A Technical Design and Documentation Workbook for Assessments Based on Modified Achievement Standards

Scott Marion

Center for Assessment

National Center on Educational Outcomes

University of Minnesota

S350 Elliott Hall, 75 East River Road

Minneapolis, MN 55113

This document was supported by the National Center on Educational Outcomes (NCEO) through its cooperative agreement with the U.S. Department of Education (Cooperative Agreement No. H326G050007). However, the opinions expressed herein do not necessarily reflect the position of the U.S. Department of Education and no official endorsement by the Department should be inferred.

A Technical Design and Documentation Workbook for Assessments Based on Modified Achievement Standards

Scott Marion

Center for Assessment

Working Draft: July 26, 2007

Preface

This document represents the current thinking of the Center for Assessment and our colleagues at this point in time. The suggestions and recommendations contained in this workbook have **not** been endorsed by the United States Department of Education. While we intend for these recommendations to help states develop technically adequate alternate assessments based on modified achievement standards, readers should not infer that these suggestions will guarantee a successful USED peer review outcome. This workbook represents first stage of technical assistance and there will be additional forthcoming more fully developed materials to assist states with the design, implementation, and evaluation of alternate assessments on modified achievement standards throughout the 2007-2008 school year.

Introduction

The U.S. Department of Education (USED) recently promulgated regulations (April 9 2007) to allow states to assess a limited number of special education students using an alternate assessment based on modified achievement standards (AA-MAS). Researchers, technical assistance providers, and state departments of education have recently begun to develop a good understanding of approaches for evaluating the technical adequacy of alternate assessments based on alternate achievement standards (AA-AAS), those assessments geared for the most significantly cognitively disabled students. A great deal of this progress can be attributed to applying the approaches for assessment design and evaluation outlined in *Knowing What Students Know: The Science and Design of Educational Assessment* (Pellegrino, Chudowsky, & Glaser, 2001) to AA-MAS (Marion & Pellegrino, 2006). Alternate assessments based on modified achievement standards pose many of the same technical challenges as AA-AAS, therefore we propose using the frameworks put forth in Pellegrino, et al. (2001) and Marion and Pellegrino (2006) to organize the technical guidelines and suggestions for the design of AA-MAS.

Validity is Central

As was argued for the design and evaluation of AA-AAS, validity needs to be the central focus of any design decisions for the development of the AA-MAS. Actually, we would take the same position for the design of any assessment or accountability system, but we focus on the AA-MAS here. Validity should be considered even more prominently in the case of the AA-MAS, because one of the stated reasons for allowing this new assessment system is that the students to be assessed with the AA-MAS have not been validly assessed with the regular grade-level assessment. Given that rationale, it makes sense

that any assessment designed for students designated to participate in the AA-MAS lead to valid inferences regarding their knowledge and skills, and for the evaluation of this assessment to be focused on such validity questions. It is beyond the scope of this brief to outline a full validity argument, but we urge the reader to see Marion and Pellegrino (2006) to see how this framework applies to the alternate assessment for the students with the most significant cognitive disabilities. It will also be important to attend to some of the primary sources of validity thinking to best determine how those apply to the AA-MAS. We think that Kane (2006) and Shepard (2003) would be especially useful.

The Assessment Triangle

We rely on *Knowing What Students Know* (KWSK) to frame our design and evaluation considerations for the AA-MAS. *Knowing What Students Know* was built from Mislevy's (1996) notion of assessment as a "process of reasoning from evidence" (p. 39) and previous National Research Council (NRC) work synthesizing research on human learning (Bransford, Brown, and Cocking, 2000). The authors of KWSK used the heuristic of an "assessment triangle" to illustrate the relationship among learning models, assessment methods, and inferences from assessment scores.

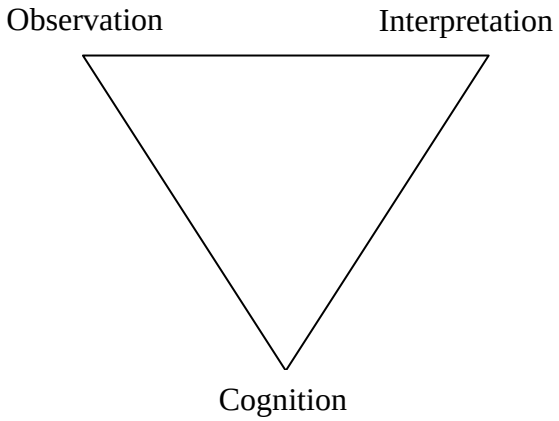

Figure 1. The Assessment Triangle (from Pellegrino et al., 2001, p. 39)

Cognition refers to the empirically-based theories and beliefs about how humans represent information and develop competence in a particular academic domain (Pellegrino et al., 2001). The theories of "learning and knowing" that help explain varying levels of performance in a particular domain are crucial for the design and interpretation of assessments. The observation vertex of the triangle refers to "a set of specifications for assessment tasks that will elicit illuminating responses from students" (Pellegrino et al., 2001 p. 42). The design of items or tasks is based upon the belief that those particular assessment events will allow students to demonstrate their understanding in a given domain, based upon a particular view of learning and knowing. The interpretation component in this diagram includes all of the methods and analytic tools used to make sense of and reason from the assessment observations (Pellegrino et al., 2001). Large-scale assessments rely on statistical models to inform these interpretations and models are designed to describe "patterns one would expect to see in the data given varying levels of student competency" (p. 43).

The Assessment Triangle Applied to Assessments Based on Modified Achievement Standards

We suggest using the assessment triangle heuristic to help us organize the technical considerations involved in designing and evaluating AA-MAS. There are substantial challenges inherent in accurately characterizing the cognition vertex for this group of students. The initial challenge for each state will be describing the learning characteristics of the students eligible to participate in the AA-MAS. This is a necessary first step in designing appropriate assessment experiences for those students[1]. We emphasize the importance of this critical first step for creating a valid assessment design. In other words, a valid assessment must be based on a clear understanding of which students will participate and how those students come to develop competence in the specific domain.

Organization of the Document

After introductory sections about validity, purposes and uses, this workbook is organized according to the vertices of the assessment triangle—cognition, assessment, and interpretations—and is intended to help states work through the design and documentation considerations associated with the implementation of an alternative assessment based on modified achievement standards. We close with a short section regarding some practical considerations about implementing an AA-MAS.

[1] We know that others are working on documents designed to help state leaders in identifying the group of students eligible to participate in the AA-MAS.

Validity

All of the design and evaluation aspects discussed in this document—and organized according to the assessment triangle—are intended to contribute to creating a more valid AA-MAS. We highlight some of the key validity issues here.

The state should be prepared to build a defensible validity argument in support of this assessment and in particular should be prepared to address validity issues related to potential unintended negative consequences as a result of implementing this assessment (e.g., lower expectations).

Rationale for this Information

Validity is clearly the most important technical criterion for the design and evaluation of assessment systems. The AA-MAS should be designed and implemented to foster more valid score interpretations for students in the "2% population" than has been the case with the regular assessment. In particular, the validity evaluation should focus on the connection among the characteristics of the tested students, the knowledge and skills assessed, and the design of the assessment.

Guiding Questions

1. What is the state's plan for evaluating the validity of the score inferences for students and schools of the AA-MAS?
    a. What is the state's plan for systematically evaluating the validity of this assessment for both the short-term (e.g., 1-3 years) and long-term (4-10 years)?
    b. How is this planned approach different from or the same as validation efforts for the regular assessment and AA-AAS?
2. How does this assessment provide a more accurate measure of the knowledge and skills of the participants compared with the regular assessment?
3. How will the inclusion of the AA-MAS as part of the state's assessment system lead to better instructional and curricular opportunities for these students? Whatever benefit (e.g., positive consequential evidence) for these students should be weighed against potential benefits from other approaches such as more appropriate accommodations and especially targeted instructional interventions.
4. What are the state's highest priority validity questions related to the AA-MAS?
    a. How does the state propose to evaluate the construct validity of the modified assessment?
    b. How does the state propose answering/addressing these high priority validity questions?
5. The state should also propose studies of response processes (AERA, APA, NCME, 1999) to document that the assessment approaches used with the AA-MAS are a better measure of students' knowledge of the grade level standards than would be the case with the regular assessment.
6. How will participation in the AA-MAS lead to improvements in the instruction for these students' in order to improve their likelihood of attaining grade-level proficiency over time?

Purposes and Uses

Purposes of the Assessment System

The state should describe the purposes for developing its AA-MAS. For example, NCLB accountability and IDEA 1997 & 2004 are often key reasons for developing statewide assessment systems that include AA-AAS. Certainly the newly promulgated regulations regarding the AA-MAS support several of the purposes for which these assessments may be used. In addition, there are almost always governing statutes and regulations at the state level and the state may often articulate other purposes of the system (e.g., instructional change).

Rationale for this Information
Validity can only be evaluated in the context of the purposes of the assessment(s) and how the results are used (next section). State leaders should be clear that if a purpose is specified in this section, evidence should be collected to evaluate the validity of the assessment related to this purpose.

Data Sources:
Enabling legislation, design documents, state board minutes, minutes from constituent group meetings (if applicable), and RFP documents are all potential sources of information to document the purposes of the assessment system and the purpose of the AA-MAS within that system. For initial design work, meetings of key stakeholder groups, especially those most familiar with the learning characteristics of the students likely to participate in the AA-MAS, should be convened so the purposes of the developing system can be developed explicitly with the involvement of such key groups.

Guiding Questions:
1. Given all the potential purposes, what are the primary purposes for implementing an AA-MAS?
2. What are the governing statutes providing the legal authority for the AA-MAS?
3. What do these legal documents require in terms of purposes for this test?
4. How are the purposes of the AA-MAS consistent with the purposes of the entire system?
5. How has the state ensured that its AA-MAS will provide coherent information for students across grades and subjects? (Peer Review Guidance p. 3).

Notes

Uses of the Assessment Information

The intended uses of the inferences drawn from the assessment results should be described for individual students, schools, and any other levels for which the results will be used.

Rationale for this Information

Specifying the intended uses of the assessment results is critical for building the validity argument. We validate assessments primarily for the ways in which the results are used and each use needs to have validity evidence to support it.

Data Sources:

Enabling legislation, design documents, state board minutes, minutes from constituent group meetings (if applicable), and RFP documents are all potential sources of information to document how the results of the assessment system are to be used. Additionally, score reports, interpretative documents, professional development workshops can all provide data to describe the uses of the assessment results.

Guiding Questions:

1. Does the state offer guidance to local educators, parents, and other stakeholders about the intended use(s) of the assessment scores?
2. Are there specific requirements for how the scores are to be used?
3. How are the data derived from the assessment system being used (e.g., accountability, program evaluation, instructional feedback)?
4. Has the state provided guidance to prevent inappropriate uses of results?

Notes

Cognition

As discussed above, the cognition vertex of the assessment triangle includes information on the nature of the students involved in the assessment, how they come to develop competent understanding of the knowledge and skills in the content domain and the specific content on which they will be assessed.

Who are the Students?

The state must describe, as completely as possible, the students participating in the AA-MAS. This is crucial for building the validity argument framed around the assessment triangle. The state should present the numbers of students participating in the AA-MAS both according to the nature of their specific disabilities and their relevant learning characteristics. More important than the quantitative information is the information about how these students learn, how they are taught, and what specific features of their learning makes them unlikely to reach proficiency on the regular assessment within the near future.

Rationale for this Information

In order to build a validity argument, we need to have a good understanding of who is participating in the AA-MAS. This is not meant to limit who participates, but simply to gain an accurate understanding of the participants (at least as much as possible).

Data Sources:

- Results of empirical studies and/or other approaches used by the state to describe the students eligible to participate in the AA-MAS.
- State and federal special education data bases indicating the counts of students participating in the AA-MAS by disability code and any other pertinent information may be used.
- Results from demographic data other than disability label that describe characteristics of assessment population are critical sources of information.
- IEP reviews may be a good source of information to gain a better understanding of the learning characteristics of students participating in the AA-MAS.
- To the extent possible, interviews with teachers and parents of students being considered for eligibility should be conducted—at least early in the program—to develop a good understanding regarding the nature of the learning characteristics (and instructional programs) of these students.

Guiding Questions:

1. How many students by specific disability category participate in the AA-MAS?
2. What are the characteristics of the learners that differentiate them from students in the regular assessment?
3. How congruent is the description of the intended population to the actual assessed population?
4. How do these students currently perform on existing academic measures?

5. What is the state's theoretical rationale for selecting this group of students to constitute the 2% (e.g., those just above the 1% compared with a group of students somewhat higher in the distribution)?
6. How will IEP teams identify these students?
   a. What "multiple valid and objective measures" will be used to identify the students?
   b. How will the information from these multiple measures be combined/weighed in order to contribute to the identification process?
7. Will the AA-MAS be offered for all grades and content areas currently tested?
8. How will states ensure that the identification procedures identify approximately 2% of the students consistently over time within each content area?
9. How are the students participating in the AA-MAS expected to acquire competence in the respective domains?

Notes

What is the Content?

The state must describe, as completely as possible, the content expectations for students participating in the AA-MAS. States will need to thoroughly describe the content and performance expectations for students participating in the AA-MAS to help define the domain for instruction and assessment.

Rationale for this Information

While the AA-MAS is required to be based on the same grade level content standards as the regular assessment, states are permitted to adjust the difficulty of the achievement standard by developing an assessment that is less difficult (but appropriately challenging for the identified group of students) than the regular test. The content and achievement domain must be defined for both instruction and assessment. Instruction is certainly a state and not federal matter, but it is in the state's and the student's best interest to focus first on the instructional needs of eligible students. Aspects of the validity argument (e.g., content validity, alignment) cannot be evaluated without these definitions (Peer Review Guidance, p. 4).

Data Sources:

State content and achievement standards (both for the regular and modified assessment), documentation of the processes used to create such standards, and research supporting the design of the standards would all be data sources for this chapter. Further, test blueprints for both the regular and modified assessments would be important information to help define the domain.

Guiding Questions:

1. How is the content of the modified test related to the content on the regular grade-level assessment?
    a. The regulations require the "same grade-level content as the regular assessment. How does the blueprint (in terms of content and cognitive processes) differ, if at all, from the blueprint for the regular assessment?
    b. What is the rationale to support the use of a different blueprint, if applicable?
2. What research was used to support the modification of test blueprint?
3. What guidance will the state provide to support the development of IEPs that are aligned with grade-level content standards for students participating in the AA-MAS?

Notes

Achievement (Performance) Level Descriptors

States may choose to describe the achievement level descriptors in the context of their standard setting discussions, but we argue that because of their critical connection to the definition of an AA-MAS this discussion is a fundamental aspect of the cognition vertex.

Rationale for this Information

The "2% regulations" explicitly state that while the content domain for the AA-MAS is required to be based on the same grade level content standards, it is the achievement standards that may be modified in order to provide a more valid assessment for students in this population. The particular methodology used to derive cutscores on whatever assessment is eventually chosen is far less important than the description of the knowledge and skills associated with each achievement level. The achievement level descriptor is the vehicle for linking the required content with the expected level of performance on the test. These descriptions are crucial for every standards-based assessment, but even more so when we are trying to clarify expectations for a newly-designed assessment system.

Data Sources:

State content and modified achievement standards, documentation of the processes used to create such standards, the qualifications of the key personnel involved in writing the descriptors, and research supporting the design of the standards would all be data sources for this chapter.

Guiding Questions:

1. How will the performance descriptors for the modified achievement standards be developed?
    a. How will they differ from and be related to the performance level descriptors for the regular assessment?
    b. How will states' current "level/rigor" of achievement standards interact with describing performance levels for the AA-MAS? In other words, how will states with large percentages of students already scoring at the proficient level or above on the regular grade-level assessment find "room" for describing performance for the AA-MAS and still have it be considered "grade level"?
2. What research or other evidence was (or will be) used to support the modification (compared to the regular assessment) of certain achievement expectations?
3. Do the modified achievement level descriptors lead to challenging expectations for the population?

Notes

Observation

A state is responsible for multiple decisions regarding the design, implementation, scoring, and reporting of its AA-MAS. Many of these considerations are similar to those for a regular assessment and are therefore familiar to states. For that reason this document does not go into as much detail regarding the observation vertex of the assessment triangle as for the cognition or interpretation vertices.

Rationale for this Information

Once state assessment teams determine who and what they intend to assess, the next phase involves designing an assessment to fulfill those intentions. Clearly, the various specific design decisions will have implications for how well the assessment can meet the needs of the population the AA-MAS is intended to serve.

Data Sources:

The data/information used to inform the assessment design decisions should come from multiple sources including (but not limited to):

- Minutes from stakeholder group meetings, directions for teachers and/or content advisory groups
- Examples of items and descriptions of the types of expected student responses
- Description of "front-end" alignment procedures
- Results of field /pilot tests
- Results of bias review
- Results of review or consideration of Universal Design
- Theoretical, best-practices, and/or instructional rationale for the selected assessment design

Guiding Questions:

There are many questions that can be used to help guide the development of an assessment system. The following are those that we consider particularly important for the development of an AA-MAS.

1. Are there special considerations regarding item development for the AA-MAS compared to the regular assessment?
2. If the AA-MAS relies on the same item pool as the regular test, how has the state ensured that the existing regular assessment items have an adequate "floor" such that the state will be able to set three cutscores between the existing cutscores and the lower end of the scale measured by this item pool?
3. Will the administration conditions and approaches for the AA-MAS differ from the administration of the regular grade-level assessment? If so, how and why?
4. Will the AA-MAS use the same or different scoring rules and procedures than the regular assessment?
5. What are the unique logistical issues the state must consider in the design and implementation of an AA-MAS?

Notes

Interpretation

The interpretation vertex encompasses those aspects of the assessment enterprise that focus on how we make sense of the scores derived from the assessment events. Topics such as alignment, scaling and equating, standard setting, reliability and reporting typically fall under the interpretation vertex.

Alignment

Alignment, as a technical criterion, has received perhaps the most attention in standards-based education. Alignment among the various aspects of the system—e.g., content, curriculum, assessments, and instruction—is thought to be a requirement for the educational system to function as intended.

Rationale for this Content

This is a legal requirement, but most importantly it is a critical educational requirement to ensure that all students are instructed and assessed on grade-level content, whether their work is evaluated against regular or modified achievement standards.

Data Sources:

- Grade-level content standards
- AA-MAS and regular grade-level test blueprints
- Modified achievement performance level descriptors
- Item specifications or materials used to train item writers

Guiding Questions:

1. What is the relationship between the blueprints for the AA-MAS and the regular assessments? If there is a difference, what is the rationale for this difference?
2. How will the state evaluate the alignment of the AA-MAS with the grade-level content standards—what protocol(s) will be used and the rationale for this choice?
    a. How will the state adapt, if at all, the typically-used alignment criteria for such components as depth of knowledge and range of knowledge to meet the needs of the AA-MAS?

Notes

Scaling and Equating

The state should discuss the methods and results for ensuring comparability among scores from this assessment within and across years. We assume that most modified assessments will report scores on the same scale as the regular assessment, but that is not necessarily a requirement or expectation. The state should describe the choice of scale, the rationale for the choice, and methods for transforming the raw scores into scale scores.

Additionally, the state should describe how it intends to ensure the comparability of score inferences across different forms of the test. Typically, this involves linking or equating, but it could include other forms of establishing the comparability of score inferences (e.g., judgmental approaches).

Rationale for this Content

If a single assessment score is a sample of behavior from which we would like to generalize to a larger domain, then we need to have some way of ensuring the comparability of score inferences from tests that are supposed to be tapping the same domain. This could mean different forms of the test administered to different students whether in the same year or across years. Depending on how the AA-MAS is constructed, this could also involve ensuring the link between the AA-MAS and the regular assessment.

Data Sources:

- Test scores and distributions, scale choice, and transformational methods.
- Item-level information and complete score distributions from multiple forms of the assessment
- Descriptions of the tasks, responses, and judgments.

Guiding Questions:

1. Will the item calibration approaches and scaling procedures differ from the procedures used for the regular assessment? This is especially critical for approaches where the number of distractors is reduced or items are selected from a subset (less difficult) of the item pool.
2. How will the scores on the AA-MAS be linked across years and to the regular grade-level assessment if applicable? How will the validity of this linking be judged?

Notes

Standard Setting

The state needs to provide a description of the methodology used to set cut scores, the reason for choosing this method(s) and for not selecting other methods.

Rationale for this Content

This is one of the most visible decisions that a state will make and it is crucial to provide a clear rationale for and clear description of the specific methodology selected, especially considering that the achievement standards are the primary means of flexibility for the AA-MAS. Documentation of the standard setting process is required (Peer Review p. xx) including: the selection of judges, methodology employed, and final results.

Data Sources:

- Student scores and frequency distributions
- Literature reviews to support methodological choice
- Item difficulty and item mapping information (depending on method)
- Student work samples (depending on method)
- Materials used to train the panelists
- Impact data and final recommended cut scores

Guiding Questions:

1. What method(s) will the state use to establish cutscores to validly reflect the performance level descriptors?
    a. Are currently documented standard setting procedures sufficient or will new approaches/modifications be required?
    b. What should the relationship be between the scores/achievement level on the AA-MAS and the regular grade-level assessment? In other words, should there be some type of mapping/linking so that we can relate judgmentally or empirically the scores from the AA-MAS to the regular assessment?

Notes

Reliability

The state should discuss the methods and results of analyses designed to characterize the measurement (broadly speaking) and sampling errors associated with test scores.

Rationale for this Content

The level of the educational system for which the scores will be used will dictate the types of reliability/consistency analyses necessary. For example, if the student level results are limited to serving as another source of information about the student, then traditional reliability/measurement error analyses are much less important than school- or district-level decision consistency analyses. However, much of the rationale for designing an AA-MAS is to enable states to provide a more valid assessment for students in the "2% population", so it is highly likely that states will want to provide reliable student-level information.

Data Sources:

- Data files of student-by-item responses
- Results of inter-rater reliability studies (if applicable)
- Data documenting the variability of performance across occasions
- Business rules for school-level accountability calculations

Guiding Questions:

1. How will the state evaluate the measurement error associated with scores on the AA-MAS?
    a. How do issues of restriction of range of test scores and restricted population interact with the measurement error evaluation? Does the state plan to make any adjustments to its reliability calculations as result of these restrictions?
2. What decisions—and at what level of the educational system—are made as a result of the assessment scores?

Notes

Reporting

The reporting section descriptions should align with purposes and uses described earlier. We anticipate that reports of AA-MAS performance will be included as a component of the assessment system reports. The process for determining the appropriate way to report the results of the AA-MAS should be described including the extent to which stakeholders were involved in the process. The information that schools and parents receive should be provided in this section. The extent to which reports for various constituencies adhere to the **Standards for Educational and Psychological Testing** (AERA, APA, NCME, 1999) should be described here. Summary scores of students, schools, and other stakeholders should be described.

Rationale for this Information

Reporting is often the last aspect of the assessment system to which we attend, yet is the most important vehicle for communicating about the assessment system to the public. Communicating about the AA-MAS will be a challenge because it is not an assessment with which many people will be familiar. Therefore, considerable attention should be devoted to the design of the reports and associated interpretation materials in order to facilitate the most effective communication possible.

Data Sources:

- Stakeholder meeting reports
- Reporting procedures from RFP
- Score interpretation guides
- Student, school data
- Parent letters
- Sample reports at the student and school levels

Guiding Questions:

1. How will the scores on the AA-MAS be reported?
   a. Will the same reporting scale be used as is used for the regular grade-level assessment?
   b. Will the achievement levels have the same names as the regular assessment? If so, how will it be communicated that these students participated in the AA-MAS?
   c. Will the report format be the same or different from the regular grade level test?
2. What constituencies receive reports?
3. What is the critical information that should be shared?
4. What level of student data may be reported and at what levels may these data be reported?
5. Do the reports comply with the recommendations found in the **Standards for Educational and Psychological Testing** (AERA, APA, NCME, 1999)?
6. What information should be included to ensure understanding by a non-technical audience and guard against inappropriate interpretation of results?

Notes:

Practical Issues

Finally, we raise some practical considerations. We are aware that these financial and logistical concerns are not necessarily part of technical design conversations, but anyone familiar with large-scale assessments knows that without attending carefully to these practical issues, the best laid technical plans can get derailed quickly.

Few states are likely to have budgeted for the expenses of developing a separate "2%" test. As states attempt to fund such efforts, it may be tempting or even necessary to under-fund some of the technical steps outlined throughout this document. While states will continue to receive assessment development funds through NCLB, most approaches for developing and implementing an AA-MAS will require an infusion of state or other funds. For example, many of the approaches being used by states currently rely on using the same item pool as the regular assessment. This might appear to represent a considerable cost efficiency—and there can certainly be an item development cost savings—but other aspects such as equating, form development, report, and standard setting will have to be paid for as if another grade-level test was being added to the state assessment system.

Adding a separate 2% test will present some logistical challenges that must be considered. Managing an expanding testing program (especially with science becoming part of the mix) for most states does not come with a comparable increase in staff and other resources. We know that state departments of education are already stretched thin and adding another set of assessments, especially one with much of the technical uncertainty surrounding the AA-MAS can "break the camel's back. States will need to evaluate their capacity to design, develop, and implement the AA-MAS before heading off down that road. For example, states must still return data to meet existing AYP reporting and school improvement deadlines. Adding another set of tests for each grade can create significant challenges with meeting these ambitious deadlines. For state leaders who feel like they must try to develop an AA-MAS, there is no requirement that it be implemented for all content areas and grades as the regular assessment. Experimenting with the AA-MAS at a few select grade-content area combinations would be a sensible initial step.

Finally, as mentioned in the validity section, states must weigh whatever benefits accrued to students as a result of implementing an AA-MAS must be weighed against other strategies for improving the educational success of the students eligible for the AA-MAS. Implementing an additional large-scale assessment program will not lead to improvements in the real education opportunities and success of these "2% students" without systematic and targeted instructional interventions. States must ensure that these most important aspects of the educational system are in place prior to or at least along with implementing an AA-MAS.