

System Implications

Chapter 12

AA-MAS in Pennsylvania: Defining the Population; Tracking their Performance

Naomi Zigmond

Amanda Kloo

Christopher J. Lemons

University of Pittsburgh

Lynda Lupp

Bureau of Special Education

Pennsylvania Department of Education

This chapter discusses findings from the General Supervision Enhancement Grant H373X070027-09 awarded to the state of Pennsylvania by the U.S. Department of Education's Office of Special Education Programs for which the authors provide research support and consultation. The views and commentary expressed herein are solely those of the authors. No official support or endorsement by the U.S. Department of Education or the Pennsylvania Department of Education is intended or to be inferred.

Executive Summary

One of the largest challenges in developing effective accountability systems is determining how to best include students with disabilities. The purpose of this paper is to report on Pennsylvania's response to the 2007 U.S. Department of Education challenge to develop an Alternate Assessment based on Modified Academic Achievement Standards (AA-MAS) for a small group of students with disabilities. A close examination of the characteristics of the population tested on the AA-MAS across the first two years of its implementation highlights the practical challenges to implementing accountability policy related to the appropriate assessment of students with disabilities. We explore the intended and unintended consequences of the state's launch of the AA-MAS including its effect on IEP team decision-making and its impact on student achievement across three years. We close with a discussion of lessons learned from the development and implementation of the AA-MAS that should inform the current restructuring of accountability systems nationwide.

Introduction

For the last decade, the accountability provisions of the No Child Left Behind Act (NCLB) of 2001 have shaped school practice. Federally mandated annual reporting of aggregate student achievement data has placed enormous pressure on public school districts and individual schools. Adding to that pressure is the additional requirement for public reporting of scores for student groups disaggregated by race/ethnicity, economic status, English language proficiency, and disability status. The result is that the academic progress of students with disabilities has quickly become the target of public and political scrutiny. Why? Students with disabilities are historically low-achievers. They are designated eligible for an Individualized Education Plan and special education services *because* their disability adversely affects educational performance (IDEIA 2004, § 300.8 (c)). But the NCLB accountability policies embody a “no excuses” model of high achievement expectations. It requires that 100% of public school students demonstrate proficiency on grade level academic achievement standards by 2014. On the one hand, such high expectations have spurred an increase in educational opportunities for school-aged populations of students historically ignored or undervalued by the educational mainstream. On the other hand, it has also placed incredible pressure on special educators, students, and their families to achieve what many consider to be impossible.

Researchers and practitioners have argued that expecting *all* students (particularly those with cognitive/intellectual disabilities) to master age-appropriate grade-level academic standards is simply unrealistic. The federal response was an immediate concession to accommodate students with the most significant cognitive disabilities. For these students, states would be encouraged to develop alternate assessments based on alternate achievement standards (AA-AAS). This “1% option,” as it came to be known, appeared to be a satisfactory accountability provision for a small group of students whose competency on grade-level academic content could not be measured in the same way (or on the same test) as their same-age peers as a result of the severity of their cognitive impairment. Unfortunately, many more students with disabilities (who did not meet AA-AAS eligibility requirements) failed to achieve proficiency on the general assessment year after year. These special education students came to be known as *the kids in the gap*—students not sufficiently cognitively impaired to qualify for an AA-AAS, yet not learning enough to reach proficiency on the regular assessment even with accommodations (Bechard & Godin, 2007; Zigmond & Kloo, 2009).

After much discussion and debate, in December 2007, the U.S. Department of Education announced flexibility in statewide assessment participation that appeared to target the kids in the gap. The flexibility applied to students for whom the “grade level assessment is too difficult” but the “alternate assessment based on alternate achievement standards is too easy” (U.S. Department of Education, *Modified Academic Standards: Non-regulatory guidance*, 2007, p. 8). The proposed flexibility would permit states to develop an alternate assessment based on modified

achievement standards (AA-MAS), aligned with grade-level content but reflecting reduced breadth or depth of that grade-level content so that persistently low performing students with disabilities would be better able to demonstrate what they know and can do on a state test. Much like the AA-AAS, the provision did not limit how many students with IEPs could be assessed against modified achievement standards and left individual student test assignment decisions to the IEP team. But states could include proficient scores from such assessments in making AYP decisions only up to a cap of 2% of the total tested population.

Pennsylvania embarked on AA-MAS implementation (hereafter referred to as the *Pennsylvania System of School Assessment-Modified*, or the *PSSA-M*) with the aim of more accurately measuring the knowledge and skill of a small group of children with disabilities by minimizing the effects of “processing (e.g., cognitive, linguistic) or physical challenges related to the students’ disabilities without significant alteration of the assessed construct” (Pennsylvania Department of Education, 2010a, p. 55). Because Pennsylvania had taken advantage of the 2004 2% proxy option, the deadline for AA-MAS implementation loomed large¹. The PA Bureau of Accountability and Assessment authorized a vendor to begin *PSSA-M* development and design while the PA Bureau of Special Education launched a separate effort, funded by a General Supervision Enhancement Grant (GSEG) awarded by the U.S. Department of Education’s Office of Special Education Programs, focused on defining the AA-MAS target population, developing state-wide training on assignment of students to the new assessment, and implementing a comprehensive standards-aligned instruction and IEP planning system for students with disabilities. The GSEG team conducted various research activities to define the target population for the *PSSA-M*. (See Lemons, Kloo, & Zigmond, 2011, for a more detailed discussion of these activities and findings).

Survey

The first information gathering activity conducted by the GSEG team was a small-scale survey of special education teachers in schools identified as key research sites for the grant. The survey’s purpose was to describe opportunities provided, in general or special education settings, to students with IEPs in grades 5, 8, and 11 that would support learning of eligible grade level content (i.e., assessed by the regular grade level test [the Pennsylvania System of School Assessment, *PSSA*]). Survey questions were divided into four topic areas: (1) *teacher information* for which teachers were asked to provide data about their professional experience and professional development related to the AA-MAS and standards-aligned instruction for students with IEPs; (2) *student information* for which teachers were asked to provide academic achievement information about a target student or students whom they considered to be “persistently low performing” and a candidate suited for the AA-MAS based on the criteria detailed in the Federal Guidance; (3) *opportunity to learn data* for which teachers were asked to quantify the target

¹The federal proxy permitted states committed to implementing an AA-MAS by 2009 to count 2% of students with disabilities who failed the general state assessment as proficient in the accountability calculations of schools who failed to meet accountability requirements due to IEP subgroup performance.

students' opportunity to learn the eligible content of the state reading test; and (4) *IEP goals and instructional access* that spoke to the students' level of academic functioning, the alignment of their IEP goals, and nature of their individualized instruction overall.

Survey respondents overwhelmingly identified their lowest performing students with specific learning disabilities as the population of students in need of an alternate assessment that could provide a clearer indication of what the students know and can do. Overall, survey findings raised serious concerns about target students' opportunities to learn grade level academic content. However, the respondents also expressed great concern regarding how to make a "modified" test challenging and aligned to grade-level standards, and at the same time "doable" for a group of children who are significantly behind a majority of their grade-level peers. Further, the limited connection between grade-level standards and instruction provided to students identified by respondents as potential AA-MAS takers raised questions regarding the appropriate content for this assessment.

Focus Groups

To extend the information provided by the survey, a series of stakeholder focus groups were convened statewide. In all, 110 participants (including parents of students with disabilities, general and special educators, state-level personnel, content area teachers, administrators, school psychologists, curriculum specialists, teacher trainers, and related service personnel) reviewed the federal guidance about the AA-MAS and discussed issues related to identifying the target population, developing modified academic achievement standards, and anticipating the implications/potential impact of the AA-MAS on educational and assessment experiences of students with disabilities. Three major themes arose from these discussions. First, focus group participants agreed that the students most appropriate for the test are those who are "far below" grade level. Second, participants had a difficult time resolving the fact that while the AA-MAS may result in improved *testing* experiences for severely struggling students with disabilities, it may not result in improved *instructional* experiences for those students. These sentiments are echoed in the literature (see Elliott, Kettler, & Roach, 2008; Marion, 2007). Finally, participants were greatly confused by changes in the federal requirements related to the rigor of the AA-MAS. Revisions to the Guidance deleted the phrase "reduced breadth or depth" in describing the scope of the modified achievement standards and inserted, instead, the phrase "must be challenging for eligible students, but may be less difficult than grade-level academic achievement standards" (§ 200.1 (e)(1)(ii)). As a result, participants found the distinction quite perplexing.

Analysis of PSSA Performance Trends for Students in Special Education.

A third activity that the GSEG group conducted to assist in making recommendations about the AA-MAS was a trend analysis. Three consecutive years (2006, 2007, and 2008) of PSSA

performance data were gathered for four cohorts of students with IEPs (enrolled in grades 3, 4, 5, and 6 during the 2006-2007 school year). Data were gathered on children who had an IEP at any point during the three-year window and who also had at least two years of reported PSSA scores. The analysis was a simple examination of movement between proficiency levels across years. These data provided some evidence that if any children in special education are likely to move from one of the failing categories of performance into a passing one, it would most likely be children who scored in the basic level on the general assessment—a possible reason to exclude such a student from taking the AA-MAS.

As a result of these GSEG activities, researchers and the Pennsylvania Bureau of Special Education partners recommended to the Pennsylvania Bureau of Assessment and Accountability the very lowest performing students with IEPs should be the students eligible for the PSSA-M; that is, students who consistently scored in the lowest performance category (i.e., Below Basic) on the regular assessment but did not meet eligibility criteria for the AA-AAS. The Pennsylvania Bureau of Special Education developed and disseminated test assignment criteria and decision-making resources to guide IEP teams' efforts in recommending the most appropriate assessment option for each tested student in their charge (ASIST, Pennsylvania Department of Education, 2010b, see Figure 1). Consistent with federal guidance and the GSEG's research findings, IEP teams were instructed to consider as eligible for the PSSA-M: (a) students not eligible for the AA-AAS; (b) students whose IEP documented standards-aligned grade level instruction; (c) students with persistent academic difficulties despite intensive intervention efforts; and (d) students whose growth data across multiple measures indicated they were unlikely to attain grade-level proficiency. The guidelines indicated various types of data that could be used to support the decision (e.g., previous year's state test, norm referenced achievement tests, classroom assessment data, etc.).

In Spring 2010, the Pennsylvania Department of Education launched an operational PSSA-M Math assessment and in Spring 2011 the PSSA-M Reading and PSSA-M Science assessments. All three used the traditional paper and pencil format. Reading and Math covered the grade span 4-8 and 11; Science was targeted to grades 8 and 11. State leaders agreed that students with IEPs should be required to participate in the regular PSSA at least one time (in third grade for reading and math, in fourth grade for science) before being considered for the modified alternative.

Figure 1. ASIST Guidelines for IEP Team Decision-Making 2010

IEP Team Recommendations for Participation in PA State Assessments

Spring 2010 Administration: Decision Guidelines for Assignment to PSSA-M

- Students must meet 3 out of 4 criteria in order for an IEP team to recommend participation in the PSSA-Modified by subject area.
- If a student does not meet 4 out of 4 criteria by subject area, IEP teams may consider recommending participation in the PSSA with accommodations.
- Potential Evidence may include what is listed below but is not limited to those listed.


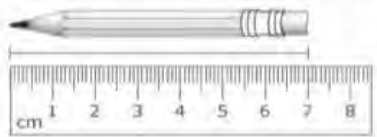
Criteria	Guideline	Potential Evidence	Decision-Math		Decision-Reading		Decision-Science	
			YES	NO	YES	NO	YES	NO
1. Ineligible for the PSSA	Students considered for the PSSA-M do not have significant cognitive disabilities and should not be held to alternate achievement standards.	<ul style="list-style-type: none"> • IEP • PASA Guidelines for Participation 			N/A	N/A	N/A	N/A
2. Standards-Aligned Grade-Level Instruction	All students considered for the PSSA-M must have a grade-level standards-aligned IEP that clearly documents that the student requires significant instructional accommodations and/or modifications to successfully access grade-level content.	<ul style="list-style-type: none"> • In applicable subject area: • Standards-aligned IEP Goals • SOI documentation • Present level documentation • Instructional accommodations and modification documentation/evidence 			N/A	N/A	N/A	N/A
3. Persistent academic difficulties	Students considered for the PSSA-M have persistent academic difficulties despite having received intensive research-based interventions.	<ul style="list-style-type: none"> • In applicable subject area: • Progress monitoring data (CBM, CBA) • Intervention documentation 			N/A	N/A	N/A	N/A
4. Lack of academic progress or growth	Students considered for the PSSA-M have established patterns of significantly low performance on multiple valid measures that indicates that even if significant growth occurs, achievement of grade level proficiency is unlikely.	<ul style="list-style-type: none"> • In applicable subject area: • State level assessment data: <ul style="list-style-type: none"> ○ Failure(s) of below basic performance on the PSSA ○ PVAAS projection to proficiency of less than 40% • District/School level assessment data: <ul style="list-style-type: none"> ○ Performance well below grade level on a norm-referenced achievement test • Grade level assessment data: <ul style="list-style-type: none"> ○ Well below average performance on progress monitoring or other classroom assessment data 			N/A	N/A	N/A	N/A

Following federal guidance, the PSSA-M was designed to assess the same grade level content knowledge as the general assessment (PSSA). However, students taking the PSSA-M would demonstrate proficiency by meeting a less rigorous standard of academic achievement. In other words, students would still be required to demonstrate mastery of grade-level skills, albeit at a less cognitively complex level. To meet this aim, test contractors modified original items from the general assessment by: (a) reducing the cognitive load (i.e., amount and complexity of information); (b) reducing the language load (i.e., construct-irrelevant language); or (c) supporting students’ processing of information (e.g., segmenting or chunking information, providing graphics that support understanding) (Pennsylvania Department of Education, 2010a). The changes were intended to address “access” needs and to increase test validity for eligible students.

Modifications focused on simplifying the language, graphics, and numbers used in question and answer choices. In some cases, extraneous information was removed. Additionally, items were reformatted by adding additional white space, increasing font size, or reordering items. Scaffolding was also provided with additional tables and keys (e.g., a timeline for organizing events), increased use of underlining and bolding of key information, and inserting helpful graphics or figures (Pennsylvania Department of Education, 2010c).

Figure 2 illustrates a modified 4th grade math item. The original item is positioned on the left, the modified item on the right. In this example, the context of the modified item was simplified and the question stem was presented in a more direct manner. A graphic representation of a ruler was added and all answer choices were clearly labeled in centimeters to eliminate a possible confusion with inches.

Figure 2. Sample Original (PSSA) and Modified (PSSA-M) Math Item

Original Item	Modified Item
<p>27. Cleo's pencil was very short.</p>  <p>Using your ruler, what is the length of her pencil, in centimeters?</p> <p>A 2 <input type="text" value="near '2 inches' mark"/></p> <p>B 3 <input type="text" value="approximately 3 inches"/></p> <p>C 7 *</p> <p>D 8 <input type="text" value="ruler end set at 1 cm"/></p>	<p>10 A pencil and a centimeter (cm) ruler are shown below.</p>  <p>What is the length of the pencil?</p> <p><input type="radio"/> 4 centimeters <input type="text" value="too short"/></p> <p><input type="radio"/> 5 centimeters <input type="text" value="too short"/></p> <p><input type="radio"/> 7 centimeters *</p> <p><input type="radio"/> 8 centimeters <input type="text" value="last number pictured on ruler"/></p>

Analysis

The GSEG collaborators set out to review the usefulness of their newly developed guidance documents by analyzing who took the Math PSSA-M in 2010 (and also in 2011 and 2012) and the performance outcomes for these students. In this report we summarize findings on the characteristics of the tested population and their achievement. We focus on the math assessment because it provides the largest amount of longitudinal data. We close with a discussion of lessons learned from math AA-MAS development and implementation in PA that we hope will inform the future restructuring of accountability systems nationwide.

2010 PSSA-M—Year One of Statewide AA-MAS Implementation in Math

Student Participation in 2010

In the spring of 2010, students in grades 4 through 8 and 11 (n=795,514) participated in the statewide accountability assessment in Pennsylvania. This included 132,790 (16.7%) students with disabilities. Not all of the students with disabilities took the general grade level PSSA assessment. At each grade, approximately 1.5% of students completed the alternate assessment for students with the most severe cognitive disabilities. And, from 1.7% to 2.6% of tested students (average 2.1%) took the modified math assessment, PSSA-M (see Tables 1 and 2). It appears that Pennsylvanians generally held fast to the federal estimates that 2% of the total tested popu-

lation would benefit from a test based on modified achievement standards. IEP teams across the state assigned nearly 10% of 4th grade students with disabilities to the M-assessment; over 17% of 11th grade students with disabilities were assigned to the test. As grade level increased, IEP teams assigned more students and larger proportions of students with IEPs to the PSSA-M. Nevertheless, the number of students assigned to the modified math assessment was fewer than anticipated. Since twice as many students can be called proficient in the AYP calculations based on scores derived from a modified assessment (AA-MAS) than based on the alternate assessment (AA-AAS), it was reasonable to expect twice the level of enrollment in the modified as compared with the alternate. Instead, only about one and a half times as many students were enrolled in the AA-MAS that first year.

Table 1. Tested Student Population in Spring 2010

		n	Percent Who Took the PSSA-M (Math)	Percent Who Took the PASA (Math)	Percent Who took the PSSA (Math)
Grade	4	130,503	1.7%	1.5%	96.8%
	5	131,004	1.9%	1.6%	96.5%
	6	130,941	2.1%	1.5%	96.4%
	7	132,541	2.1%	1.5%	96.3%
	8	135,045	2.2%	1.5%	96.3%
	11	135,480	2.6%	1.5%	95.9%
Total		795,514	2.1%	1.5%	96.4%

Table 2. Tested Population of Students with Disabilities in Spring 2010

		n	Percent Who Took the PSSA-M (Math)	Percent Who Took the PASA (Math)	Percent Who took the PSSA (Math)
Grade	4	22,637	9.6%	8.8%	81.6%
	5	22,752	11.2%	8.9%	79.8%
	6	22,249	12.1%	8.8%	79.1%
	7	22,054	12.8%	9.2%	78.0%
	8	22,488	13.4%	9.1%	77.5%
	11	20,610	17.2%	9.9%	73.0%
Total		132,790	12.6%	9.1%	78.2%

Table 3 summarizes the demographic characteristics and test participation rates of students with IEPs who took the 2010 PSSA-M Math. The majority of students assigned to the test were white (60%), males (60%), and students whose primary language was English (97%). Nearly half (46%) were eligible for Free/Reduced Lunch. These data mirror that of Pennsylvania’s IEP population. The majority of students taking the test were those with a designation of specific learning disability (71%)

Table 3. Demographic Characteristics of Students with IEPs Tested on the 2010 PSSA-M Math Spring 2010

Demographic Category	Percent
Male	40%
Female	60%
White	60%
Minority	40%
English Language Learner	3%
Economically Disadvantaged	46%
Primary Disability	Percent
Specific Learning Disability	71%
Other Health Impairment	10%
Emotional Disturbance	7%
Autism	4%
Mental Retardation	6%
Speech Language Impairment	2%
Hearing Impairment/Deafness	1%
Orthopedic Impairment	<1%
Traumatic Brain Injury	<1%
Multiple Disabilities	<1%
Visual Impairment/Blindness	<1%

Further investigation of “who” took the PSSA-M Math test indicated that the majority (65.1%) of the students assigned to the PSSA-M Math in 2010 were in fact the lowest performers at their grade level on the 2009 general assessment (Table 4). However, as many as 3,347 students (27.8%) were assigned to the modified assessment who had scored in the Basic category, and another 858 (7.1%) who had scored Proficient or Advanced on the regular assessment the previous year.

Table 4. 2009 PSSA Performance Levels of Students Assigned to the 2010 PSSA-M

		Below Basic	Basic	Proficient	Advanced
		<u>Percent (n)</u>	<u>Percent (n)</u>	<u>Percent (n)</u>	<u>Percent (n)</u>
Grade	4	50% (915)	40% (737)	9% (174)	1% (20)
	5	64% (1461)	24% (547)	10% (237)	2% (38)
	6	64% (1579)	30% (752)	4%(109)	1% (26)
	7	72% (1897)	24% (626)	4% (97)	1% (20)
	8	71% (1994)	24% (685)	4% (126)	<1% (11)

There was some fluctuation by grade level. In general, the numbers of very low performing students assigned to the PSSA-M increased as grade level increased. This is in contrast to the relatively stable enrollment rate across grade levels for the alternate assessment, the PASA (see Table 1).

It seems that overall, IEP teams across the state shared our theoretical vision that the 2% target population are exceptional children who are very far from proficient for whom a shorter, less challenging, more accessible test makes sense. Interestingly, however, this theoretical vision was counterbalanced by a very practical one. Some IEP teams assigned students “close” to proficiency to the test. These teams perhaps considered realistically which group of students would attain proficiency given the grade-level content of the test. This practical interpretation of accountability is reflected in the nearly 28% of students in the “Basic” column of Table 4. These are students who were assigned by an IEP team to the 2010 PSSA-M who had “failed” the state test the year before but were closer to reaching proficiency standards than those who scored Below Basic. Ultimately, while we had intended for the eligibility criteria to guide IEP teams toward assigning Below Basic performers to the PSSA-M, IEP teams’ inclusion of students scoring in the Basic range made practical sense for attaining proficiency and did not explicitly violate any state or federal regulation. After all, test assignment decisions are individualized and require IEP teams to examine a variety of student data in addition to the previous year’s state test scores and both Basic and Below Basic scores are considered to be “failures” by AYP standards. However, most surprising was the number of students assigned to the 2010 PSSA-M who had reached or exceeded achievement expectations on the 2009 general assessment. Table 3 indicates that 7% of the students assigned to the easier test were students who had scored Proficient/Advanced on the regular state test the year before. Although IEP teams were not prohibited from assigning high achievers to the AA-MAS, the state considered these students to be “misassigned” to the test and conducted follow-up research to better understand IEP team rationale for these decisions (see Lemons et al., 2012). Bureau of Special Education staff distributed a survey to the directors of special education of each school district that had placed Proficient or Advanced students into the PSSA-M. The survey was sent to 220 school districts and 218 responded. In the survey, districts were asked to indicate who participated in the decision making process, whether participants had been trained on the state guidelines, and whether there were extenuating circumstances that influenced the test assignment decisions. Interestingly, only 76.6% of respondents indicated that the IEP team had been involved in the assignment decision, even though this is a federal requirement. The other 23.4% of responses indicated that decisions were made, in varying degrees, by individual special education teachers or school/district leadership (e.g., principals, special education supervisors, or assessment coordinators). A majority (77.1%) indicated that decision makers had been trained on the state guidelines. More than half (58.7%) of respondents indicated that extenuating circumstances influenced the assignment. Explanations of the “extenuating circumstances” made clear that in many cases the “misassignments” were intentional and the decisions had been made thoughtfully.

Many reported using data other than the state assessment (e.g., local assessments, curriculum-based measurement) that suggested the child would be unlikely to pass the general assessment in 2010 despite performance in 2009. Some made the assignment on non-academic grounds (e.g., emotional, behavioral, or medical concerns). In a few cases there were no available data on a child who recently transferred to the district. In only two cases was the decision the result of a parent request.

Student Performance in 2010

The central tenet of the 2% flexibility rests on the assumption that the modified assessment option will afford low-performing students with disabilities to show what they know and enable them to achieve proficiency on an assessment that is more sensitive to their unique learning needs than is the general accountability test. Unfortunately, the group performance data summarized in Table 5 suggest that despite the state’s effort to make grade-level math items less cognitively complex and more accessible (Pennsylvania Department of Education, 2010a), fewer than half the students with IEPs assigned to the PSSA-M Math (about 44.9%) scored at Proficient or above. The remaining 55.1% of students with IEPs who took the PSSA-M Math scored in the Basic or Below Basic range. As grade level increased so did the proportion of students failing to reach the modified proficiency standards. For example, nearly two-thirds of the grade 11 students were not proficient on the test.

Table 5. Performance Levels of Students with IEPs on 2010 PSSA-M Math

		Percent Advanced	Percent Proficient	Percent Basic	Percent Below Basic	Percent “Passed” (Proficient+Advanced)	Percent “Failed” (Basic+Below Basic)
Grade	4	21.2%	38.3%	35.8%	4.8%	59.5%	40.6%
	5	13.0%	38.0%	43.4%	5.5%	51.0%	48.9%
	6	20.0%	23.4%	22.8%	33.8%	43.4%	56.6%
	7	8.1%	33.2%	50.4%	8.3%	41.3%	58.7%
	8	5.6%	35.2%	49.2%	10.0%	40.8%	59.2%
	11	5.6%	27.6%	45.3%	21.6%	33.2%	66.9%
Overall						44.9%	55.1%

A primary rationale for the AA-MAS option was that a more accessible, less cognitively taxing test than the general assessment would enable the targeted group of low achieving students with disabilities to demonstrate proficiency of grade-level academic content. Moreover, NCLB policy defines school accountability through the calculation of Adequate Yearly Progress that hinges on the percentage of students who meet or exceed scores at the proficient level on state tests. Therefore, the achievement data summarized herein are categorized as Proficient (i.e., Proficient + Advanced) vs. Not Proficient (i.e., Below Basic + Basic). We recognize that this pass/fail conceptualization of student performance does not communicate discrete movement between

or among performance categories (see *Lessons Learned* at the end of this chapter); nonetheless, these analyses do speak to whether or not AA-MAS participation influenced proficiency attainment for the 2% target population in Pennsylvania. Table 6 reports the assessment results for the student sample profiled in Table 4—those students in Pennsylvania who took the 2010 PSSA-M who also had 2009 general PSSA assessment scores. Results suggest that for a majority of students (57.4%) the outcome on the modified test was no different from the outcome on the regular assessment—students who were not proficient in 2009 remained not proficient in 2010 and students who were proficient in 2009 remained proficient in 2010. Neither the modifications made to test items nor the less rigorous achievement standards were enough to positively impact test performance of 46% of the students unable to pass the general assessment the year before. In contrast for 41% of students, the PSSA-M accomplished what it was supposed to accomplish: 4,936 students who were not proficient in 2009 on the regular assessment scored in the proficient range on the modified assessment in 2010. The greatest movement from not proficient to proficient occurred in 4th grade (54%), the initial year of eligibility for the modified assessment in Pennsylvania. Improved achievement occurred for a slightly smaller proportion of students in grades 5 and 6 (46%), and for an even smaller proportion of students in grades 7 and 8 (39%). These results suggest that while achievement outcomes for the majority of students remained the same on the new test, the PSSA-M enhanced the performance of nearly half of the elementary-aged children taking the test.

Table 6. Analysis of Achievement Changes from 2009 PSSA Math to 2010 PSSA-M Math for Matched Student Sample

		Not Proficient to Proficient	Proficient to Not Proficient	Remained Not Proficient	Remained Proficient
		<u>Percent (n)</u>	<u>Percent (n)</u>	<u>Percent (n)</u>	<u>Percent (n)</u>
Grade	4	54% (900)	12% (24)	46% (752)	88% (170)
	5	46% (931)	20% (55)	54% (1077)	80% (220)
	6	46% (1071)	32% (43)	54% (1260)	68% (92)
	7	39% (987)	27% (32)	61% (1536)	85% (73)
	8	39% (1047)	27% (38)	61% (1632)	73% (99)
Overall		Positive Impact 41.0%	Negative Impact 1.6%	No Proficiency Status Change 57.4%	

Moving into 2010-2011 School Year

To prepare for the full-scale implementation of an AA-MAS in reading, math, and science in 2011, several activities were undertaken to better understand why students were assigned to the PSSA-M Math in 2009-10. We conducted a follow-up focus group to determine how schools were making the decisions regarding which students were to be placed into the PSSA-M. Electronic surveys were sent to 12 teachers in eight school districts in Pennsylvania. These teachers were selected because they were either participating in a larger research project being conducted by

the authors (n=8) or they were participating in a graduate level class taught by one of the authors (n=4). Email correspondence and phone interviews were used to follow up with teachers as needed to clarify and add additional detail to responses. Due to the limited sample size and the purposive sampling, caution should be taken in generalizing findings to the state as a whole. However, the responses do provide insight into the implementation of the PSSA-M.

Assignment process. The aim of the focus group was to better understand the process schools were using to place students into the test and to determine which group of students schools were targeting. Respondents reported three different processes for PSSA-M test assignment. For the first process, respondents indicated that their school used the ASIST guidelines and that decisions were made during an IEP team meeting. In the second process, respondents indicated that school administrators (e.g., director of special education, building principals) reviewed various data sources and had then provided the special education teachers with a list of students to be placed into the PSSA-M. In the third process, respondents indicated that the decision was left to the special education teacher. In one of those instances, the teacher responsible for decision-making indicated that administration explicitly asked special education teachers not to involve parents in fear that parent requests would inflate test assignment.

Target group. Next, we asked teachers whether any particular group of students with IEP (i.e., those scoring Below Basic, Basic, Proficient, or Advanced) represented the PSSA-M target group at their school. Again, variation was observed. Four teachers believed that their district had targeted any student with an IEP scoring in the Basic range. In those instances, teachers reported reviewing data to determine which students would most likely answer grade-level questions accurately if modifications were applied. Three teachers indicated that only students scoring in the low range of Below Basic (but not eligible for the AA-AAS) were the targets for the modified assessment. Previous PSSA data were used as the primary source of data to make those decisions. Two teachers speculated that any student with an IEP who failed the previous year's PSSA was the target. Three teachers reported they were unsure of the students targeted at their school. One 11th grade teacher reported that test assignment decisions were challenging because almost all students on her caseload met state criteria.

Teacher Perceptions of the AA-MAS. In addition to asking teachers about test assignment and target population, they were asked their opinions about PSSA-M math implementation overall. Teachers who reported that test assignment was an IEP team decision or special education teacher decision were satisfied. These teachers reported that the process was understandable and clear. Teachers in schools where administrators selected students to participate were not satisfied. These teachers thought the process was unclear and they desired input. They wanted improved communication and team decision-making. One teacher felt that parents and students should be involved.

When asked about students' testing experiences, a majority of teachers provided positive feedback. In general, teachers reported that students were less frustrated and appeared to be more confident about their performance. Students shared that they liked being able to write answers directly in the book and they liked the reduced length of the assessment. Most teachers felt that the PSSA-M items were more closely matched to classroom-based assessments than to PSSA items. One teacher expressed concern that despite modifications the math items were still too challenging for her very low achieving students. Generally, however, teachers felt that, no matter what achievement outcomes were, students had a more positive test experience with the PSSA-M than with the general assessment.

Finally, additional teacher feedback about the PSSA-M reflected three themes. First, teachers felt that if the modifications made to the assessment were valid that there should be no limit on the number of students who could count as proficient on this exam. Teachers expressed frustration about administrative pressures to assign a small number of children to the assessment (resulting from confusion about the 2% cap) because teachers genuinely believed that more students would benefit from a less stressful testing experience. Second, teachers in schools where administrators made test assignment decisions wanted improved communication about the decision-making process. They also believed that test assignment decisions were made too close to the testing window to adequately prepare students. Third, some teachers wondered if it was justifiable to call students "proficient" on an easier assessment. This group reported that it was difficult to explain to parents what proficiency against modified academic achievement standards means.

Overall, most teachers were happy to have an additional assessment option for students with IEPs. They believed that decision-making would improve after multiple years of implementation. They also felt that more experience with the PSSA-M would better equip them to assign the "right" students, those who could achieve proficiency, to the test. Over time, they planned to incorporate PSSA-M-like modifications into instruction to better prepare students for the test.

2011 PSSA-M—Year Two of State-wide AA-MAS Implementation

Student Participation in 2011

In the 2010-11 school year, PSSA-M was administered in all three tested content areas at grades 4 through 8 and 11. In order to examine longitudinal participation and performance, the GSEG analyzed math data only.

In the spring of 2011, nearly 23,000 fewer Pennsylvania students participated in statewide accountability assessments (a decrease from 795,514 in 2010 to 772,645 in 2011) although the number of students with disabilities and IEPs who participated actually grew by 1,918 (from 132,790 in 2010 to 134,708 in 2011). Now, in year two of the test, nearly 3% of the total tested population took the PSSA-M Math subtest, with almost a 1% increase in participation rates at

every grade level (see Tables 7 and 8). This time, however, within the tested population of students with IEPs, the distribution of students across the three accountability assessments changed substantially, especially in the upper grades. While only 10% of 4th graders with IEPs were again assigned to the modified math assessment, over 20% of 11th graders with IEPs were assigned to the PSSA-M Math. Still, most students with disabilities took the general PSSA math subtest (74.5%, with a range from 89% in 4th grade to 69.8% in 11th grade) while 16% took the PSSA-M and 9% took the PASA. The majority of students (60.5%) who took the PSSA-M in 2010 took the modified math test again the following year. IEP teams also decided that the AA-MAS was not the best “fit” for some students in the original cohort and assigned 37.8% of them to the PSSA Math in 2011 and 1.7% of them to the AA-AAS that year.

Table 7. Student Population Tested in Spring 2011

		n	Percent Who Took the PSSA-M (Math)	Percent Who Took the PASA (Math)	Percent Who took the PSSA (Math)
Grade	4	126,918	1.9%	1.7%	96.4%
	5	128,333	2.6%	1.6%	95.8%
	6	128,571	2.8%	1.7%	95.5%
	7	129,187	3.1%	1.6%	95.3%
	8	129,636	3.2%	1.7%	95.2%
	11	130,000	3.3%	1.6%	95.1%
Total		772,645	2.8%	1.6%	95.6%

Table 8. Distribution of Students with Disabilities on 2011 Assessment Options

		n	Percent Who Took the PSSA-M (Math)	Percent Who Took the PASA (Math)	Percent Who took the PSSA (Math)
Grade	4	23,387	10.2%	9.3%	80.5%
	5	23,335	14.4%	8.9%	76.7%
	6	22,155	16.2%	9.7%	74.0%
	7	22,255	17.8%	9.3%	72.9%
	8	22,636	18.2%	9.4%	72.4%
	11	20,940	20.4%	9.8%	69.8%
Total		134,708	16.1%	9.4%	74.5%

The demographic characteristics of the 2011 PSSA-M Math population mirrored those of the year-one cohort with a slight increase in the proportion of students eligible for Free/Reduced Lunch (58%). As expected, the majority of students taking the test were those with a specific learning disability (66%) (Table 9).

Table 9. Demographic Characteristics of the 2011 Students Tested on PSSA-M Math

Demographic Category	Percent
Male	59%
Female	41%
White	67%
Minority	33%
English Language Learner	3%
Economically Disadvantaged	58%
Primary Disability	Percent
Specific Learning Disability	66%
Other Health Impairment	10%
Emotional Disturbance	6%
Autism	6%
Mental Retardation	5%
Speech Language Impairment	1%
Hearing Impairment/Deafness	<1 %
Orthopedic Impairment	<1%
Traumatic Brain Injury	<1%
Multiple Disabilities	<1%
Visual Impairment/Blindness	<1%

Our exploration of the characteristics of the 2011 AA-MAS tested population centered on two groups of students: (1) students reassigned to the PSSA-M and (2) students newly assigned to the PSSA-M. Table 10 focuses on the first group.

Students reassigned to the PSSA-M math. Approximately 38% of students taking the modified math test in spring 2011 had also taken the modified test in 2010. Most of the students reassigned to the PSSA-M had scored in the Basic and Proficient ranges in the previous year (see *Basic* and *Proficient* columns of Table 10.) Among those reassigned to the modified assessment, only 7% had scored in the Below Basic range the previous year. No clear explanation could be found for this pattern of reassignment. Perhaps IEP teams deemed the modified math test the best fit for students who were close to Proficient. Or perhaps, IEP teams thought reassignment to the modified test might give Proficient students an opportunity to score at the Advanced level. Perhaps IEP teams thought students previously scoring Basic and Proficient on a modified math test were the most likely to achieve Proficient status in 2011 and ultimately boost their school or school district's AYP status. Or, perhaps schools were resigned to the fact that students who performed in the Below Basic range were unlikely to positively affect the 2% cap so fewer were reassigned to the test. Overall, 43.5% of the students that IEP teams reassigned to the PSSA-M

math test had already demonstrated proficiency on a modified math test the previous year; 56.5% of those reassigned had not scored in the Proficient range the previous year.

Table 10. 2010 PSSA-M Performance Levels of Students Assigned Again to the 2011 PSSA-M

		Below Basic	Basic	Proficient	Advanced
		<u>Percent (n)</u>	<u>Percent (n)</u>	<u>Percent (n)</u>	<u>Percent (n)</u>
Grade	5	5.2% (60)	40.6% (473)	38.7% (451)	15.5% (181)
	6	6.4% (85)	49.7% (658)	34.1% (451)	9.8% (130)
	7	7.7% (117)	47.8% (731)	37.4% (572)	7.1% (109)
	8	9.1% (155)	56.2% (958)	30.2% (515)	4.5% (77)
Overall	7.2% (417)	49.3% (2820)	34.8% (1989)	8.7% (479)	

Table 11. 2011 PSSA Performance Levels of Students Newly Assigned to the 2011 PSSA-M

		Below Basic	Basic	Proficient	Advanced
		<u>Percent (n)</u>	<u>Percent (n)</u>	<u>Percent (n)</u>	<u>Percent (n)</u>
	4	50.3% (988)	42.4% (832)	6.3% (124)	<1.0% (5)
Grade	5	62.9% (1081)	26.8% (460)	8.0% (137)	1.2% (20)
	6	66.0% (1241)	28.8% (530)	3.9% (74)	<1.0% (7)
	7	67.5% (1408)	26.7% (557)	3.4% (70)	<1.0% (16)
	8	75.4% (1602)	19.0% (403)	2.8% (60)	<1.0% (13)
Overall	65.6% (6320)	28.9% (2782)	4.8% (469)	0.6% (61)	

Students newly assigned to the PSSA-M. Table 11 reports the previous year’s achievement data for the second group of students—those students taking the PSSA-M Math subtest for the first time in 2011. With respect to this group, IEP teams seem to have understood the eligibility criteria for participation in the assessment and assigned the group for which the assessment was intended. 94.5% of those newly assigned to the PSSA-M math were *not* proficient in math the previous year, and about 70% of those students had scored in the Below Basic range on the regular test in the previous year.

Overall, these data are strikingly similar to the 2010 test assignment data discussed earlier. IEP teams across the state generally held fast to the state guidance that the majority of students assigned to the modified test should be those students very far from proficiency on the general assessment (i.e., scoring Below Basic). Again, it was evident that IEP teams also considered the feasibility that students who scored in the Basic range on the regular grade level test would benefit from the reduced cognitive load, improved accessibility, and modified achievement standards of the PSSA-M. Despite the Pennsylvania Department of Education’s efforts to discourage including high performers in the PSSA-M, we see that some Proficient and Advanced students were assigned to test in 2011 (see columns 3 & 4 of Table 11). Though there were fewer instances

of these students being assigned, an average of 5.4% of students with IEPs considered to have mastered grade-level content were assigned to the less challenging test.

Student Performance in 2011

Examination of student performance on the 2011 PSSA-M Math shows, again, that fewer than half the students (42.2%) assigned to the modified test scored at Proficient or Advanced, this rate actually down from the previous year (44.9%; see Table 12). The data also show that most students (i.e., 71.1% of those taking the PSSA-M Math a second time and 56.5% of those taking a modified PSSA for the first time) did not change their proficiency status from 2010 to 2011. Only 11% of students taking the modified math assessment for the second year moved from not-proficient to proficient status; 42% of those taking the modified math assessment for the first time moved from not-proficient to proficient. The PSSA-M was the better assessment option than the general assessment for these low performing students with IEPs.

Table 12. 2011 PSSA-M Math Achievement

		Percent Advanced	Percent Proficient	Percent Basic	Percent Below Basic	Percent "Passed" (Proficient+Advanced)	Percent "Failed" (Basic+Below Basic)
Grade	4	17.5%	36.2%	40.5%	5.8%	53.7%	46.3%
	5	9.5%	35.7%	44.4%	10.4%	45.2%	54.8%
	6	9.7%	33.2%	45.8%	11.4%	42.9%	57.2%
	7	8.5%	32.5%	53.5%	5.5%	41.0%	59.0%
	8	6.7%	31.7%	48.7%	13.0%	38.4%	61.7%
	11	7.0%	25.4%	43.1%	24.4%	32.4%	67.5%
Overall						42.2%	57.8%

Analysis of changes in individual student achievement across two years of the PSSA-M revealed some interesting trends. Table 13 reports the assessment results for students with two years of Math scores on the modified assessment. Here we see that achievement was pretty stable from year to year, meaning students who were not proficient (i.e., scoring in the Basic or Below Basic ranges) in 2010 remained not proficient in 2011 and students who were proficient (i.e., scoring Proficient or Advanced) in 2010 remained proficient in 2011. Data displayed in the third column shows that approximately 80% of students at each grade level failed the test two years in a row. Moreover, the percentage of students maintaining proficiency from year to year ranged from 54% at 5th grade to 66% at 8th grade (Column 4). Analysis also showed that many students did "worse" on the 2011 PSSA-M than on the 2010 PSSA-M (see the *Proficient to Not Proficient* column). Specifically, almost half (46%) of the students who were Proficient on the 4th grade math test were not proficient at the end of 5th grade. Similar patterns of depressed achievement occurred from grades five to six, six to seven, and seven to eight (41%; 39%; 34% respectively).

In contrast, improved achievement occurred for about 20% of fifth through eighth graders from 2010 to 2011 (see the *Not Proficient to Proficient* column).

Table 13. Matched Student Sample Analysis of Achievement Changes from 2010 PSSA-M Math to 2011 PSSA-M Math

		Not Proficient to Proficient	Proficient to Not Proficient	Remained Not Proficient	Remained Proficient
		Percent (n)	Percent (n)	Percent (n)	Percent (n)
Grade	5	20.6% (110)	46.4% (293)	79.4% (423)	53.6% (339)
	6	22.1% (164)	41.3% (240)	77.9% (579)	58.7% (341)
	7	20.4% (173)	38.8% (264)	79.6% (675)	61.2% (417)
	8	19.0% (212)	33.6% (199)	81.0% (901)	66.4% (393)
Overall		Positive Impact 11.5%	Negative Impact 17.4%	No Proficiency Status Change 71.1%	

Table 14 details interesting patterns of student achievement across two years of the PSSA-M. Analysis of student movement between proficiency categories revealed that most non-proficient students moved from Below Basic status to Basic status on the next grade level's test. For example, 38% of the Below Basic 4th grade students now taking the 5th grade test scored in the Basic range. The same improvement was evident for 37% of the 2010 5th grade cohort, 68% of the 6th grade cohort, and 47% of the 7th grade cohort. Students' overall static performance as they matriculated to the next grade level raises some important questions about student achievement and the value added of the AA-MAS. Although they still failed the state test, all of those students edged closer to proficiency status in 2011.

Unfortunately, some of their non-proficient peers moved even further away from proficiency status. For example, 22% of the students scoring Basic on the 2010 PSSA-M 4th grade Math subtest were Below Basic on their 5th grade test. The same was true for 17% of the 2011 6th graders; 7% of 7th graders, and 15% of 8th graders. Positive movement occurred most frequently from the Basic achievement category to the Proficient achievement category. Between 20-22% of Proficient students in 2011 had scored in the Basic range in 2010. Most of the 2010 proficient group (students scoring Proficient or Advanced) who dropped achievement levels scored in the Basic range in 2011 with as many as 51% of students moving from Proficient in 4th grade to Basic in 5th grade; 39% moving from Proficient in 5th grade to Basic in 6th grade; 41% moving from Proficient in 6th grade to Basic in 7th grade, and 34% moving from Proficient in 7th grade to Basic at the end of 8th grade. Little movement was evident between the other performance categories across the two years.

Table 14. 2010 to 2011 PSSA-M Math Matched Student Sample Analysis of Movement Between Proficiency Categories

			2011 Performance Level				
			Percent Below Basic	Percent Basic	Percent Proficient	Percent Advanced	
2011 Grade Level	5	2010 Performance Level	Below Basic	55.0%	38.3%	3.3%	3.3%
			Basic	22.0%	55.6%	20.7%	1.7%
			Proficient	2.9%	51.2%	37.3%	8.6%
			Advanced	1.1%	26.0%	48.6%	24.3%
	6		Below Basic	60.0%	36.5%	3.5%	>1.0%
			Basic	16.9%	58.7%	22.5%	2.0%
			Proficient	4.0%	39.2%	47.9%	8.9%
			Advanced	3.8%	30.8%	45.4%	20.0%
	7		Below Basic	29.9%	67.5%	1.7%	>1.0%
			Basic	7.3%	69.5%	20.2%	3.0%
			Proficient	1.4%	40.6%	47.6%	10.5%
			Advanced	>1.0%	21.1%	49.5%	28.4%
8	Below Basic	51.6%	47.1%	1.3%	>1%		
	Basic	15.4%	62.6%	19.8%	2.1%		
	Proficient	2.1%	34.4%	52.2%	11.3%		
	Advanced	>1%	14.3%	55.8%	29.9%		

Table 15 reports the assessment results for the students profiled in Table 11—those students in Pennsylvania who took the general PSSA Math test in 2010 and the PSSA-M Math test in 2011. Much like the year-one results, these data suggest that in many cases, students who previously “failed” to make AYP on the general assessment met or surpassed proficiency standards on the PSSA-M. The greatest “movement” from non-proficiency to proficiency occurred in 4th grade (50%). Improved achievement occurred for approximately the same proportions of students in grades 5 through 8 ranging from a 40% to 45.5% jump in proficiency from 2010 to 2011. Overall, 44% of students who were not proficient on the regular test demonstrated proficiency on the modified test. Unfortunately, a small number of students sacrificed proficiency when moving from the PSSA to the PSSA-M. Across grades four through eight, 192 students (22%)

of the students who scored Proficient or Advanced on the 2010 general assessment scored in the Basic or Below Basic range on the 2011 modified test. Columns 3 & 4 of Table 15 display the numbers of students who saw no achievement level change from the general PSSA to the alternate PSSA-M. That is, students who failed the general assessment in 2010 and also failed the modified assessment the next year or those students who passed the test taken each year. The “Remained Not Proficient” statistics reported in Table 15 coupled with the “Remained Not Proficient” statistics in Table 13 diminish the returns of the AA-MAS option to enhance the achievement of the lowest persistently low performing students with disabilities.

Table 15. Matched Student Sample Analysis of Achievement Changes from 2010 PSSA Math to 2011 PSSA-M Math

		Not Proficient to Proficient	Proficient to Not Proficient	Remained Not Proficient	Remained Proficient
		Percent (n)	Percent (n)	Percent (n)	Percent (n)
Grade	4	49.7% (905)	12.4 (16)	50.2 (913)	87.6 (113)
	5	45.5% (701)	26.8 (42)	54.3 (836)	73.2 (115)
	6	44.3% (784)	30.9 (25)	55.6 (984)	69.1 (56)
	7	41.0% (806)	20.9 (18)	58.5 (1150)	75.6 (58)
	8	40.0% (802)	20.5 (15)	53.9 (1081)	79.5 (58)
Overall		Positive Impact 42.2%	Negative Impact 1.2%	No Proficiency Status Change 56.6%	

Exploration of student achievement changes between and among performance categories again tells an interesting story about the perceived impact of the PSSA-M for students with disabilities in Pennsylvania. Very few, 11.5%, of the students who scored Below Basic on the PSSA Math in 3rd grade remained Below Basic on the 4th grade PSSA-M Math test. In addition, 56% of that group moved into the Basic range on the Modified test, while over 1/4 of their peers (26%) performed well enough to reach Proficient—6% even scored in the advanced range. Similar trends occurred at higher grade levels. Between 36.5% and 67.5% of 2010 PSSA Below Basic achievers scored Basic on the PSSA-M the following year. The most significant performance boost was experienced by students who had scored in the Basic range on their regular grade level test in 2010 with as few as 46.6% (4th grade) and as many as 59% (8th grade) of these students achieving Proficient on the 2011 Modified test. For the 22% of students who scored Proficient or Advanced on the PSSA in 2010 but did poorly on the PSSA-M one year later, a move from Proficient to Basic was the most common trend.

2012 PSSA-M—Year Three (and Final Year) of State-wide AA-MAS Implementation

Participation in 2012

Despite the fact that GSEG funding ended on October 31, 2010, the GSEG research and practitioner community continued to monitor participation and performance of students with disabilities on the state’s AA-MAS. During the 2011-12 school year, IEP teams again had the responsibility to assign students with IEPs who had persistent academic problems to one of three versions of the annual statewide accountability assessment: the regular assessment (PSSA); the modified assessment (PSSA-M); or the alternate assessment (PASA). A preliminary study of the PSSA-M target population indicated that after 3 years of decision-making IEP teams continued to make some very idiosyncratic decisions.

Data were available for two subsamples of students: (a) students who had taken the modified math test in 2011, but whose IEP team assigned the student to take the regular assessment in 2012; and (b) students who had taken the regular assessment in 2011 whose IEP team assigned them to take the PSSA-M Math assessment in 2012. What follows is a brief description of these groups.

Students newly assigned to the PSSA-M in 2012. Table 15 shows that 8,940 students were newly assigned to the PSSA-M Math test in the 2011-12 school year, with the largest number of these students entering the modified assessment in 4th grade. As in the previous year, IEP teams seemed to understand the ASIST guidelines and primarily assigned students who had scored in the Below Basic range the previous year on the regular assessment to the modified assessment. However, as in previous years, the teams also assigned students who appeared not to need a modified assessment to demonstrate math proficiency.

Table 16. 2011 PSSA Performance Levels of Students Assigned to the PSSA-M Math in 2012

		Below Basic	Basic	Proficient	Advanced
		<u>Percent (n)</u>	<u>Percent (n)</u>	<u>Percent (n)</u>	<u>Percent (n)</u>
Grade	4	53.7% (1315)	38.3% (938)	7.2% (177)	0.8% (19)
	5	60.2% (1103)	26.9% (492)	10.7% (196)	2.2% (40)
	6	56.9% (862)	35.0% (531)	6.8% (103)	0.1% (20)
	7	70.3% (1139)	24.6% (398)	3.6% (59)	1.4% (23)
	8	72.8% (1110)	23.2% (354)	3.7% (56)	0.3% (5)
Overall		61.9% (5529)	30.4% (2713)	6.6% (591)	1.2% (107)

Students assigned “back” to the regular assessment. About 28% of students with IEPs in grades 5 through 8 who had taken the PSSA-M in grades 4 through 7 in 2011 were assigned back to the regular assessment in 2012 (see Table 17). A majority of those reassigned (58.1%) had scored in the Proficient or Advanced ranges in the modified assessment in 2011, but 42% had not achieved

proficiency on the modified assessment in 2010. Since the modified assessment was designed to enable more students with disabilities to demonstrate proficiency, we were perplexed as to why IEP teams would assign Proficient PSSA-M students back to the PSSA.

Table 17. 2011 PSSA-M Performance Levels of Students Assigned Back to the Regular Assessment in 2012

		Below Basic	Basic	Proficient	Advanced
		<u>Percent (n)</u>	<u>Percent (n)</u>	<u>Percent (n)</u>	<u>Percent (n)</u>
Grade	4	-	-	-	-
	5	3.0% (18)	30.3% (181)	40.1% (240)	26.6% (159)
	6	5.0% (43)	34.6% (295)	42.9% (366)	17.5% (149)
	7	6.4% (60)	38.2% (360)	37.2% (350)	18.3% (172)
	8	2.2% (23)	44.2% (457)	37.0% (383)	16.6% (172)
Overall		4.2% (144)	37.7% (1293)	39.1% (1339)	19.0% (652)

Perhaps students who scored Advanced on the modified assessment the previous year were considered to have been “misplaced” in the assessment options and reassigned to a harder test. Perhaps too, in their annual review of student test participation criteria, IEP teams decided that these students were not eligible for the PSSA-M in 2012 and therefore assigned them to the PSSA. It is also difficult to understand why students who scored in the Basic range on the modified assessment would be reassigned to the ‘harder’ assessment the next year. Additional analyses of test assignment trends across years are needed to dismiss or confirm these speculations.

Performance in 2012

As in the two previous years, switching to the modified math assessment did not change the earned performance level for a majority of students (see Table 17); 5,115 (57.2 % of students who took the modified assessment for the first time in Spring 2012) remained in the same performance category as the year before on the general assessment. However, the switch to the PSSA-M math did result in 3,643 students (40.7% of those students with IEPs taking the AA-MAS for the first time) demonstrating proficiency in math.

Table 18. Matched Student Sample Analysis of Achievement Changes from 2011 PSSA Math to 2012 PSSA-M Math

		Not Proficient to Proficient	Proficient to Not Proficient	Remained Not Proficient	Remained Proficient
		Percent (n)	Percent (n)	Percent (n)	Percent (n)
Grade	4	42.5% (1042)	1.9% (46)	49.4% (1211)	6.1% (150)
	5	39.3% (719)	3.1% (52)	47.8% (976)	9.8% (179)
	6	44.0% (667)	2.6% (40)	41.3% (626)	5.5% (83)
	7	38.9% (629)	1.5% (25)	56.1% (908)	3.5% (57)
	8	38.4% (586)	0.9% (14)	57.6% (878)	3.1% (47)
Overall		Positive Impact 40.7%	Negative Impact 2%	No Proficiency Status Change 57.2%	

The final data presentation in Table 19 illustrates an important GSEG discovery that sends an important message to the assessment community as we move away from AA-MAS implementation. First, the data suggest that students with disabilities who were competent in grade level content were able to demonstrate that grade-level content knowledge no matter the assessment (see the *Remained Proficient Column*). By the same token, the students with disabilities who had not mastered grade level math content were unable to demonstrate proficiency on either test (see the *Remained Not Proficient Column*). That is, for 55.1% of students reassigned back to the regular assessment in 2012, there was no change in performance status. Second, many students (44%) who had scored in the Proficient range based on the modified achievement standards in 2011 failed the following year when held to general achievement standards of the PSSA. Those 15,007 students scored Proficient on the more accessible test but could not demonstrate proficiency once the modifications were removed. Clearly, while the improved accessibility of the PSSA-M was not the only ingredient for success on the annual accountability assessment, it was certainly an important ingredient for a substantial number of students with disabilities. These data are important for states to consider as they move ahead with new assessment options next year. What impact will dropping the AA-MAS have on the achievement of students with disabilities in Pennsylvania as PSSA-M participants return to the general assessment in 2013?

Table 19. Matched Student Sample Analysis of Achievement Changes from 2011 PSSA-M Math to 2012 PSSA Math

		Not Proficient to Proficient	Proficient to Not Proficient	Remained Not Proficient	Remained Proficient
		Percent (n)	Percent (n)	Percent (n)	Percent (n)
Grade	4	-	-	-	-
	5	0.5% (3)	56.8% (340)	32.8% (196)	9.9% (59)
	6	0.9% (8)	49.1% (419)	38.7% (330)	11.3% (96)
	7	1.7% (16)	37.4% (353)	42.9% (404)	17.9% (169)
	8	0.9% (9)	38.2% (395)	45.5% (471)	15.5% (160)
Overall		Positive Impact 0.8%	Negative Impact 44.1%	No Proficiency Status Change 55.1%	

Discussion: Lessons Learned

The apparent vagueness of the federal guidance led many states to pull together a group of advisors to assist in determining whether their state would develop and implement the AA-MAS. And, if so, to answer these questions: Who would it target? What would it look like? What would it measure? In 2007, the Pennsylvania’s Bureau of Special Education was awarded a General Supervision Enhancement Grant (GSEG) by the U.S. Department of Education’s Office of Special Education Programs to engage in research and inquiry about issues related to the development and implementation of the AA-MAS in the state. The primary goals of the grant were to: (a) identify the target population for the test, (b) to investigate those students’ opportunity to learn grade level content so as to inform and improve the state’s framework for standards-aligned instruction and assessment, and (c) determine the impact of the AA-MAS on the distribution of student performance across the four achievement categories (Below Basic, Basic, Proficient, and Advanced). The GSEG team, which included the chapter authors, was one group charged with providing guidance to state leaders as they moved forward with the development and implementation of the AA-MAS in Pennsylvania. The team conducted various activities to gather information related to the three objectives, and continued to work on these issues long after the GSEG funding ended.

In the previous pages, we summarized the process of defining the target population for a Pennsylvania AA-MAS and the development of guidelines for IEP teams to facilitate test assignment decisions for the state’s three annual accountability assessment options: the PSSA, the PSSA-M, or the PASA. We also summarized the characteristics and achievement of students assigned to the AA-MAS in the first three years of implementation of the modified math assessment. Based on the GSEG activities and the data reported, we feel confident in making the following generalizations about lessons learned.

1. The target population should be fully understood and clearly defined prior to test development. For reasons of expedience, Pennsylvania sanctioned two different groups to work on aspects of the AA-MAS initiative. The Bureau of Assessment and Accountability contracted test developers to create the modified assessment itself. The Bureau of Special Education tasked the GSEG with defining the target population for the AA-MAS and developing the IEP team decision-making guidelines for identifying those target students. The result of these separate activities was an assessment that many teachers felt was too challenging for the very low-performing students they were directed to assign to the test. The GSEG team believes that it makes more sense to create assessments *after* the target population has been selected. If tests are developed without a clear understanding of who will be taking them, a mismatch is likely and results are less educationally useful for teachers.

2. IEP team decisions about students with a disabilities' participation in accountability assessments are not made in isolation. Individual student test assignment decisions are part of the total plan that defines an appropriate education for the student. As such, decisions should be made at the annual IEP meeting, with all the important members present, and in conjunction with the development of IEP goals. If the student is making adequate progress in the general education curriculum, the recommendation regarding participation is indisputable—the default decision is the general assessment, with or without accommodations. The discussion at the IEP team meeting should focus on the nature of the accommodations routinely used to enhance this student's typical educational experience. These accommodations should be recommended to make the assessment more accessible and to have the assessment results better reflect what the student knows and is able to do.

The second decision is also fairly straightforward. If the student has a significant cognitive disability, has IEP objectives and is receiving instruction based on extended or alternate academic content standards, and requires significant scaffolding to participate meaningfully in the general education curriculum, the team should recommend that the student participate in annual state-wide accountability assessments through the alternate assessment based on alternate academic achievement standards (AA-AAS).

Of course, if the student is not making adequate progress in the general education curriculum, but is not eligible to be assigned the AA-AAS, the IEP team may consider a different assessment option by reviewing and following guidelines provided by their State Department of Education. This may mean using flow charts, decision trees, or checklists designed uniquely for their state. However, the decision of whether a student should take the AA-MAS is not nearly as clear-cut as assignment to either the general test or to the AA-AAS. Within the parameters of the state guidelines, IEP teams would be wise to consider the following:

- Make certain all members of the IEP team are clear about what the participation decision for the individual student is supposed to accomplish.
- Don't prejudge the outcome of the decision based on the student's diagnostic label or the setting in which special education services are provided.
- Don't let the team decision be unduly influenced by the quality and alignment with grade level standards of previous years' IEPs or the limited educational accomplishment of the student that resulted from those previous educational experiences.
- Craft the IEP so that the student has the opportunity to learn what will be tested (i.e. this section of the IEP has ramifications for other sections of IEP).
- Remember that an IEP team cannot "stop the clock," that is, suspend participation in the annual accountability assessments until the student has appropriately written IEP goals and access to appropriate instruction and accommodations.
- Don't be intimidated by the "cap" on how many students' scores from alternate assessments can be counted as proficient.
- IEP team members must understand that there will be intended and unintended consequences to the assessment participation decision.

IEP teams in annual IEP meetings are not making grand assessments of the efficacy of the standards based accountability system. They are making recommendations and decisions to ensure an appropriate education for a particular student in a particular district, in a particular state. Each member of the team—the LEA representative, the educators, and the student's parents—should be fully informed about the nature of the decisions that need to be made and the implications of those decisions in shaping the student's educational experiences. There are federal, state, and perhaps even district guidelines that should influence the decision, but in the end, the decision needs to be a very personalized one, reflecting only the team consensus of what is appropriate for this particular student at this particular time in his/her educational career, until the next year when the decision is made again.

IEP teams make decisions for individual students using the guidelines set forth. Although some decisions may appear to be misguided (i.e., a student is assigned to an assessment for which he does not qualify), IEP teams try to have the best interests of children in mind. Careful monitoring of the outcomes of assessment decisions is essential to ensuring that the intended positive consequences are occurring and unintended negative consequences are not.

3. Given the litigious nature of special education, if there are opportunities to make life easier for students with disabilities and their families, states likely will pick up the gavel.

Therefore it is not surprising that 14 states, including Pennsylvania, developed an AA-MAS when the option was put forward. We sincerely hope that the lessons learned about suitable accommodations for students with disabilities will not be forgotten. Many of the accommodations did, indeed, make the AA-MAS more accessible to students, and provided opportunities for students to more accurately demonstrate what they know and can do. It would be a waste in development monies for the AA-MAS to simply disappear. Many of the accommodations in format and wording should be phased into the regular assessment in the spirit of universal design for assessment. Furthermore, one fundamental advantage of designing tests with students with disabilities in mind from the beginning is that tests will not have to be retrofitted for those students after the fact.

4. Practitioners (teachers and school administrators) are still unclear, after several years, about the arithmetic involved in the 1% and 2% options for calculating AYP.

Confusion about the “2% cap” abounds! Even repeated statewide in-service preparation did not make administrators confident in what to tell teachers, stakeholders, or advise IEP team members. Even with fairly intensive efforts, some schools did not follow AA-MAS test assignment guidelines as the state had intended. Some schools did not involve IEP teams in the decision-making as the USDE had intended. Increased and improved training and an emphasis on adhering to guidelines might have been helpful. Substantial training is necessary.

5. Non-aggregated results would paint a clearer picture of the achievement of students with disabilities.

One of the greatest concerns of the Pennsylvania GSEG with the onset of the AA-MAS was the resulting increased complexity of interpreting student performance data. Using the same performance descriptors (i.e., Advanced, Proficient, Basic, and Below Basic) across assessments, despite significant differences in achievement standards and cognitive complexity, misrepresents student achievement and hinders schools’ ability to clearly and honestly discuss student competency of grade-level material. Following the guidance of Phillips (2011), we believe that a better approach may be to report results on multiple assessment options separately. Perhaps unique performance level descriptors would help reduce stakeholder confusion as to what “proficiency” of grade level content really means on the general assessment vs. alternate assessment options. Acknowledging that there are different expectations for some (but not all) students with disabilities and giving credit to students and teachers when these expectations are met appears to be a more rational, honest approach to accountability. Moreover, non-aggregated result reporting would better reflect “who” participated in each assessment option and “how” they did considering that students in Pennsylvania can participate in the state assessment in a variety of ways including taking one or more subtests of the general assessment and modified assessment with or without accommodations. For example, one student with disabilities might take the PSSA-M Math without accommodations, the PSSA Reading with accommodations and

the PSSA-M Science with accommodations, but his same grade peer with disabilities takes all PSSA subtests without accommodations and yet another peer takes all PSSA-M subtests with accommodations.

6. Moving forward requires states to look back. As the federal government moves toward the next generation of assessments, the Pennsylvania Department of Education will further examine the performance analyses and trends of students on the PSSA-M as presented herein to determine how these data can be used to improve future test design and administration of the general assessment for all students including those with disabilities. The PSSA-M was last offered in Pennsylvania in the Spring of 2012. Pennsylvania was keenly aware that although NCLB has not been reauthorized, language related to the AA-MAS or the 2% has not been included in draft revisions of the law. Additionally, many states have joined Race to the Top Assessment Consortia, where “next generation assessments” are being universally designed and made accessible for all students. Pennsylvania recognized that the landscape of state-wide assessment and accountability had changed, and therefore, the state had to change. The state will move forward, like other states with an AA-MAS, by reflecting on what the test results mean for future assessment design and related educational reform initiatives such as measuring student growth on summative assessments and adopting student-achievement-based teacher evaluation systems. As this paper reports, although student achievement did not change enough in one year to impact proficiency status, growth indeed occurred for many students—a finding that should heavily influence future test development for low achieving students with disabilities. Therefore, in depth analysis of the movement of students taking the PSSA-M across (and within) performance categories will continue.

The PDE’s, Bureau of Special Education is working diligently to improve educational performance of students with disabilities on its general assessment. Despite the myriad of research on the AA-MAS target population and our lessons learned about test design that will be applied to the next generation of assessments, a lingering question remains. *If collective state AA-MAS results show that a modified assessment and modified academic achievement standards did not impact AYP, how will placing students with disabilities into the general assessment with only test enhancements make a difference in their achievement?* Pennsylvania is hopeful that many of the on-going, federal educational initiatives will answer that question and converge into a coherent accountability system attuned to the needs of students with IEPs.

In closing, it should go without saying that the achievement of students with disabilities is a critical part of state accountability systems. AA-MAS policy brought the achievement of “the kids in the gap” to the forefront of education reform—a good thing! Response to this assessment option was not without its challenges or concerns. However, the work was not without merit and has taught us much about what the field should do (and not do) to enhance assessment experiences of students with disabilities. Given what we learned about the depressed achieve-

ment of high performing students on the PSSA-M who moved back into the general PSSA, it is important for test developers to recognize the value of improved item accessibility to capture students' skill competence. Much work on improving accessibility and applying principles of Universal Design for Learning (UDL) to general assessments evolved from research related to the AA-MAS (Hall et al., 2011; Kettler et al., 2011). Abandoning the AA-MAS does not mean abandoning that knowledge—applying what we've learned about item enhancements on “modified” assessments is essential when designing the next generation of general assessments. As we move forward, we urge policy makers to consider our lessons learned and give states time to learn more. As with the implementation of many new policies, time is of the essence. States' research and development around the AA-MAS skyrocketed ahead almost as quickly as they now retreat from those efforts and pilot universal assessments on the Common Core Standards. Nonetheless, taking time to collect empirical data and conduct rigorous evaluations of outcomes for students with and without disabilities in initial phases of test design and implementation is important. Moreover, it is important to conduct thorough analyses of changes in student achievement across and within each performance category over time to inform growth analyses and better capture discrete, yet significant, changes in achievement trajectories of students with disabilities. Perhaps the assessments could be implemented for multiple years prior to having them count under the accountability system to give school systems, parents, and practitioners time to fully understand changes to the assessment and accountability system and their impact on the educational lives of children with IEPs. It would also give federal and state education agencies time to consider the impact of the intended and unintended consequences of these changes to the school achievement and post-school outcomes for children with IEPs in our charge.

References

Bechard, S., & Godin, K. (2007). Identifying and describing students in the gaps in large-scale assessment systems. In New England Compact (Ed.), *Reaching students in the gap: A study of assessment gaps, students in those gaps, and assessment alternatives to lessen the gap*. Newton, MA: Education Development Center, Inc.

Elliott, S. N., Kettler, R. J., & Roach, A. T. (2008). Alternate assessments of modified achievement standards: More accessible and less difficult tests to advance assessment practices? *Journal of Disability Policy Studies, 19*(3), 140–152.

Hall, T., Arathoon, I., Kloo, A., & Taub, D. (2011). Assessment of learning in universal design for learning. A presentation at the annual Council for Exceptional Children Conference, National Harbor, MD, April.

Individuals with Disabilities Education Act (IDEA) Amendments 1997. Pub. L. No. 105–117, §1–619, 111 Stats. 3–1047.

Individuals with Disabilities Education Improvement Act (IDEIA) 2004. Pub. L. No. 108–446, § 1–306, 118 Stats. 2647–2808.

Kettler, R. J., Rodriguez, M. C., Bolt, D. M., Elliott, S. N., Beddow, P. A., & Kurz, A. (2011). Modified multiple-choice items for alternate assessments: Reliability, difficulty, and differential boost. *Applied Measurement in Education, 24*(3), 210–234.

Lemons, C. J., Kloo, A., Zigmond, N., Fulmer, D., & Lupp, L. (2012). Implementing an alternate assessment based on modified academic achievement standards: When policy meets practice. *International Journal of Disability, Development, and Education, 59*(1), 67–79.

Lemons, C. J., Kloo, A., & Zigmond, N. (2011). Implementing modified achievement tests: Questions, challenges, pretending, and potential negative consequences. In S. N. Elliott, R. J. Kettler, P. A. Beddow, & A. Kurz (Eds.), *Handbook of accessible achievement tests for all students: Bridging the gaps between research, practice, and policy* (pp. 295–317). New York, NY: Springer.

Marion, S. (2007). *A technical design and documentation workbook for assessments based on modified achievement standards*. Minneapolis, MN: National Center on Educational Outcomes. Retrieved April 2, 2012, from <http://www.cehd.umn.edu/nceo/Teleconferences/AAMASteleconferences/AAMASworkbook.pdf>

No Child Left Behind (NCLB) Act 2001. Pub. L. No. 107–110, § 1–1076, 115 Stat. 1425–2094.

Pennsylvania Department of Education. (2010a). *2010–2011 Assessment handbook*. Harrisburg, PA: Pennsylvania Department of Education.

Pennsylvania Department of Education. (2010b). *Guidelines for IEP teams: Assigning students with IEPs to state tests (ASIST)*. Harrisburg, PA: Pennsylvania Department of Education.

Pennsylvania Department of Education. (2010c). *PSSA-M: Assigning students to the test and understanding participation*. Retrieved May 30, 2012, from [http://www.portal.state.pa.us/portal/server.pt/community/pennsylvania_system_of_school_assessment_\(pssa\)/8757/resource_materials/507610](http://www.portal.state.pa.us/portal/server.pt/community/pennsylvania_system_of_school_assessment_(pssa)/8757/resource_materials/507610)

Phillips, S. E. (2011). U.S. legal issues in educational testing of special populations. In S. N. Elliott, R. J. Kettler, P. A. Beddow, & A. Kurz (Eds.), *Handbook of accessible achievement tests for all students: Bridging the gaps between research, practice, and policy* (pp. 33–67). New York: Springer.

U.S. Department of Education. (2007). *Modified academic achievement standards: Non-regulatory guidance*. Washington, DC: Author. Retrieved May 22, 2012, from <http://www.ed.gov/policy/speced/guid/nclb/twopercent.doc>

Zigmond, N., & Kloo, A. (2009). The “two percent students”: Considerations and consequences of eligibility decisions. *Peabody Journal of Education*, 84(4), 478–495.