Considerations for the Alternate Assessment based on Modified Achievement Standards (AA-MAS): Understanding the Eligible Population and Applying that Knowledge to their Instruction and Assessment

Executive Summary

Project Manager & Editor: Marianne Perie, Center for Assessment

Authors:

Jamal Abedi, University of California, Davis

Christopher Domaleski, Center for Assessment

Stephen Dunbar, University of Iowa

Meagan Karvonen, Western Carolina University

Scott Marion, Center for Assessment

Jim Pellegrino, University of Illinois, Chicago

Marianne Perie, Center for Assessment

David Pugalee, University of North Carolina, Charlotte

Rachel Quenemoen, National Center on Educational Outcomes

Robert Rickelman, University of North Carolina, Charlotte

Catherine Welch, University of Iowa

Reviewers:

Howard Everson, Fordham University

Claudia Flowers, University of North Carolina, Charlotte

Brian Gong, Center for Assessment

Suzanne Lane, University of Pittsburgh

Katherine Ryan, University of Illinois, Urbana-Champaign

Gerald Tindal, University of Oregon

Phoebe Winter, Pacific Metrics

Paper commissioned by the New York Comprehensive Center in Collaboration with the New York State Education Department

Completed July 16, 2009

EXECUTIVE SUMMARY

In April 2007, the U.S. Department of Education (USED) released new regulations for the *No Child Left Behind* Act*,* that allowed for the use of an alternate assessment based on modified achievement standards (AA-MAS). These regulations supplemented the most recent Elementary and Secondary Education Act legislation regarding the development of grade-level assessments and alternate assessments based on alternate achievement standards. States could use this new assessment for students with disabilities to count up to two percent of students as "proficient" for purposes of Adequate Yearly Progress (AYP). These regulations were in response to state concerns that there were students with disabilities who were not able to show proficiency on the general assessment and yet would not be assessed appropriately by the alternate assessment based on alternate achievement standards. It supplemented the option of developing alternate assessments based on grade-level achievement standards which only a few states used.

The regulations state that "there is a small group of students whose disability has precluded them from achieving grade-level proficiency and whose progress is such that they will not reach grade-level proficiency in the same time frame as other students" (34 C.F.R. Part 200). However, this statement has raised countless questions as state policymakers try to determine who this "small group" is within that larger group of students who are not eligible for the AA-AAS but who are not performing well on the grade-level assessment. An emphasis of the regulations and the non-regulatory guidance was that this assessment must be challenging for these students. The assessments are required to cover the same breadth and depth as the other grade-level assessments. Modified achievement standards were described as being challenging for eligible students although defining a less rigorous expectation of mastery of grade-level

academic content standards. They could not be linked to content from a lower grade level or exclude content standards that were assessed by the grade-level general assessment. Again, these conditions raised challenges for states to develop assessments that measured the same content, including the same depth of knowledge, but would be “less difficult” for students with disabilities.

In spring 2008, six states submitted documentation of their “modified” assessments for Peer Review to determine whether they could be used for purposes of AYP. None of the states received approval for their AA-MAS, but lessons learned from the review of their designs provided useful information for all states. In June 2008, USED released a paper written by Janet Filbin that described and summarized the issues raised during the Peer Review of the six state alternate assessments based on modified achievement standards. Filbin identified five areas that were challenging to states:

1. Identifying students eligible to take the AA-MAS
2. Providing guidelines for writing standards-based IEPs and then monitoring the implementation of those guidelines
3. Designing an assessment based on grade-level content standards that is of an appropriate difficulty and depth of knowledge for this population
4. Determining the relationship between the AA-MAS, the general assessment, and the alternate assessment based on alternate achievement standards (AA-AAS)

   Writing appropriate modified achievement level descriptors

In Fall 2008, the New York Comprehensive Center (NYCC) applied for and received supplemental funding from USED to collaborate with the New York State Education Department (NYSED) to study these issues further. Their proposal involved convening national research experts to provide guidance to NYSED regarding the feasibility of developing an alternate assessment based on modified achievement standards and advice on how to design standards and assessments for students with

disabilities who are part of the “2%” population. NYCC partnered with the National Center for the Improvement of Educational Assessment (Center for Assessment) to convene a panel of experts and write a white paper on this topic.

In January 2009, a group of 17 research experts were identified and brought together in New York City to discuss the issues surrounding the design and development of an AA-MAS. This report is a result of that and several other meetings. Nine of the experts authored chapters of this report and the remaining eight experts reviewed the chapters and provided support and feedback to the authors. Each author is a nationally recognized expert on the issue discussed in his/her chapter and the reviewers possess similar qualifications allowing for a full review both within and across chapters. The expert panel used the five issues identified through peer review as a starting point during the initial planning meeting, and also brought their own experiences and expertise to bear on each of the major issues.

A driving question for New York State (and other states) is whether the development of this assessment will yield useful information to guide instruction and be cost effective. More specifically, in times of budget cutbacks, how can the limited funding available be best allocated to support the learning of these students? The first issue in answering this question is determining who “these students” are. Subsumed within that question is the possibility of expanding this report beyond the current federal regulations, focusing on students who may not be receiving grade-level content. In addition, it is important to consider the challenges of using the data to “guide instruction” when the primary focus of many of these assessments is simply to provide an additional measure for purposes of accountability.

NYSED indicated a desire to follow the regulations of assessing the same content on the AA-MAS as on their general assessment, to better understand how to make the assessment less rigorous, and to learn how to modify the achievement

standards while maintaining the reliability and validity of the results. Specific questions raised by NYSED include:

1. Which students are best served by this assessment?
2. How different are they from the rest of the special education population?
3. What is an “appropriately challenging” achievement standard?
4. Which modifications make the most sense in the context of the AA-MAS?
5. How do the modifications affect the validity and reliability of the score interpretations?
6. What is the credential (e.g., diploma) that is most appropriate for students participating in the AA-MAS and what are the implications in terms of post-secondary potential?

Most of the issues raised by NYSED are general issues that many state policymakers are confronting, and many of these match closely with the issues raised by Filbin (2008). The one exception is the last question regarding student credentialing. Because the Regents Examinations are used both for AYP purposes (thus open to modification) and graduation requirements,

Summary of the Report and Recommendations

The full report provides more detail on the regulations and describes the primary challenges in developing an AA-MAS based on the Filbin (2008) paper, existing research, and the panel’s own experiences. It provides a research-based analysis of the design and development issues and focuses on the theory, wherever possible, behind each issue. In addition, this report explores the existing research and best practices in identifying and assessing students who could be eligible to participate in an AA-MAS. Specifically, the goal of this report was to make recommendations to NYSED about developing an AA-MAS, with the expectation that these recommendations could be

generalized to other states. The authors each approached their chapters with an intention to help states think through the issues, make appropriate decisions regarding the allocation of resources, and ultimately improve opportunities for students with disabilities. Recognizing that people disagree on the assumptions behind this 2% population, this paper is written from the assumption that all students can learn (and should be taught) grade-level content standards with appropriate instruction and support. However the degree to which all students achieve the grade level content standards may vary.

Upon the completion of the second drafts of these chapters, the expert panel recognized the utility of the information beyond the application of fulfilling the federal regulations regarding an alternate assessment for purposes of accountability. Much of the discussion in this report relates to instruction and assessment of all low achievers and specifically low achievers with disabilities. Therefore, even if the regulations were rescinded, the panel believes the information provided in this report will continue to be applicable as the field works to improve our knowledge and understanding of how low achieving students with disabilities—and perhaps those without disabilities who are also struggling with grade-level achievement standards—learn, organize information, and communicate their understanding.

Summary of Recommendations

Although each chapter focused on a specific issue or theme, as described in the next section, some overarching recommendations flow throughout the report. First, there is a clear link between assessment, instruction, and student cognition, as first described in the 2001 NRC report *"Knowing What Students Know*." All of the authors recognize that the assessment must be designed to reflect what is known about the cognition of low achievers. Furthermore, there was a general acknowledgement that what is broken in instruction cannot be fixed through assessment. Therefore several of the chapters

focus on the importance of bringing lessons learned in studying the students and designing the assessments into the classroom. Similar to accommodations, certain modifications will only be successful to the degree they are incorporated into a student's daily instruction.

Another common thread throughout the report is the importance of developing a validity argument for this assessment early in the process, and testing the various assumptions throughout. The first section highlights the importance of determining who the students are and why they might need to be assessed using a separate instrument. Testing the assumption that a new assessment is needed is one of the first important recommendations. Then, within each chapter, there is a discussion of how to collect and evaluate validity evidence to determine that the development is in line with the expected goals.

Finally, the paper includes suggestions on how to incorporate the recommendations into an existing assessment and accountability system, including how to work with current state content standards and grade-level achievement standards.

Overview of Chapters

The direction to the expert panel from the Assistant Commissioner of the Office of Standards, Assessments, and Reporting in NYSED was to provide information on current research and best practices and make recommendations on the steps NYSED should take towards designing an AA-MAS (or to recommend not to do it at all).

Later, the specific questions from NYSED were added and divided among the different chapters as appropriate. Ultimately, though, this report was organized into three sections focusing on different aspects of designing and developing the AA-MAS, with three chapters in each section. Within the ten chapters (including an introduction as chapter 1), all of the issues described by Filbin (2008) and the questions raised by NYSED were addressed.

Section I. Identifying and Understanding the Population. The first challenge any state faces is to determine who the students are who are in need of a new assessment. The focus is on students with disabilities who are not achieving proficiency on grade level standards and who do not appear to be making significant progress towards achieving that proficiency. But beyond that, it is important to explore various characteristics of these students, including the nature of their disability and why it might hinder learning. And, we would have been remiss not to explore the issues of curriculum and instruction to see whether opportunity to learn is having a larger impact on performance than the nature of the disability. During the initial expert panel meeting, the experts decided that the issues of identifying the students were wrapped up in the NRC assessment triangle of assessment, instruction, and cognition (Pellegrino, Chudowsky, & Glaser, 2001). Thus it was decided that this first section should discuss the issues of identifying the students and understanding their cognitive abilities, including the interaction between instruction, cognition, and assessment.

The three chapters in this first section describe the population, the necessary elements of their instruction, and how they might be expected to develop competence in the domains. Each chapter provides, to the extent possible, the theoretical underpinnings of their positions.

Chapter 2, written by Rachel Quenemoen, focuses on the notion of the least dangerous assumption when considering exclusionary criteria. She provides a history of regulations regarding students with disabilities and discusses applications of the current regulations to the school environment. She next describes the policy context and summarizes research related to the teaching and learning of students with disabilities. Most importantly, she lays out a framework for state policymakers in considering how to identify students who might benefit most from an alternate assessment based on modified achievement standards. Included in this framework is a discussion on

improving student access to grade-level curriculum and providing more opportunities to learn.

Chapter 3, by Meagan Karvonen, takes this argument one step further by examining various instructional strategies for teaching students with disabilities. She focuses on procedural integrity by providing an overview of standards-based individualized education programs (IEPs) and describing how to promote improved opportunities to learn the standards-based curriculum with specialized instruction, services, and supports based on individual student learning characteristics. She also discusses the importance of aligning the curriculum with the grade-level content standards and providing supports for students to access this curriculum within the IEPs. She describes ways to promote quality of instruction and provide guidance to IEP teams.

Then, Chapter 4, by Jim Pellegrino--the last chapter of this section—helps us understand explanatory constructs by discussing broader conceptions of student cognition, describing possible sources of differences among students that have implications for learning, instruction, and assessment. He follows the previous discussions on learning and instruction and provides information on the third vertex of the triangle, student cognition. A strong focus of this chapter is on exploring issues related to barriers to learning for low achievers. He discusses the importance of understanding student learning characteristics and cognitive processes in assessment and goes on to describe possible sources of differences among students that have implications for learning, instruction, and assessment.

Section II. Test Development. This section moves the discussion from one about the students—who they are, how they learn, and how they should be instructed—to the assessment itself. Once we have a grasp on which students might be best served by an alternate assessment based on modified achievement standards, we need to determine how to take our understanding of the students and apply it to good test design. The

main question the authors wrestled with was how to make the assessment more accessible for students with a wide-range of disabilities but maintain the reliability and validity of the results. A deep understanding of the content and test design was necessary as well as an understanding of what is meant by modified achievement standards. Of critical importance is to understand how to cover the same breadth and depth as a general assessment and yet make it less difficult. These modified achievement standards can be less rigorous, but what does that truly mean?

Thus, chapter 5, written by David Pugalee and Bob Rickelman, begins this section with a discussion of the content domains of reading and mathematics. They continue the discussion from the first section regarding aligning curriculum, instruction, and assessment but focus specifically on issues related to reading and mathematics. They focus on content standards and curriculum and describe how content standards are developed. Then they move to the key issue of how to maintain the same content, only modifying the achievement standards. They describe important issues regarding sampling the domain and making the content accessible to students with disabilities. The chapter ends with some suggestions on ways to enhance or revise items to provide scaffolding for students who may need additional supports in order to show what they know and can do. The authors point out that the scaffolds described only benefit students if they are incorporated both in instruction and assessment.

Chapter 6, by Cathy Welch and Steve Dunbar, picks up where Chapter 5 leaves off, focusing more specifically on issues of good test development. They discuss the challenge of developing items and test forms in reading and mathematics that better match the learning characteristics of the population identified for the AA-MAS, focusing on reducing the difficulty while maintaining the reliability of the assessment. Welch and Dunbar include many suggestions of types of modifications that can be made to the general assessment to make it more appropriate for low achieving students with

disabilities. They also provide an overview of best practices for item and test development and use these considerations to frame the discussion of areas for modification. The authors then address the psychometric consequences of test modifications as they play out in the assembly of test forms and in the analysis of technical characteristics of items and test forms.

Then, in Chapter 7 by Marianne Perie, the focus turns to the modified achievement standards. She focuses on the main components of achievement standards—numbers and names of levels, achievement level descriptors, and cut scores—and provides guidance on how to develop each component. She then discusses how the modified achievement standards might "fit" between the grade-level achievement standards and the alternate achievement standards, and provides practical advice for writing achievement level descriptors and setting cut scores, discussing the rationale behind each. The theories behind each approach to modifying the achievement level descriptors are linked back to both the test design from Chapter 6 and student cognition, discussed in Chapter 4.

Section III. Technical Considerations and Practical Applications. The third section of this report addresses three issues related to the technical quality and use of the assessments: examining the validity of these assessments, determining the comparability of these assessments to the general assessment, and understanding how these assessments will be operationalized and used in a state accountability system. Its original intent was to provide information on examining the technical adequacy of these assessments as a logical follower of the section on assessment design. However, it soon became clear that specific issues needed to be addressed, far beyond most technical considerations of item analysis, reliability and equating. Thus, a previous chapter (Welch & Dunbar, Chapter 6, this volume) began the discussion of technical adequacy, focusing on item analyses and psychometric characteristics of the test. This

section, then, focuses on very specific questions regarding the technical quality of this assessment, not as a standalone assessment, but as it fits into a larger assessment program.

In chapter 8, Jamal Abedi explores issues of comparability of the AA-MAS with the general assessments and grade-level achievement standards. Abedi writes from the perspective of ensuring students who take this assessment should have the same opportunities for success as students who take the general assessment. Issues concerning comparability of assessments are of paramount importance for inclusion, as states may not produce valid outcomes if the degree of comparability across the assessments has not be clearly explored and described. Likewise, descriptions of any differences in interpretations of achievement levels of the same name across each type of assessment need to be provided. The author examines several components of comparability, including content and construct, psychometrics, scale and score, linguistic structure, basic text features, and depth of knowledge. Finally, he discusses the appropriateness of the accommodations used for students with disabilities based on their IEP.

Chapter 9, by Scott Marion describes how to develop a validity argument for the implementation and use of the AA-MAS. He emphasizes the importance of stating the policymakers' values explicitly and articulating the theory of action particularly in light of the uncertain conceptual framework supporting this AA-MAS initiative. He then describes methods for gathering evidence from the design process through implementation. This evidence can then be used to evaluate the assumptions of the validity argument. He argues that a synthesis will provide information about how to improve the program as well as determine the value of AA-MAS in terms of the instructional and social benefits given the costs.

Finally, in Chapter 10 by Chris Domaleski, the focus turns to the practical application of these ideas in a state assessment system. This chapter focuses on how the AA-MAS fits into a pre-existing state assessment and accountability system. Domaleski provides specific advice on how to develop participation guidelines, evaluate the reliability and validity of accountability decisions made using these assessments, and how to operationalize the “2% cap”. He considers the current state context in reviewing operational considerations and discusses ways to produce rich and useful score reports and other data that could inform policy decisions, such as those related to diploma eligibility.

At the end of the full report are three appendices followed by a glossary of terms. Appendix A simply provides information on the team that developed this white paper as the chapters were shaped by the entire expert panel. Appendix B provides suggested resources, available on the internet, for effective curriculum and instruction. Appendix C (appended to this summary as well) is a tool that state policymakers can use as they are considering whether and how to develop an AA-MAS. This tool consists of questions for state policymakers and educators to consider at each phase of assessment development as well as a link back to resources within this report that will inform the discussions. Many of the questions come from the validity framework that guides much of the discussion in these chapters (c.f., Marion, 2007). Finally, a glossary of terms is included that encompasses vocabulary used in both the assessment and disabilities worlds.

References

Filbin, J. (2008). *Lessons from the initial peer review of alternate assessments based on modified achievement standards* .Washington DC: U.S. Department of Education, Office of Elementary and Secondary Education (OESE) Student Achievement and School Accountability Program.

Marion, S. (2007). *A technical design and documentation workbook for assessments based on modified achievement standards.* Minneapolis: National Center on Educational Outcomes. Retrieved April 2009 from http://cehd.umn.edu/nceo/Teleconferences/AAMASteleconferences/AAMASworkbook.pdf

Pellegrino, J. W., Chudowsky, N. J., & Glaser, R. (Eds.) (2001). *Knowing what students know: The science and design of educational assessment.* Washington, DC: National Academy of Sciences.

U. S. Department of Education (2007a). *Final Rule 34 CFR Parts 200 and 300: Title I—Improving the academic achievement of the disadvantaged; Individuals with disabilities education act (IDEA).* Federal Register. 72(67), Washington DC: Author. Retrieved July 12, 2008, from the World Wide Web: http://cehd.umn.edu/NCEO/2percentReg/Federal-RegApril9TwoPercent.pdf

U. S. Department of Education (2007b). *Modified achievement standards: Non-regulatory guidance.* Washington DC: Office of Elementary and Secondary Education (OESE). Retrieved from the World Wide Web on August 27, 2008, at http://www.ed.gov/policy/speced/guid/nclb/twopercent.doc

APPENDIX: TOOL FOR STATE POLICYMAKERS

This tool is simply a list of guiding questions for state policymakers considering the development of an AA-MAS. Beyond providing information to think about in deciding whether an AA-MAS fits well into a state's current system, it also provides guidance during the design and development process. Each question is linked back to a chapter for further information about topical considerations.

Topic	**Guiding Question(s)**	**Relevant Chapter(s)/Section(s)**
Appropriateness of Developing an AA-MAS	Are there students who cannot be appropriately assessed with the state's current large-scale assessment system?	Chapter 2, pages 23–39 Chapter 4, pages 103–105 Chapter 10, pages 345–347
	How do we know that the problem is the format or design of the general assessment rather than a lack of opportunity to learn the material on the assessment?	Chapter 2, pages 32–39 Chapter 3, pages 52, 60
	Will this state reap more benefits from developing a new assessment targeted towards eligible students rather than focusing on their instruction through another means, such as professional development of teachers?	Chapter 3, pages 78–79 Chapter 5, pages 163–167
	What is your theory of action for how this assessment will improve student outcomes?	Chapter 9, pages 317–323
Identifying the Target Population	How do we identify the students who are eligible to take the AA-MAS?	Chapter 2, pages 30–39 Chapter 10, pages 343–347
	What are the characteristics of these students? (*This question may need to be answered for several different groups of students.*)	Chapter 2, pages 23–39 Chapter 10, pages 345; 356; 358
	What are your assumptions about these students' ability to learn grade level content and to show what they know?	Chapter 4, pages 103–105; 125–134
	Are these students different from students without	Chapter 2, pages 23–39 Chapter 4, pages 103–105

	disabilities who have performed poorly on the large-scale assessment? If so, how?	
Appropriateness of current curriculum and Instruction	Do these students have standards-based IEPs that promote opportunity-to-learn the standards-based curriculum? How do you know?	Chapter 3, pages 57–59; 61–63; 69–74
	What evidence exists to support a determination that these students will not achieve grade level proficiency in the current year because of the effects of their disability and not because of lack of opportunity-to-learn?	Chapter 2, pages 32–39 Chapter 3, pages 57–63 Chapter 9, pages 325–326
	What evidence exists to support the policy assumptions that these students are provided high quality access to the standards-based curriculum, through specialized instruction, services, and support?	Chapter 3, pages 59–69 Chapter 5, page 163
	What types of training and support are available for teachers of these students to improve participation and performance in the standards-based curriculum?	Chapter 3, pages 78–79 Chapter 5, pages 164–167
	What training, oversight, and monitoring processes are built into the system to ensure that IEP teams make high quality decisions about who participates in AA-MAS?	Chapter 3, pages 78–81
Appropriateness of AA-MAS for improving student outcomes	What is the nature of the barriers to these students' participation on the general assessment?	Chapter 4, pages 103–105; 122–124; 135; 140 Chapter 5, pages 163–167 Chapter 6, pages 210–215
	How will this assessment provide a more accurate measure of the knowledge and skills of the participants compared with the general assessment?	Chapter 6, pages 224–232
	How will development of an AA-MAS yield more valid	Chapter 2, pages 27; 45–48

	inferences about the students than other assessment approaches, such as improved general assessment design, appropriate accommodations, or development of an AA-GLAS?	Chapter 3, pages 76–78; 82 Chapter 5, page 176 Chapter 6, pages 214–215 Chapter 8, pages 293–296; 300
	What are the relative costs and benefits of assessment development and implementation compared with other uses of resources, such as targeted staff development on instructional and curricular interventions for teachers of struggling learners?	Chapter 6, page 197 Chapter 9, page 309 Chapter 10, page 342
	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 participating students?	Chapter 3, pages 59–69
Modified achievement standards	How do the performance expectations of the AA-MAS relate to those in the general assessment and the AA-AAS? Is Proficient on the AA-MAS similar in nature to Proficient on the general assessment? Is it closer to Basic? Or is it somewhere in between?	Chapter 7, pages 240–251 Chapter 8, pages 272–273; 276
	Is there an expectation that the AA-MAS may provide a stepping stone for students to reach Proficient on the general assessment? Or, is the expectation that students taking the AA-MAS are a unique population that will always need the modifications provided? Is a student who scores Advanced on the AA-MAS prepared to take the general assessment or an AA-GLAS or are they simply exceeding the criterion on their own assessment?	Chapter 2, pages 30–39, 45–47 Chapter 7, pages 242–243
Test Design	How will you carry your philosophy regarding the description of the students and their barriers to participation in a general assessment to your design of	Chapter 5, pages 186–192 Chapter 6, pages 208–224 Chapter 7, pages 240–246

	the AA-MAS?	
	What type of assessment best fits your philosophy—a modification of your general assessment? AA-GLAS or a modification of the AA-GLAS?	Chapter 2, pages 45–48 Chapter 6, page 252
	If you choose to modify your general assessment, which types of modifications best match your philosophy regarding the students' barriers to participation in the general assessment?	Chapter 4, pages 140–142 Chapter 5, pages 186–192 Chapter 6, pages 208–224
	How do you intend to maintain the depth and breadth of the assessment but reduce the difficulty?	Chapter 4, page 135 Chapter 5, pages 161–162; 180–192 Chapter 6, pages 206–2224
	How will you measure and demonstrate the degree of comparability between the AA-MAS and the general assessment?	Chapter 8
Documenting the technical quality or validating the AA-MAS	What are the important features of technical quality and validity that should be evaluated and documented throughout this process?	Chapter 9
Incorporating an AA-MAS into an existing assessment and accountability system	In which subjects/grades should we develop an AA-MAS?	Chapter 10, pages 340–342
	How does the AA-MAS fit between the AA-AAS and the general assessment? Do we expect to see smooth transitions from one assessment to the next?	Chapter 2, pages 48–49 Chapter 3, pages 52–53, 64–65, 85–86 Chapter 7, pages 246–247 Chapter 10, page 346
	How will you report results in a manner that will provide maximum information to teachers and parents?	Chapter 4, page 138 Chapter 10, pages 359–360