriculum, and the special education student population.

**Statistical analyses** such as classical measurement statistics, including test means, standard deviations, standard errors, p-values, discrimination indices, point biserials, and omission rates, have been examined to inform decisions about the appropriateness of the content and to determine whether access has indeed been increased. If sample sizes permit, IRT parameters also have been reviewed. Additionally, the relationship of AMAAS student performance with other outcome measures (e.g., course grades, performance on other tests such as previously administered general education assessments, teacher evaluation of student academic achievement) has been examined.

There is an interaction between AMAAS student population character-

istics (e.g., abilities/disabilities, access needs) and the content assessed. Using methods like the ones listed above can help validate the appropriateness of the grade-level academic content included on the AMAAS.

### **Access**

As mentioned previously, the regular general education assessment and the AMAAS are similar in breadth and depth of grade-level content coverage. Differences between the regular and AMAAS assessments include a range of possible item format and page layout changes to the AMAAS that are intended to facilitate students' access to the assessed content so that they can better show what they know and can do. Such changes include, for example, modifications to item layout and formatting that are intended to address student processing needs—increased white space around items, separating sentences onto different lines, etc. Modifications also are implemented to address processing load/demand—multiple-choice items have three rather than four answer choices, texts are shorter and less complex, etc. Methods used to inform the degree to which changes to the AMAAS actually increase student access as intended (i.e., actual vs. intended effect of changes to items on student access) include the following.

**Access is a major threat to validity of assessments for this student population. Methods used to inform the degree to which changes to the AMAAS actually increase student access as intended include expert judgment and cognitive interviews.**

There is an interaction between AMAAS student population characteristics (e.g., abilities/disabilities, access needs) and the content assessed.

**Expert judgment** has been used to verify the degree to which changes to the AMAAS appropriately address student access needs. Experts need to have in-depth knowledge of assessment design and validation (for both traditional and alternate/performance-based), measurement, standards development and implementation, instruction and curriculum development, IEPs, and the special education student population. These experts are convened in committees so that cross-disciplinary discussions can occur to address the complex issues related to student access and issues of validity, reliability, and fairness.

**Cognitive interviews** have been conducted to examine the accessibility of test items. For the cognitive interview, trained researchers or practitioners guide individual students through a think-aloud protocol as students work on items. Students receive training on and are given the opportunity to practice the think-aloud protocol prior to the administration of the actual test items. Once the student is trained and the researcher is confident that the student understands the requirements of the task, the student will complete think-alouds for a purpose-

fully selected set of items.

Some states have used their existing regular assessment item pool and selected items that meet their AMAAS item/test specifications, making slight modifications to the items as necessary and appropriate. Other states have developed items specifically for their AMAAS. An approach that combines the selection of items from an existing item pool with the development of items specifically for the AMAAS also has been used in order to ensure appropriate coverage of content. Regardless of a state's approach to the development of its AMAAS, methods like the ones listed above can help validate the manner in which a state addresses the access needs of the AMAAS student population. The effectiveness with which a state addresses student access needs has real implications for the validity of the assessment's results.

As mentioned at the beginning of this paper, the Assessment and Accountability Comprehensive Center (AACC) will update these guidelines as new, relevant research, guidance, and strategies become available. Additionally, these guidelines are designed to evolve with the changing needs of RCCs and states. ❖

*For questions related to these guidelines, please contact:*

*Edynn Sato, Ph.D.*

*Director, Special Populations  
Assessment and Accountability  
Comprehensive Center, WestEd  
email: [esato@wested.org](mailto:esato@wested.org).*

*For more information  
about the AACC, visit  
[www.aacompcenter.org](http://www.aacompcenter.org).*

## References

Abedi, J., Courtney, M., & Leon, S. (2003). *Research-supported accommodation for English language learners in NAEP* (CSE Technical Report). Los Angeles: University of California, Los Angeles; Center for the Study of Evaluation; National Center for Research on Evaluation, Standards, and Student Testing.

Abedi, J., Hofstetter, C., & Lord, C. (2004). Assessment accommodations for English language learners: Implications for policy-based empirical research. *Review of Educational Research*, 74(1), 1–28.

Almond, P., & Bechar, S. (2005). *In-depth look at students who take alternate assessments: What do we know now?* Retrieved June 13, 2007, from <http://www.measuredprogress.org/resources/inclusive/articlespapers/StudentsTakingAltAssess.pdf>

*The contents of this report were developed under a grant from the Department of Education. However, those contents do not necessarily represent the policy of the Department of Education, and you should not assume endorsement by the Federal Government.*

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: AERA.

Bailey, A., Butler, F., & Sato, E. (2007). Standards-to-standards linkage under Title III: Exploring common language demands in ELD and science standards. *Applied Measurement in Education*, 20(1), 53–78.

Bielinski, J., Sheinker, A., & Ysseldyke, J. (2003, April). *Varied opinions on how to report accommodated test scores: Findings based on CTB/McGraw-Hill's framework for classifying accommodations* (NCEO Synthesis Report 49). Minneapolis, MN: National Center on Educational Outcomes.

Elliot, S. N., Kratochwill, T. R., McKeivitt, B., Schulte, A. G., Marquart, A., & Mroch, A. (1999). *Experimental analysis of the effects of testing accommodations on the scores of students with and without disabilities: Mid-project results*. Paper presented at the CCSSO Large-Scale Assessment Conference, Snowbird, UT.

Green, D. R. (1998). Consequential aspects of the validity of achievement tests: A publisher's point of view. *Educational Measurement: Issues and Practices*, 17(2), 16–19, 34.

Hafner, A. L. (2001, April). *Evaluating the impact of test accommodations on test scores of LEP students and non-LEP students*. Paper presented at the annual meeting of the American Educational Research Association, Seattle, WA.

Helwig, R., Rozek-Tedesco, M. A., Tindal, G., Heath, B., & Almond, P. (1999). Reading as an access to mathematics problem solving on multiple-choice tests for sixth grade students. *The Journal of Educational Research*, 93(2), 113-125.

Individuals with Disabilities Education Act Amendments of 2004, Pub. L. No. 108–446, 111 Stat. 2647 (codified as amended at 20 U.S.C. § 1400 *et seq.*).

Johnstone, C. J. (2003). *Improving validity of large-scale tests: Universal design and student performance* (Technical Report 37). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes. Retrieved February 1, 2007, from <http://education.umn.edu/NCEO/OnlinePubs/Technical37.htm>

Kopriva, R., Samuelson, K., Wiley, D., & Winter, P. (2003, April). *Evidentiary logic in the assessment of diverse learners*. Paper presented at the annual conference of the National Council on Measurement in Education, Chicago, IL.

Lewis, D. M., Mitzel, H. C., & Green, D. R. (1996, June). *Standard setting: A bookmark approach*. Paper presented at the Council of Chief State School Officers Large-Scale Assessment Conference, Colorado Springs, CO.

Livingston, S. A., & Zeikey, M. J. (1982). *Passing scores: A manual for setting standards of performance on educational and occupational tests*. Princeton, NJ: Educational Testing Service.

- Messick, S. (1993). Validity. In R. L. Linn (Ed.), *Educational measurement, 3rd edition* (pp. 13–103). New York: Macmillan.
- National Research Council. (2001). *Knowing what students know: The science and design of educational assessment*. Washington, DC: National Academy Press.
- Rabinowitz, S. & Sato, E. (2005). *The technical adequacy of assessments for alternate student populations*. San Francisco: WestEd.
- Rabinowitz, S. & Sato, E. (2006). *The technical adequacy of assessments for alternate student populations: Guidelines for consumers and developers*. San Francisco: WestEd.
- Sato, E., Lagunoff, R., Worth, P., Bailey, A. L., & Butler, F. A. (2005). *ELD standards linkage and test alignment under Title III: A pilot study of the CELDT and the California ELD and content standards*. (Final Report to the California Department of Education). San Francisco: WestEd.
- Sato, E. & Rabinowitz, S. (forthcoming). *Assessing students in the gap: Critical issues and considerations impacting the validity of assessments based on modified academic achievement standards* [working title]. San Francisco: WestEd.
- Sato, E., Rabinowitz, S., & Gallagher, C. (forthcoming). *Access and special student populations—The similarities/differences in the needs of English language learners and students with disabilities: Implications for standards, assessment, and instruction* [working title]. San Francisco: WestEd.
- Sireci, S. G., Li, S., & Scarpati, S. (2003). *The effects of test accommodations on test performance: A review of the literature* (Center for Educational Assessment Research Report No. 485). Amherst, MA: School of Education, University of Massachusetts.
- Thompson, S. J., Blount, A., & Thurlow, M. (2002). *A summary of research on the effects of test accommodations: 1999 through 2001* (NCEO Technical Report No. 34). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.
- Thompson, S. J., Johnstone, C. J., Thurlow, M. L., & Altman, J. R. (2005). *State special education outcomes: Steps forward in a decade of change*. Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes. Retrieved April 13, 2007, from <http://education.umn.edu/NCEO/OnlinePubs/2005StateReport.htm>
- Thurlow, M. L., & Wiener, D. (2000). *Non-approved accommodations: Recommendations for use and reporting* (Policy Directions No. 11). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes. Retrieved February 1, 2007, from <http://education.umn.edu/NCEO/OnlinePubs/Policy11.htm>
- Tindal, G. & Fuchs, L. (2000). *A summary of research on test changes: An empirical basis for defining accommodations*. Lexington, KY: Mid-South Regional Resource Center.

Title I—Improving the Academic Achievement of the Disadvantaged; Individuals with Disabilities Education Act (IDEA); Final Rule, 72 Fed. Reg. 17,748 (April 9, 2007) (to be codified at 34 C.F.R. pt. 200 and 300).

Towles-Reeves, E., Muhomba, M., & Kleinert, H. (2006). *Alternate assessment: Have we learned anything new?* (Draft paper). Lexington, KY: University of Kentucky, used with permission.

U.S. Department of Education. (2006). *Peer review decision letters*. Retrieved December 28, 2006, from <http://www.ed.gov/admins/lead/account/nclbfinalassess/index.html>

U.S. Department of Education. (2007). *Measuring the achievement of students with disabilities*. Washington, DC: Author.

U.S. Department of Education. (2007). *Modified academic achievement standards: Non-regulatory guidance*. Washington, DC: Author.

Webb, N. (1999). *Alignment of science and mathematics standards and assessments in four states* (Research Monograph No.18). Madison, WI: Wisconsin Center for Education Research.

Webb, N. L., Horton, M., & O'Neal, S. (2002, April). *An analysis of the alignment between language arts standards and assessment in four states*. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.

WestEd. (2004). *Alignment study: Louisiana content standards and the Louisiana alternate assessment (LAA)*. San Francisco: WestEd.

Yovanoff, P., & Tindal, G. (2007). Scaling early reading alternate assessments with statewide measures. *Exceptional Children*, 73(2), 184–201.