Chapter 2

Understanding Low-Performing Students with Disabilities and Their Barriers to Success on Traditional Assessments: A Southern Tale

Marianne Perie
National Center for the Improvement of Educational Assessment

Melissa Fincher
Georgia Department of Education

John Payne
South Carolina Department of Education

Suzanne Swaffield
South Carolina Department of Education

This manuscript was supported, in part, by the U.S. Department of Education Office of Elementary and Secondary Education (Grant Nos. H373X070007 and H373X070009). However, the opinions expressed do not necessarily reflect the position or policy of the U.S. Department of Education and no official endorsement should be inferred.

Overview

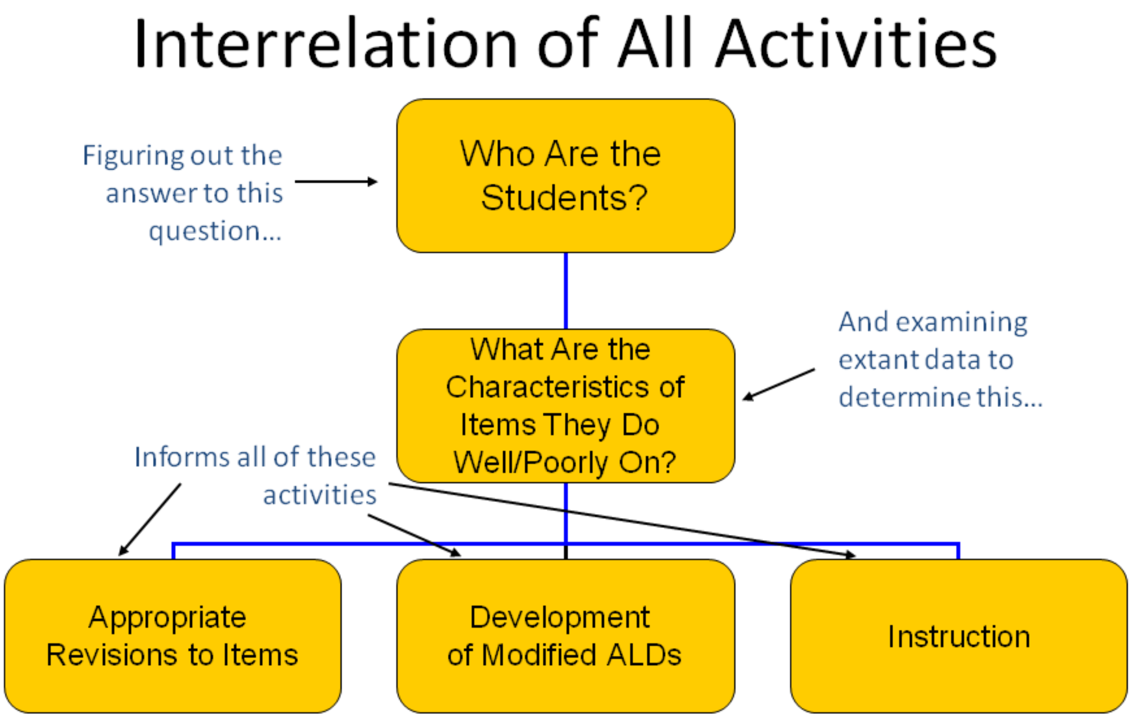
When the General Supervision Enhancement Grants (GSEGs) were first awarded there was much discussion across the states about who the students were that these “2%” assessments were supposed to serve. According to the U. S. Department of Education, “there is a small group of students whose disability precludes 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” (USED, 2007). Among the state policymakers, there was general agreement that there was a gap between the general assessment and the AA-AAS and students who fell into that gap. However, there was disagreement as to whether this new AA-MAS was supposed to serve students at the high end of the AA-AAS or the low end of the general assessment, as many believed those specifications characterized different students. And, in some states, the biggest driver was accountability. These states were interested in developing an assessment that would help students with disabilities who were performing just below the proficient level attain proficiency.

There was, however, a group of states whose policymakers felt that the best parameters for defining students who would be well served by an AA-MAS where those students with disabilities who consistently scored at the lowest level of the general assessment. We refer to these students as “persistent low performers.” In this chapter, we will compare the efforts of Georgia and South Carolina in identifying the persistently low performing students and determining characteristics associated with learning and demonstrating their knowledge and skills. The chapter will then explore how this understanding of the student population drove other aspects of the project from supporting IEP team decisions, to modifying items to better suit their needs, to developing descriptors of what proficiency might look like for this population. (See Figure 1 for a graphic of how all the activities fit together.)

Although the Georgia and South Carolina Departments of Education received separate grants, they were both supported by the National Center for the Improvement of Educational Assessment and information was continually shared across the two states. Therefore, there are many parallelisms to the work that lend themselves to a common analysis and report.

Both Georgia and South Carolina began their projects by attempting to gain a better understanding of the characteristics of students who were not succeeding on the general assessment and then focused specifically on students with disabilities. The work in Georgia actually began earlier, under an Enhanced Assessment Grant (EAG), where most of the work defining the student population occurred. The GSEG covered the work on further item modifications based on lessons learned in the EAG, pilot testing the items, conducting cognitive laboratories, and developing performance level descriptors

Figure 1. How the Student Characteristics Drive All Aspects of Developing an AA-MAS



Georgia started by examining the population, understanding the characteristics of items that these students struggled with, modifying these items and testing the revisions and enhancements. They administered a new AA-MAS in spring 2011. Although the South Carolina Department of Education (SCDE) stopped short of developing a full AA-MAS, they followed many of the same procedures as Georgia, defining the population, identifying and analyzing problematic items, modifying items and conducting small-scale tryouts, and developing guidelines for IEP teams for identifying this population. In addition, they spent time analyzing the individualized education programs (IEPs), providing guidance for improving them, and conducing training seminars for their teachers of students with disabilities. Starting with a description of the process for understanding the students, each section of this chapter will demonstrate how the knowledge of student characteristics drove each of these steps in test design and professional development.

Defining the Student Population

Both Georgia and South Carolina agreed that the target population for the AA-MAS was students with disabilities who performed at the lowest level of achievement on the general assessment for at least two consecutive years. Because the general assessments in the two states differed in terms of the number of performance levels and the number of years the current version had been operational, the specific parameters for defining the two populations were different. However, they both began with similar data mining strategies followed by a closer look at the characteristics of the persistently low performing students.

Statistical Analyses

Both states began their data mining by examining student performance on the general assessment. Georgia was able to use data from the Criterion-Referenced Competency Tests (CRCT) for the years 2004 through 2006. In those years, Georgia reported data out in three performance levels. The initial analysis identified students who scored in the lowest performance level all three years. To narrow the scope, the researchers focused on students in grades 5 and 8.

Georgia's results identified 3% of the student population at grade 5 reading, 4% at grade 4 mathematics, 4% at grade 8 reading, and 9% at grade 8 mathematics. The numbers ranged from 4,016 persistent low performers in grade 5 reading to 12,252 persistent low performers in grade 8 mathematics. As a reference, Georgia averages approximately 125,000 students per grade. Interestingly, only half of the students identified as persistent low performers had documented disabilities; the proportions ranged from 40% to 55% across the different grades and subjects.

South Carolina mined data from their Palmetto Achievement Challenge Tests (PACT) for the two years of 2007 and 2008. The test had undergone revisions between 2006 and 2007 making any earlier comparisons unadvisable. Results from the PACT were reported in four achievement levels, and the initial analysis focused on identifying students who scored in the lowest performance level both years. Researchers in South Carolina chose to focus on grades 4 and 7.

Initially, the numbers in South Carolina were higher than those in Georgia. Eleven percent of students were identified as persistent low performers in grade 4 ELA. In the other grades and subjects the numbers were 14% at grade 4 mathematics, 22% at grade 7 ELA, and 16% at grade 7 mathematics. The numbers ranged from 5,470 in grade 4 ELA to 10,979 in grade 7 ELA. South Carolina has a smaller population than Georgia and averages about 53,000 students per grade.

Because those numbers were too high when considering a test that was intended for approximately 2% of the population, the researchers looked at the accountability report, which further subdivided the lowest performance level into Below Basic 1 and Below Basic 2. The cut score for Below Basic 2 was set at two standard errors of measurement below the Basic cut score making Below Basic 1 the lowest category. When the analyses were redone focusing on the students scoring at Below Basic 1 two years in a row, the numbers were more comparable to Georgia's. Table 1 shows the final numbers used to describe the targeted population.

Table 1. Persistent Low Performers in South Carolina

Content Area	SWD		Non-SWD		Total	
	N	%	N	%	N	%
Gr. 4 ELA	1,398	2.9	801	1.6	2,199	4.5
Gr. 4 Math	1,602	3.3	941	1.9	2,543	5.2
Gr. 7 ELA	1,794	3.7	766	1.6	2,560	5.3
Gr. 7 Math	1,482	3.0	778	1.6	2,260	4.6

Once the student population had been identified, both states examined the group of persistent low performers by demographic. Table 2 shows that there was a greater proportion of male, black, low socioeconomic, and limited English proficiency students identified than expected given their proportions in the general population.

Table 2. Characteristics of Persistent Low Performers

	Georgia		South Carolina	
Characteristic	Base-line	Identified	Base-line	Identified
Male	51%	60-65%	51%	70-74%
Black	40%	60-65%	37%	59-60%
Free/reduced lunch	50%	75-80%	55%	82%
ELL	5%	7-12%	4%	5-8%

Next, each state examined the persistent low performers with disabilities—referred to as the target population—by primary disability. Overall, the most prevalent types of disability in this target group appear to be mild intellectual disabilities and specific learning disabilities. In Georgia, approximately 10% of students in the general population have mild intellectual disabilities, but up to 30% of the target population have that listed as their primary disability. In addition, around 33% of the target population with a disability has a specific learning disability listed as their primary disability. In South Carolina between 16-21% of the target population listed mild mental disabilities as the primary disability, and approximately 59% had a learning disability listed as the primary disability.

Focus Groups

Consultants working with the Georgia Department of Education ran focus groups with Georgia teachers. The facilitators asked the following questions of the teachers:

- Who are these students? How would you describe them, particularly in the context of when they read or do mathematics in your classroom?
- What are their greatest struggles in reading and mathematics?

- What instructional strategies seem to best support their engagement and learning in reading or mathematics?

The teachers described the target population as passive learners, who were hesitant to take risks. In addition, they struggled with generalizing skills and concepts to new situations or problems, making connections, and changing topics. With reading, they had limited vocabulary and poor decoding, fluency, and comprehension skills. In math, the students demonstrated poor number sense and difficulty accessing and applying strategies.

In discussing effective instructional strategies, the teachers mentioned using guided practice and incorporatng visuals and tools such as number lines, place value charts, graphic organizers and multiple representations, some of which were concrete. They described grouping, chunking or summarizing information, and previewing questions prior to reading a passage.

Consultants with the South Carolina Department of Education also ran two focus groups: one with teachers and one with administrators. The teacher group was comprised of both special education and general education teachers and the administrators were primarily special education coordinators with one assistant principal. They started with a list of typical behaviors identified by OSEP and other states and narrowed it down to what they saw as the most salient characteristics of the target population. These characteristics included:

- limited vocabulary compared to same age peers,
- limited working memory,
- limited meta-cognition,
- difficulty with problems requiring multi-step solutions, and
- difficulty understanding the intended meaning of texts.

Other behaviors identified by South Carolina educators included students who are easily distracted, have trouble organizing work and completing assignments, and need frequent one-on-one assistance. The teachers also discussed specific issues seen in the classroom. For instance, when a student made a mistake working a problem, the teacher often struggled to correct the misconception because the student could not explain how he/she had arrived at the solution. The student would also understand a strategy when working with a teacher but then could not apply it later when working independently.

A second focus group identified priorities for instructing persistently low performing students with disabilities, including common techniques such as formative assessment, selection of effective accommodations, and inclusion in general education classes. They also thought it was very important for teachers to have a deep understanding of how content progresses across grades, to understand how to prioritize skills to be taught, and to know when to move on when a student is “stuck” but could benefit from other concepts.

The focus group from both states identified similar issues. Limited vocabulary and poor comprehension were mentioned by both. The South Carolina educators discussed limited metacognition, specifically noting the inability to explain thinking or to apply what they had learned when working independently, while the Georgia teachers listed difficulties generalizing, making connections, and accessing and applying strategies. Both groups of educators also noted difficulties in math. Many of these characteristics lend themselves to different strategies for teaching the students and could be applied to a new assessment as well.

Surveys

In addition to collecting data from focus groups, the South Carolina Department of Education (SCDE) also contracted with the University of South Carolina (USC) to develop and administer a survey to better understand persistent low performers with disabilities. The survey was adapted from a special educator survey developed in Oklahoma by Nagle, Cameto, Almond, and Morrison (2006). Staff from SCDE and USC modified, deleted, and added questions to the survey to make it more specific to South Carolina and the topics of interest. A total of 1,541 educators responded to the survey, although the number of answers to any one question varied from 644 to 1,534. Approximately 69% of the educators responding indicated that they had students in their classroom who fit the definition of students for whom the AA-MAS was intended; 84% indicated they taught students with disabilities who would take the general assessment and 13% indicated that they taught students who would take the AA-AAS.

When focusing on students with disabilities eligible to take the AA-MAS, the teachers most often identified students with a specific learning disability, followed distantly by intellectual disability and speech/language disability. In considering a student they had in class who met the criteria for participation in AA-MAS, over 65% of teachers noted the following behaviors occurring at least weekly:

- Reads slowly.
- Has difficult drawing inferences from grade-level text.
- Has difficulty answering comprehension questions on long passages.
- Requires frequent clarification of instructions and one-on-one support.
- Has difficulty identifying the main idea of grade-level texts.

- Is easily distracted.
- Has trouble with organization and keeping track of work.
- Has difficulty with problems requiring multi-step solutions.

One distinct finding was that these students were not students with noted behavior problems. Teachers were much less likely to report these students as having troubling following rules, considering the consequences of their behavior, or fidgeting.

Identifying Barriers to Success on the General Assessment

Each state used statistical analyses to identify items that appeared problematic for persistent low performers and then brought in content expects to review the characteristics of the items. A goal of this set of studies was to tie the understanding of student characteristics to item characteristics to learn more about barriers on a general assessment that may prevent some students from showing what they know and can do.

Statistical Identification of Items

Both states also engaged in a data mining process to identify items on the general assessment with which persistent low performers struggled. Different statistical analyses were used to identify problematic items, but an important finding in both states was that more than half of the items on the general assessment performed well, even for persistent low performers with disabilities. That is, the items seemed to distinguish between students who knew the material and those who did not, they correlated well with total score, and they did not show any bias against students with disabilities or persistent low performers.

Georgia

Georgia wanted to identify both items that were problematic for the target students and those on which they performed well. By examining both types of items, they hoped to learn what characteristics of items intersected with student characteristics to help determine which type of modifications were more likely to provide better access to the target population.

To identify items that seemed to function well or poorly for the target population, Georgia researchers ran both classical statistical analyses and item response theory (IRT) on all items administered in 2004, 2005, and 2006 at grades 5 and 8 in reading and mathematics. First, the researchers started with the difficulty index (p-value, or percent correct). Traditionally, acceptable items are expected to have difficulty values ranging from 0.35 to 0.95. Second, they looked at discrimination, which is the correlation between the item score (0/1) and the total score. Items should show positive discrimination, meaning that a correct score on an item is correlated with a higher total score. The third factor examined was the point biserial for the distractors. One would not expect to see a particular wrong answer choice to be more highly correlated with the total score than the correct choice. Those three statistics are used in traditional item analyses. When the items were re-analyzed focusing just on the persistent low performers, any items with results that did not conform to traditional item statistics were flagged as potentially problematic, while items meeting the traditional criteria ($0.35 > p < 0.95$ & $r > 0$) were classified as potentially effective.

The second set of analyses was based on IRT statistics. First, because persistent low performers are unlikely to answer correctly items that map to the proficient or advanced levels, item parameters from the general population were used to identify items located below the proficiency cut. Then, item parameters for those items were recalculated using only the results of the persistent low performers. Traditionally, acceptable items are expected to have positive discrimination indices and difficulties falling between -4 and +4. Therefore, using the new calculations, items with a negative discrimination or with a difficulty parameter that exceeded +/- 4 were considered ineffective and flagged as particularly problematic for this population.

Focusing first on effective items, the Georgia researchers looked at the IRT parameters and classical parameters simultaneously. Those that were borderline problematic on one statistic but effective on another were included in the final pool of effective items. As shown in Table 3, of the 70 items in grades 5 and 8 math, 30 and 38, respectively, were deemed effective. Reading had 40 items in grades 5 and 8, and 14 and 15, that respectively, were determined to be effective in the final analysis. While, at first glance, the number of items may appear small, it is important to remember that these results imply that 35 to 54 percent of items function well for the target population. That means that wholesale changes to the entire assessment are unnecessary. Rather, making specific modifications to individual items that are deemed problematic is the better approach.

Table 3. A Summary of Effective Items

Statistical Criteria	Number of Test Items	IRT Parameters	Classical Parameters	Final Pool of Effective Items
Grade 5 Math	70	12	22	30
Grade 8 Math	70	26	16	38
Grade 5 Reading	40	9	14	14
Grade 8 Reading	40	12	6	15

The researchers next looked at the balance of representation of items for the total test and final pool of effective items. Looking at the percentage of items in each content strand, the results were fairly consistent in both pools of items at grade 8. However, at grade 5, problem solving and statistics were slightly under represented and computation and estimation over represented in the Grade 5 math final pool of effective items. This seemed to indicate that persistent low performers in Grade 5 are better able to answer the computation and estimation questions but have difficulty with the problem solving and statistics and probability questions. In reading, the target students appear to have more success with the items requiring them to read for information and less success with the items that involve reading for comprehension than would be expected given the balance of representation on the total test.

Focusing next on ineffective items, they found that distractor analyses identified more items as ineffective in math, while the IRT statistics identified more items as problematic in reading (see Table 4). Again the next analysis examined the proportion of items by content domain.

Table 4. A Summary of Ineffective Items

Statistical Criteria	Number of Test Items	Distractor Proportions	Distractor Point Biserial	IRT parameters	Final Pool of Ineffective Items
Grade 5 Math	70	17	3	11	25
Grade 8 Math	70	16	6	9	21
Grade 5 Reading	40	8	0	20	22
Grade 8 Reading	40	1	0	13	14

Seventy items were identified as ineffective: 20 in grade 5 mathematics, 20 in grade 8 mathematics, 17 in grade 5 reading, and 13 in grade 8 reading. Overall, the balance of representation of the ineffective items appears to approximately parallel that of the total test, with the exception of Grade 8 mathematics. In this latter case, it appears that the persistently low performing students have difficulty with problem solving, as eight of the fourteen problem solving items were flagged as ineffective.

South Carolina

In South Carolina, researchers focused on differential item functioning (DIF) analyses. DIF occurs when people from different population groups with the same abilities have a different probability of answering an item correctly. In this case, if student ability (measured by test score) is held constant, we would expect the same percentage of students from each subgroup to answer the items correctly. The analysis was therefore run for all students who scored in the Below Basic 1 performance category on the general assessment, and the performance of students with disabilities was compared to those with no identified disability to see if there were any items that showed significant differences. For this study, the focus was on grades 4 and 7 mathematics and ELA. Researchers used the Cochran-Mantel-Haenszel (CMH), Fisher's exact test, logistic regression, and one-, two-, and three-parameter logistic item response theory (IRT)-based procedures.

The 10 items with the most difference between the two groups were identified at each grade level. While the majority of these items were classified by SMH as displaying non-significant DIF, the Fisher's exact test results identified most as statistically significant. In addition, items on which students with disabilities outperformed their non-disabled peers were identified and patterns of response options were provided for analysis.

Content Review of Items

After a set of items had been identified that were problematic for persistent low performers with disabilities, groups of special educators and content experts were convened to determine if any commonalities across items could be identified. Much time was spent in these meetings discussing the target population and what had been learned from examining test scores, holding focus groups, and surveying teachers about their learning characteristics. In these meetings, teachers identified problematic features of items and recommended revisions and enhancements to the items.

Georgia

Seventy items were identified as problematic based on the statistical analysis: 20 in grade 5 mathematics, 20 in grade 8 mathematics, 17 in grade 5 reading, and 13 in grade 8 reading. The research team began the process of identifying the potential barriers to access in each item by conducting a literature review to identify potential factors of presenting text and assessing students with characteristics similar to those determined by the study of the characteristics of students eligible for the AA-MAS. Specifically, they focused on students with learning disabilities in reading and students with mild intellectual disabilities. Table 5 presents the findings from the literature review that guided all modifications made to items and tests.

Georgia convened a committee of 26 educators—13 each in reading and math. Committees were comprised primarily of classroom teachers, with some curriculum experts, administrators, and educators with special expertise in understanding the diverse needs of the student population struggling to meet expectations.

The committee reviewed the items focusing on the following questions:

- What skills and concepts are being tested?
- What is the focus of each item: application of skills, fact-based information, conceptual understanding, or problem solving *and* what might make this difficult?

- What is the vocabulary load of each item and the overall readability and text "density" for each item?

- Do the charts, tables, graphs, artwork, visuals, or spacing used support *or* perhaps detract from understanding?

- What is the "closeness" of distractors to each other?

- What is the complexity or abstractness of ideas presented (e.g., use of figurative language vs. literal meanings; theme vs. main idea)?

For the mathematics items, it was also noted whether they were single-step or multi-step problems and whether they required "extensive" reading. For reading items, the genre, text features, text structure, readability, and length of each reading passage were also examined. Committee members also indicated whether or not the item was a candidate for revision.

Table 5. Using a Research Base for Revising and Enhancing Test Items and Test Formats

Factors to consider	**What the research literature says…**
Use of pronouns	Sentences with excessive pronoun use may cause a student to lose track of the main point of reference in an item.
Use of negatives	Difficulty of text may vary due to "complex Boolean expressions." Such expressions are challenging because "the respondent needs to keep track of different options and possibilities." In the case of negative expressions, an unnecessarily high cognitive loading may be added to items that employ negatives within items (e.g., "Which of the following is not a reason why the captain wanted to turn the ship around?").
Vocabulary load	There is an inverse correlation between the level of vocabulary in text and its readability. (For purposes of the study, unless the construct of an item was to test a particular vocabulary skill, it seemed reasonable that vocabulary demands could be reduced as a principle of access.) Vocabulary load can be a significant factor in accessibility of test items (Johnstone, Liu, Altman, & Thurlow, 2007).
Non-construct subject area language (specialized vocabulary)	Each content area has a specialized vocabulary of its own (e.g., characterization, denouement, alliteration). When these words are part of the intended construct, it is appropriate to include them; however, when these terms are extraneous to the intended construct, they may introduce "nuisance variables" (Haladyna & Downing, 2004).
3 "Tiers" of Vocabulary	Beck, McKeown, and Kucan (2002) identify 3 tiers of a literate individual's vocabulary. Tier 1 includes most basic words (e.g., clock, baby) that rarely require direct instruction. Tier 2 includes high frequency words used by most "mature language users" found across most domains (e.g., absurd, fortunate, anxious). These words often have multiple or nuanced meanings—intended meaning that directly impacts verbal functioning and comprehension. Tier 3 words include low-frequency, domain-specific words (e.g., land forms, classifications of plants). These words are learned for a specific purpose or need.

Table 5. Using a Research Base for Revising and Enhancing Test Items and Test Formats (continued)

Complex sentences, dense text	Sentences with "dense clauses" are sentences that "pack too many constituents or idea units (i.e., propositions) within a single clause"…and/or with "dense noun phrases" as sentences with "too many adjectives and adverbs modifying the head noun." Items with either one of the above sentence types may be problematic to reading comprehension (Rand, p. 96).
Text Structure	There are text structures that are easier to understand and use to organize and recall information (sequence, chronology, enumeration/description, definition, compare-contrast). More complex text structures require understanding of entire texts, such as cause-effect, problem-solution, and proposition-support (Hess, 2008).
Graphic organizers	Graphic organizers are effective instructional supports used during a lesson to assist students in understanding such things as the text structure (e.g., story map for narrative, timeline for chronology, Venn diagram for compare-contrast) (Schumm, 2006). The layout or format of a table, graph, or graphic organizer visually organizes information for conceptual understanding (Robb, Richek, & Spandel, 2002).
Chunking text	"Chunking" of texts is an effective instructional support used to assist students in conceptualizing ideas presented in longer texts, reducing the demand on working memory. Research related to the use of chunking has focused on chunking parts of sentences (phrases), rather than on chunking full texts. Overall comprehension may not be significantly improved by dividing sentences into smaller parts; however, students with reading disabilities have identified a preference for shorter chunks while reading (Dickson, Simmons, & Kame'enui, 1995).
Segmenting text	"Segmenting" text is a term generally used as it applies to breaking text into subparts to reduce the length and visual appearance of the reading load. Segmented texts group segments with their corresponding test items. Recent studies have examined the impact of segmenting on students' non-cognitive domains such as anxiety, fatigue, frustration, and motivation as well as on their comprehension. Results suggest segmented text may be more accessible to students with disabilities; building in short "breaks" for the reader may also reduce the need for added accommodations (Abedi, Leo, & Kao, 2008).
Spacing (the amount of space between each character)	Letters that are too close together are difficult for partially sighted readers. Spacing needs to be wide between both letters and words (Gaster & Clark, 1995). And we assume also for numbers and symbols.
Typeface (characters, punctuation, and symbols that share a common design)	Italic is far less legible and is read considerably more slowly than regular lower case (Worden, 1991). Boldface is more visible than lower case if a change from the norm is needed (Hartley, 1985).

Table 5. Using a Research Base for Revising and Enhancing Test Items and Test Formats (continued)

Text Justification (text is either flush with left or right margins [justified] or staggered/ragged [unjustified])	Staggered right margins are easier to see and scan than uniform or block style right justified margins (Arditi, 1999; Grise et al., 1982; Menlove & Hammond, 1998). Justified text is more difficult to read than unjustified text—especially for poor readers (Gregory & Poulton, 1970; Zachrisson, 1965). Justified text is also more disruptive for good readers (Muncer, Gorman, Gorman, & Bibel, 1986). A flush left/ragged right margin is the most effective format for text memory. (Thompson, 1991). Unjustified text may be easier for poorer readers to understand because the uneven eye movements created in justified text can interrupt reading (Gregory & Poulton, 1970; Hartley, 1985; Muncer, Gorman, Gorman, & Bibel, 1986; Schriver, 1997). Justified lines require the distances between words to be varied. In very narrow columns, not only are there extra wide spaces between words, but also between letters within the words (Gregory & Poulton, 1970).
Line Length (length of the line of text; the distance between the left and right margin)	Lines that are too long make readers weary and may also cause difficulty in locating the beginning of the next line, causing readers to lose their place (Schriver, 1997; Tinker, 1963).
Blank Space (Space on a page that is not occupied by text or graphics)	Blank space around paragraphs and between columns of type helps increase legibility (Smith & McCombs, 1971) A general rule is to allow text to occupy only about half of a page (Tinker, 1963). Too many test items per page can make items difficult to read.

Source: Hess, McDivitt, and Fincher, 2008. For the complete citations for the references included in this table, see Hess, McDivitt, and Fincher.

South Carolina

Georgia conducted their meeting over a year earlier than South Carolina, so the team in South Carolina was able to learn from the research done in Georgia. In addition, they used the Test Accessibility and Modification Inventory (TAMI) established by Beddow, Kettler, and Elliott (2008). South Carolina assembled a review committee of general and special educators. Eighteen participants were divided into four groups to examine the items from the grade and subject area in which they had expertise. They were provided with the 10 items as well as the response patterns and difficulty values and were asked to align each item to a grade level content standard, at the indicator level. They then examined the accessibility, cognitive complexity, and challenges of each item in their subject and grade. Specifically, each committee rated the accessibility level of different aspects of the item (passage, visual, item stem, answer choices), coded cognitive complexity (working memory, metacognition, construct, procedural knowledge, declarative knowledge, and extract[1]), analyzed response patterns and identified reasons why students with disabilities may have chosen an incorrect response more frequently than students without disabilities.

[1] Extract was defined as “decode, form mental representation.”

In conducting a distractor analysis, the committee determined that examinees were drawn to each response option at the chance level. Grade 4 language arts had the most issues with accessibility, but they were linked more to the passage than to the items, and the passage was not open for modification. For grade 4 language arts, working memory (defined as cognitive resource for processing information) and metacognition (defined as thinking about thinking) appear to be the main cognitive aspects that affected whether students correctly responded to an item. The main cognitive aspects for grade 7 included construct and procedural knowledge (defined as how things are done). In mathematics, grade 4 participants coded working memory (defined as cognitive resource for processing information) and construct (defined as understanding the intended meaning of text) as the two main cognitive aspects that could have contributed to students not successfully answering an item. In grade 7 math, the two most often chosen cognitive aspects were procedural knowledge and declarative knowledge (defined as information about how things are). Interestingly, there was little in common in terms of the cognitive challenges across subjects and grades; the issues seemed much more item specific.

Modification of Items

Both Georgia and South Carolina had the same group of educators who identified the barriers in the items also recommend the revisions or enhancements to be made to each item. Both tasks were completed in the same meeting, so there was a strong connection among the understanding of the students, the barriers to access, and the recommended changes to items.

Georgia

Again, with the 70 items identified for alteration based on the statistical analysis—20 in grade 5 mathematics, 20 in grade 8 mathematics, 17 in grade 5 reading and 13 in grade 8 reading—and the problematic features identified in the first stage of the workshop, the Georgia educators next recommended revisions to each item.

In Georgia, the researchers distinguished between "revisions," which involved changing the item to make it more accessible without changing the construct, and "enhancements," which involved adding something to the item to help the student understand the question.

Examples of revisions include:

- Simplify language in question/stem and/or distractors
- Simplify graphics, visuals, etc.
- Eliminate extraneous information
- Substitute another (more familiar) word without changing the construct

- Reformat items or passages (e.g., adding more white space, size of text)
- Adjust layouts (e.g., reorder items or passages)
- Separate reading passages into chunks, followed by related items

Examples of enhancements include:

- Add a Helpful Hint in a “Thought Balloon”
    - Definition, key word or phrase
    - Reminder of approach to help solve a multi-step problem (e.g., circle the information you need to solve this)
- Provide a Scaffold
    - Graphic organizer (e.g., timeline for organizing chronology)
    - Table, graph, chart, or visual to enhance conceptual understanding (e.g., input-output chart for finding pattern)
    - Key words, phrases, definitions (e.g., sidebar glossary), introductions to passages
- Enhance the General Presentation: underline, bold, enlarge key words/phrases/symbols; number paragraphs or lines

The committee was cautioned against making the following types of changes:

- Shortening passages
- Adapting passages in any way that would make them “below grade level”
- Adding more items to fewer passages
- Simply removing a distractor without evidence that it was problematic
- Revising questions or answer options in such a way as to change the nature of the item and/or the construct

Each of the 70 items was revised using the guidance provided. The committee also agreed to shorten the test by approximately 30 percent. However, the blueprint of the AA-MAS maintained the content coverage and the distribution of content, meaning that if the general assessment had

40% of the items assessing numbers and operations in 3rd grade, then the AA-MAS also had 40% of the items assessing numbers and operations in 3rd grade.

South Carolina

In South Carolina, a committee of teachers was convened to focus on grades 4 and 7 ELA and math. Each sub-committee worked to revise 10 items in their grade and content area. They, too, used the literature review findings presented in Table 4, as well as the TAMI to guide the revision and enhancement process.

For grade 4 language arts, participants modified item stems by simplifying vocabulary, shortening the item stem, adding a context clue, or removing extraneous information in seven items with a low accessibility rating. Participants changed the distractors on four of the items rated as having problematic distractors by making them more parallel to one another in format (e.g., all written as full sentences), making the answer choices have similar vocabulary difficulty, or altogether eliminating an answer choice. Additionally, for half of the items, modifications included the bolding of words or phrases in the item stem to draw attention to these key points.

Grade 7 language arts had the fewest items modified. The grade committee thought that increasing spacing on four of the ten items was important, given the text load students were expected to read and cull through. Although only two items were rated below the highest accessibility level in both the item stem and answer choices, this committee made modifications to further improve how the item stem and distractors were understood by bolding key words and phrasing or changing the distractors to be parallel in vocabulary difficulty and format of each answer choice. Looking across both grades of ELA, the two most often used modifications included focusing students' attention on key words or phrases through bolding in the item stem, or changing the distractors to be similar in difficulty and layout.

The fourth-grade mathematics committee coded the visuals in four items, six item stems, and two sets of answer choices as having issues with accessibility. As a result, the committee edited three of the visuals and increased spacing on one item to raise the accessibility of the visuals for these items. To improve the item stems, committee members shortened the item stem, removed extraneous information, or bolded key information or phrases. They also made modifications to the distractors for four items. There were only two items with no recommended changes.

The visuals in grade 7 math were all maximally accessible for most or all students. Nine of the ten item stems and four of the sets of answer choices were coded as maximally accessible for most students. To improve the accessibility of the item stems for all students, the seventh-grade mathematics committee shortened the item stem, bolded key words/phrases, and removed extraneous information. They also edited the visual for one item. Distractors were simplified or changed to better assess common misconceptions or errors.

Across the two math grades, the most common modification was bolding key words and phrases, followed by simplifying and removing irrelevant information in the item stems. Across the two subjects in both grades 4 and 7, most modifications were also performed on the item stems by shortening the item stem by removing extraneous information as well as bolding key words and phrases. See Table 6 for the summary of modifications made to language arts and math items.

Table 6. Types of Item Modifications Recommended by South Carolina Educators

Subject/ Grade Level	Simplified Vocabulary/ Shortened item stem	Bolded key word/ Phrase	Increase spacing	Changed distractors/ Eliminated distractor	Removed extraneous information	Simplified/ Added visual
ELA Grade 4	3	5	0	4	4	1
ELA Grade 7	1*	2	4	3	1	0
Math Grade 4	4	2	1	4	2	3
Math Grade 7	2	6	1	2	4	1

* In grade 7 ELA, the committee recommended adding a context clue for an item stem. Instead of developing a separate category for this one instance, that modification was included in the summary of how teachers modified the item stem.

Modified Item Tryouts

Georgia first piloted the items in a small-scale study in fall 2008 (under the prior Enhanced Assessment Grant [EAG]), followed by a field test in spring 2009. Approximately 5,000 students participated in the pilot study, which included two forms in each grade (5 and 8) and subject (ELA and math) that were counterbalanced with original and modified items. Eight to ten items, previously identified to have reasonable characteristics for both the general population and the persistently low-performing (PLP) population were included on both forms to serve as a link between forms. While the same alterations were successful in both grade content areas, it appears that the changes were more successful in reading than in math. (See Table 7 for the exact numbers of items.)

Table 7. Summary of item changes

	Number of items modified	Modified items classified as effective or problematic	
		Effective	Problematic
Grade 5: Math	20	10	8
Grade 8: Math	20	11	8
Grade 5: Reading	17	13	3
Grade 8: Reading	20	20	0

A few of the item alterations made the items less effective for the PLP population in three of four grade content areas. However, distinguishing which modifications were effective and which were not was problematic as the results were not consistent across items. While most of the changes had a positive impact on the ability of the students in the target population to respond to the item, some of the changes had a negative effect. But, it was often difficult to tease out the feature responsible for the effect. For instance, chunking passages was one modification made. Simultaneously, some of the items associated with those passages were also revised, making it difficult to distinguish the effect of each type of modification. However, while not every item associated with a chunked passage showed improved performance in the target population, none of the items showed worse item statistics. Therefore, at worst, the chunking did no harm, and at best it improved the performance of some students in the target population.

After using their understanding of the characteristics of students eligible for the AA-MAS and modifying items based on that understanding, South Carolina tested the modified items using cognitive labs with retrospective think alouds. They studied the effects of the modifications both on students with disabilities who fit the participation criteria for an AA-MAS as well as on students without any identified disability using both the original items and the modified items.

Overall, the studies showed a very small positive effect of the modifications on total test score for the students with disabilities and either no effect or slightly negative effect on students without disabilities. In follow-up questions, about 23% of students with disabilities found the modifications helpful while more than half (58%) indicated the modified items were still as difficult as the original items. The only changes students with disabilities responded positively to were presenting information visually and bolding key words in the item stem. Students with no identified disabilities also found those changes helpful. Students had more mixed responses to changes like reducing the wordiness of an item or removing a distractor. In some cases, reducing the number of words made the item difficult to understand. An item where the first sentence was meant to personalize and provide context was perceived as easier when that sentence was removed.

As with all cognitive laboratory studies, the sample size was small, making generalizations difficult. However, the only modifications that appeared worth including were adding graphics to present information visually, enlarging graphs when they are key to answering the question, bolding key words, and removing extraneous text meant only to personalize the item.

Using Pilot Results to Better Understand Target Student Characteristics

Georgia first piloted the items in a small-scale study in fall 2008, followed by a larger field test in spring 2009. After the spring 2009 administration, 31 students participated in cognitive labs, allowing test developers to learn more about the modified items.

Approximately 700 students per grade participated in the pilot study, which included two forms that split original and modified items. Researchers used a linear regression of the AA-MAS pilot performance on matched CRCT to determine the difference between predicted and observed performance. They then focused specifically on lower achieving students whose observed performance was higher than the predicted performance as these are the students that the AA-MAS could most benefit. Next, they examined characteristics of this group compared to all students with matched performance and students with matched performance who were low achievers whose observed performance was lower than predicted.

They found no discernible patterns with respect to gender and ethnicity, although as noted earlier males were overrepresented in the full AA-MAS group. In terms of disability status, more students with learning disabilities performed higher than predicted while students with mild intellectual disabilities tended to perform worse than expected on the AA-MAS. However, given the nature of the categories, it can be difficult to interpret this finding and generalize from it.

In spring 2009, a field test was conducted to determine the suitability of two forms that contained the modified items. Based on the results of this study, revisions were made and two forms were created for a calibration sample. In fall 2010, a larger pilot test for grades 3 through 8 in reading, ELA, and math was conducted with approximately 1,000 students per grade. In this study, participation was limited to students with disabilities who did not meet the standard on the spring 2009 general assessment. Because it was a fall administration of a spring assessment, the fourth-grade test was given to incoming fifth-graders, and the fifth-grade test was given to incoming sixth-graders, and so on. The purpose of this study was to provide descriptive and psychometric information about the items and forms to inform test development decisions.

This study resulted in dropping six items from the test. Once those items were dropped, the distributions of score were generally normal or slightly positively skewed across all grades and content areas. Factor analysis showed one dominant component on each form, providing support for an underlying unidimensional scale. The items statistics were in line with typical item statistics, and the reliability of each test form ranged from 0.77 to 0.87, which is very much in line with the reliability of most general assessments. Of the 3,576 items examined across all subjects and grades, only 6 items were classified in category C of the Mantel-Haenszel DIF analysis of white/black students and male/female students. After this pilot, the Georgia Department of Education felt that they had a sufficient number of items in the item bank to build two solid forms ready for operational administration.

In fall 2009, 31 students from 10 districts participated in cognitive laboratories. They were selected based on the criteria listed in the federal regulations as interpreted by their teachers. Students were in grades 4, 7, and 10 taking items from grades 3, 6, and 9, as this was a fall study. They were each given 9–10 items that had appeared on the spring 2009 field test. Items selected for the cognitive labs had showed some anomalous statistical characteristics, such as a very low p-value, a negative point bi-serial correlation, a distractor analysis showing that a distractor was more often chosen than the correct answer, or an item demonstrating bias through the DIF statistic.

Results of the study indicate four themes that were present across all grade levels. First, students' reading fluency was a major factor in their struggles with reading comprehension items. That is, regardless of how the item was modified, students did not have the reading skills to decode the meaning. Second, in many cases, students did not have an opportunity to learn content on the test prior to taking the item. Third, at times a single distractor had a lower vocabulary level than the other three choices in the item, which made it more attractive to students who struggled with content of items. Finally, the helpful hints provided in items were ignored by many students. In some cases, when students struggled with content, students were led to incorrect answers by their interpretation of the hint. Overall, the assessment characteristics appeared to be less of a factor in student incorrect answers than the students' own challenges with reading and content. The only lesson learned that could influence item development was ensuring that the text complexity was parallel across all four answer choices. However, while the researchers could not always explain what caused the improvements in performance or point to specific modifications that were consistently successful or ineffective, there were improvements. Overall, it may be the combination of changes that had the biggest effect on making items more accessible to students who had poor decoding skills, difficulty generalizing, find long passages challenging, and who struggle to apply the best strategy to solve a problem (Johnstone, 2009).

Although South Carolina chose not to develop a new alternate assessment, the SCDE did consider ways to improve items on their general assessment to better support persistent low performers with disabilities.

Defining "Modified Proficiency

Georgia used the information about the characteristics of persistently low performing students with disabilities and information gained from trying out modified performance to develop "modified" performance level descriptors. South Carolina explored options for doing so, commissioning a paper by Dr. Marianne Perie of the Center for Assessment on a methodology for developing performance descriptors for an AA-MAS. However, once the decision was made not to develop an AA-MAS, the work was not continued.

Georgia policymakers felt it was important to begin with the performance level descriptors (PLDs) for the general assessment and modify them based on expectations for students who meet the eligibility criteria for the AA-MAS and on the changes made to the items themselves. The process of writing the modified PLDs occurred in several stages. There were two committees, first drafting the modified PLDs and then reviewing them. In between those two committee meetings, content experts reviewed the modified PLDs and identified areas of inconsistencies either within or between grade levels.

At the first committee meeting the first half-day was spent on training. First, the facilitator reviewed the federal guidelines for the AA-MAS, including statements about who the test was meant to serve, and then reviewed the data collected to date on the grant regarding the eligible population. The facilitator explained the concept of persistent low performers and explained that while only 50-60% of them had disabilities, this test was only meant to serve those with disabilities. She then described the results of the focus groups and had a further discussion about the characteristics of the students. Next, the item analysis was described. The committee was given examples of modifications that were made to make the item more accessible without changing the construct being measured.

In drafting the modified PLDs, the committee was encouraged to consider the characteristics of the students, the modifications made to the items, and the final test blueprint. They were asked to think about how a student learns as they move from novice understanding to mastery of a content area within a grade level as well as how they learn moving from one grade level to the next. They were also asked to consider how the AA-MAS fit between the AA-AAS and the general assessment. Georgia policymakers shared their belief that the AA-MAS may provide a stepping stone for students to reach proficiency on the general assessment

Next, the facilitator shared the PLDs from the general assessment. The committee talked about the “space” between the description for “Does not Meet Standard” and “Meets Standard” and agreed that filling that space was the place to begin considering a description for “modified” proficiency. The committee also agreed to supplement the PLDs with examples of how scaffolding was providing for items on the AA-MAS. They worked in grade pairs for a day and a half to craft modified PLDs. After the meeting, content experts reviewed the PLDs and pointed out areas where articulation across grades was an issue or where abstract terms were used that would be difficult to apply in a standard setting. A second committee was convened to address the concerns raised and finalize the PLDs.

They ultimately agreed to the following names and general definitions for each level:

Performance Level 1: Below Proficiency—The student at the below proficiency level in <subject> does not meet the modified expectation set for students in X grade.

Performance Level 2: Emerging Proficiency—The student at the emerging proficiency level in <subject> meets the modified expectation set for students in X grade.

Performance Level 3: Basic Proficiency—The student at the basic proficiency level in <subject> surpasses the modified expectation set for students in X grade.

An example of a full PLD is provided in Figure 2.

Current Status in Georgia and South Carolina

Georgia currently has an AA-MAS administered operationally, starting in 2011. While the 2012 results were not available at the time of this writing, the 2011 results are presented in Tables 8 and 9. Overall fewer than 2% of students took the reading and ELA assessments, while just over 2% of students took the CRCT-M in mathematics. Students scoring at or above the proficiency cut ranged from 64% in ELA to 69% in mathematics. Given that the target student had been performing at the lowest level for three years in a row, these results indicate that something is working. The Georgia Department of Education plans to continue to use the CRCT-M until the consortia assessments become operational in 2015 or until legislation is passed prohibiting their use, whichever comes first.

Figure 2. Sample PLD for the Georgia CRCT-M in Grade 3 Reading

Level 1:

The student at the below proficiency level in reading does not meet the modified expectation set for students in the third grade. When scaffolding and supports are provided (e.g., thought bubbles, segmented texts, simplified language, graphic organizers, and timelines), students performing at this level demonstrate minimal ability to infer, draw conclusions, and make judgments about grade-level literary and informational texts. Students may be able to understand and acquire some new vocabulary. They may be unable to isolate root words from affixes and then define them. They may be able to make obvious connections between the text and their own experiences but make minimal connections beyond their own experience. Their interpretation of graphics is minimal.

Level 2:

The student at the emerging proficiency level in reading meets the modified expectation set for students in the third grade. Students performing at this level demonstrate understanding of a variety of grade-level texts when scaffolding and supports are provided (e.g., thought bubbles, segmented texts, simplified language, graphic organizers, and timelines). Their general facility with literary texts may exceed their competence with informational text at this performance level. Students use supports provided to understand new vocabulary and are able to isolate root words from affixes in order to define word meanings. They have some ability to demonstrate literal as well as inferential understanding, and are able to draw conclusions and make some judgments about literary and informational texts. They can link main ideas with supporting details and provide a simple summary, when given supports to help them organize information. Students can make obvious connections between the text and their own experiences. They use graphics to enhance their understanding.

Level 3

The student at the basic proficiency level surpasses the modified expectation set for students in the third grade. Students performing at this level consistently demonstrate comprehension of a variety of grade-level texts when scaffolding and supports are provided (e.g., thought bubbles, segmented texts, simplified language, graphic organizers, and timelines). They generally recognize the author's purpose and can sometimes delineate the various elements of literary and informational texts. Students understand and acquire new vocabulary and correctly use it in reading with supports to help them. They demonstrate an adequate understanding of grade-level appropriate reading concepts and skills. They gain meaning from textual elements, exhibiting literal, as well as emerging, inferential cognizance. They are able to consistently draw conclusions and make judgments about grade-level literary and informational text. Typically, they are able to make important connections among ideas, develop new understanding, and summarize detailed literary and informational texts when supports are provided. They consistently use graphics to enhance their understanding.

Table 8. Participation Statistics for the 2011 Georgia CRCT-M

	Number of Students Enrolled during Test Window (Grades 3 through 8)	Number of Students taking CRCT-M	Percent Participating in CRCT-M
Reading	760,627	11,357	1.49%
English	760,625	12,258	1.61%
Mathematics	761,372	15,824	2.08%

Table 9. Results of the CRCT-M in 2011

	Total	Basic	Proficient	Advanced
Reading	11357	32% (3639)	54% (6157)	14% (1561)
English	12258	36% (4448)	47% (5765)	17% (2045)
Mathematics	15824	31% (4949)	56% (8905)	12% (1970)

South Carolina chose not to pursue the path of developing an AA-MAS, and instead focused their efforts on taking the lessons learned from the GSEG studies on improving items on the general assessment to better support persistent low performers and for providing additional professional development to their educators. One Research to Practice workshop was held in 2008 focusing on training teachers on progress monitoring and data collection. They held a second Research to Practice workshop in July 2011 focusing on four topics:

1. The changing curricular context of students with significant cognitive disabilities; addressing functional skills with standards-based instruction.

2. Applying a learning progressions schema to achieve more coherent curricular and instructional planning.

3. Accommodations addressing the needs of students with disabilities.

4. Understanding challenges for students with disabilities on state assessments in English language arts and mathematics.

Experts from the National Center for the Improvement of Educational Assessment, University of Kentucky, WestEd, Minnesota State University, and University of Illinois lead the various workshops.

South Carolina also did an extensive analysis of the Individualized Education Programs (IEPs) of students who would be eligible for the AA-MAS. The resulting information is also being used to plan professional development activities for teachers with students in this target population and to provide additional guidance on developing standards-based IEPs (Karvonen, Rao, & Morgan, 2010). Thus, this grant provided an opportunity for South Carolina policymakers to work with researchers to better understand low achieving students with disabilities. The work led to additional training opportunities, focused professional development, and an understanding of barriers on the general assessment that can be reduced through improved item writing and review.

Discussion of Lessons Learned in Two Southern States

A difficulty in studying the effects of change on student performance is that no change is made in a vacuum. Georgia and South Carolina both chose to identify the target population for the AA-MAS as students who consistently scored in the lowest performance level on the general assessment. The teachers provided remarkably similar profiles of the students through focus groups and surveys.

One piece that this chapter did not cover in depth was the immediate work done in both states to better inform teachers and IEP teams on setting goals and providing scaffolding in instructing these students. While, the extent of changes in the classroom is unknown, they may be affecting student performance.

Likewise, both states made numerous changes to the assessment items. In Georgia, they actually produced an AA-MAS, which was a shorter test, with a more appealing layout (larger font and more white space), and containing items that had been revised or enhanced to make them more accessible to the targeted students. The majority of items were changed in more than one way. Subsequent studies to determine which revisions and enhancements were most effective did not provide consistent information. Although South Carolina never created a full test form, the item modifications produced mixed results in the cognitive labs. They clearly made the items easier overall, but it was difficult to parse out exactly which modifications had helped and how.

In both states, teachers working on modifications wondered why they could not make many of these changes on the general assessment, an improved implementation of universal design rather than modifications that altered the construct. In truth, they developed new items with these changes or changed current items, as long as the items are not linking items currently used for equating. In general, both states gained knowledge and experience on how to create better assessment items, focused on the construct, and without irrelevant and distracting information.

Regardless of the future of the AA-MAS, much of what was learned through these grants will help instruct students with disabilities and produce better assessment items for all students.

Implications for Building Future Assessments

More research is needed to tease out exactly which combination of item revisions and enhancements produced the effect of better performance. However, as states and groups of states develop assessments aligned to the Common Core State Standards, they should consider the lessons learned from the AA-MAS. For instance, while many would have argued that they applied universal design principles to their general assessments, the work on item modifications taught us that more could be done. Language can be simplified, key words can be highlighted, and graphic organizers can be used judiciously to focus students on the target construct. Chunking passages and associating items directly with the section of the passage to which they pertain is one of the modifications that at least appeared to be most different. There needs to be additional research on whether it changes the construct to ask the question in closer proximity to the location where the answer can be found or inferred. If it does not take away from what is being assessed, then future assessments should consider including one segmented passage to help ensure that there is a range of reading passages for all students to access.

Moving to online assessments will reduce the need for larger font and more white space in a test booklet. Universal design can truly be applied when every student can choose the font size appropriate for them. However, the field will need to research a student's ability to select the font size that is most appropriate for him or her, particularly when a consequence of increasing the font size of a passage results in the need to scroll more to read the full passage. Technology can help by magnifying certain sections to both help with vision and scrolling. Likewise, students could select scaffolds such as hints, vocabulary definitions, or graphic organizers that have been pre-determined not to interfere with the construct being assessed. Those who do not wish to use such supports will simply not access them.

New assessment development should draw heavily on the lessons learned since the AA-MAS was introduced in 2007, but as new opportunities arise, more research is needed as well. One lesson learned from this experience is that just because we can change items, add scaffolds, or reduce item options does not mean that we should.

References

Beddow, P. A., Kettler, R. J., & Elliott, S. N. (2008). *Test accessibility and modification inventory (TAMI)*. Nashville, TN: Vanderbilt University. Retrieved May 18, 2011, from http://peabody.vanderbilt.edu/tami.xml

Hess, K., McDivitt, P., & Fincher, M. (2008). *Who are the 2% students and how do we design test items and assessments that provide greater access for them? Results from a pilot study with Georgia students.* Paper presented at the CCSSO National Conference on Student Assessment Orlando, FL. Retrieved May 18, 2011, from http://www.nciea.org/publications/CCSSO_KHPMMF08.pdf

Johnstone, C. (2009). *Cognitive Interview Study Results: Georgia AA-MAS*. Minneapolis, MN: National Center on Educational Outcomes.

Karvonen, M., Rao, V., & Morgan, G. (2010). *Curricular priorities and instructional practices for students who are eligible for AA-MAS.* Paper presented at the 2010 annual meeting of the American Educational Research Association, Denver, CO.

Nagle, K. M., Cameto, R., Almond, P., & Morrison, K. M. (2006). *Oklahoma special education teacher survey.* Menlo Park, CA: SRI International.

USED. (2007). *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.