

MEANINGFUL MEASUREMENT

The Role of Assessments in Improving High School Education in the Twenty-First Century

June 2009



© 2009 Alliance for Excellent Education. All rights reserved.

Suggested citation:

L. M. Pinkus, ed., *Meaningful Measurement: The Role of Assessments in Improving High School Education in the Twenty-First Century* (Washington, DC: Alliance for Excellent Education, 2009).

Ordering information:

Copies of *Meaningful Measurement: The Role of Assessments in Improving High School Education in the Twenty-First Century* can be downloaded from the Alliance's website at www.all4ed.org. To request print copies of the report, please visit http://www.all4ed.org/publication_material/order_form. The first copy of the report is complimentary. Additional copies are available at a charge of \$1 per copy to cover shipping and handling costs.

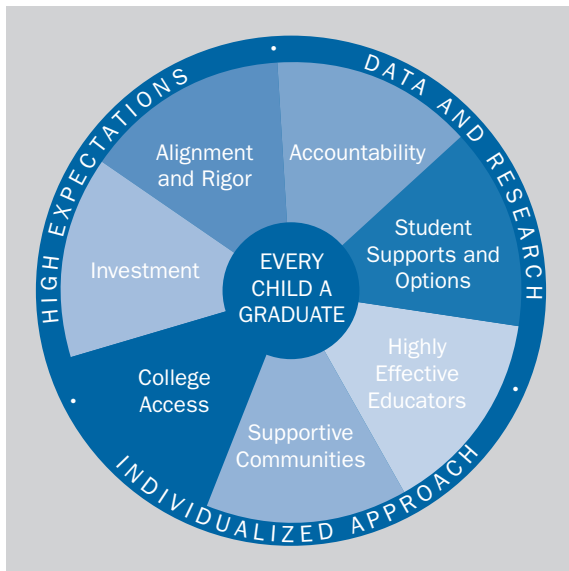
About the Alliance for Excellent Education

The mission of the Alliance for Excellent Education is to promote high school transformation to make it possible for every child to graduate prepared for postsecondary learning and success in life.

The Alliance for Excellent Education is a national policy and advocacy organization, based in Washington, DC, working to improve national and federal policy so that all students can achieve at high academic levels and graduate high school ready for college, careers, and citizenship in the twenty-first century.

The Alliance has developed a “Framework for Action to Improve Secondary Schools” that informs a set of federal policy recommendations based on the growing consensus of researchers, practitioners, and advocates about the challenges and solutions for improving secondary student learning.

The framework, shown graphically here, encompasses seven policy areas that represent key leverage points in ensuring a comprehensive, systematic approach to improving secondary education. The framework also captures



three guiding principles that apply to all of the policy areas. Although the appropriate federal role varies from one issue area to another, they are all critically important to reducing dropouts and increasing college and career readiness.

About the Editor

Lyndsay M. Pinkus is director of strategic initiatives at the Alliance for Excellent Education. Since joining the Alliance in January 2002, she has served in a variety of research, coordination, and advocacy roles, where her work has included managing policy and grant work on a range of issues including graduation rates, data, secondary school accountability, and secondary school improvement, and authoring a number of publications for the Alliance. Prior to rejoining the staff in January 2006, Ms. Pinkus served as a legislative associate at Washington Partners, LLC, providing government relations and policy research and analysis for a variety of clients, including the Alliance. She is a graduate of the School of Public Affairs at American University as a presidential scholar; the Public Affairs and Advocacy Institute at the Center for Congressional and Presidential Studies; and the Institute for Educational Leadership's Education Policy Fellowship program.

Acknowledgments

The Alliance for Excellent Education is greatly appreciative of the authors for sharing their time and expertise in writing the following chapters, as well as of the multiple Alliance staff members and advisors whose dedication contributed significantly to this volume.

The Alliance for Excellent Education is also grateful to Carnegie Corporation of New York for the financial support that made this publication possible.

The views expressed in this volume are those of the authors and do not necessarily represent those of the Alliance for Excellent Education or the funder.

CHAPTER

8

Students with Disabilities: Expectations for Academic Achievement, and the Critical Role of Inclusive Standards-Based Assessments in Improving Outcomes

Rachel Quenemoen

National Center on Educational Outcomes, University of Minnesota

Prologue: Story from San Diego—The Dilemma

Excerpt from E. Alpert, “Deterred from Diplomas for Better or Worse,” Voiceofsandiego.org, October 2, 2008:

Just after starting high school, Lance Rogers was told he wouldn’t earn an ordinary diploma. He struggled with attention deficit hyperactivity disorder and other disabilities, and had trouble focusing in big classes at Point Loma High School.

Instead he took special education classes that were smaller and easier, but wouldn’t help him earn a degree. His mother Ruth Rogers hoped he would flourish there, even if he was “non-diploma bound.” It is a label given to thousands of San Diego Unified students with disabilities who focus on skills that will help them live independently instead of prepping for college or beyond, studying shopping lists and sales tax instead of calculus or Cervantes. But Lance Rogers grew depressed and bored in those classes. He

can't remember what he learned—only that he was often asked to draw pictures or maps—and ultimately ditched school.

“I was downhearted,” said Lance Rogers, now 16 years old. “I didn't do my work, because what was the point of doing it? I didn't get any credit. So I didn't go to school.”

Yet when the Rogers family moved to Texas, their son thrived in a school with a mixture of small classes and counseling. His grades rose from Ds to Bs. And when the family returned to San Diego, teachers at another school said Lance Rogers was perfectly capable of earning a diploma.

“I was blown away,” Ruth Rogers said. “I was shocked that he was in the classroom, doing what he's supposed to be doing.” ...

No educator means to shortchange children with disabilities, but an overburdened and underfunded system causes mistakes when diagnosing and placing children in classes, said parent Joyce Clark, chairwoman of a San Diego Unified committee on special education. Clark said some children are funneled into easier classes instead of making ordinary classes accessible through technology or other aids.

“Teachers are wonderful but they get weary of trying to address all the needs they are asked to do,” Clark said. “And somehow [some students] just fall through the cracks.”

This story illustrates a dilemma that creates false choices for students, parents, and teachers. This dilemma stems from common misconceptions held by the public, policymakers, school leaders, and even teachers of how specific learning needs related to identified disabilities affect a student's ability to learn and to earn a regular diploma. It results in persistently low expectations for the achievement of students with disabilities, unwillingness of schools to be held accountable for their progress (or lack of it), and low levels of achievement and postschool success for many of these students. These erroneous but pervasive misconceptions of high school students with

disabilities are deeply affected by policies related to standards, assessments, and accountability systems, as described in this chapter. To improve expectations, teaching, learning, and outcomes—for all students, including students with disabilities—it is critical that policies at all levels help leverage implementation of an inclusive assessment system that supports these goals.

This chapter describes issues concerning assessing students in a standards-based accountability system and related federal policies. It also describes ways to evaluate assessments that are inclusive of all students in the accountability system. It concludes with recommendations for policymakers.

Misconceptions and Low Expectations

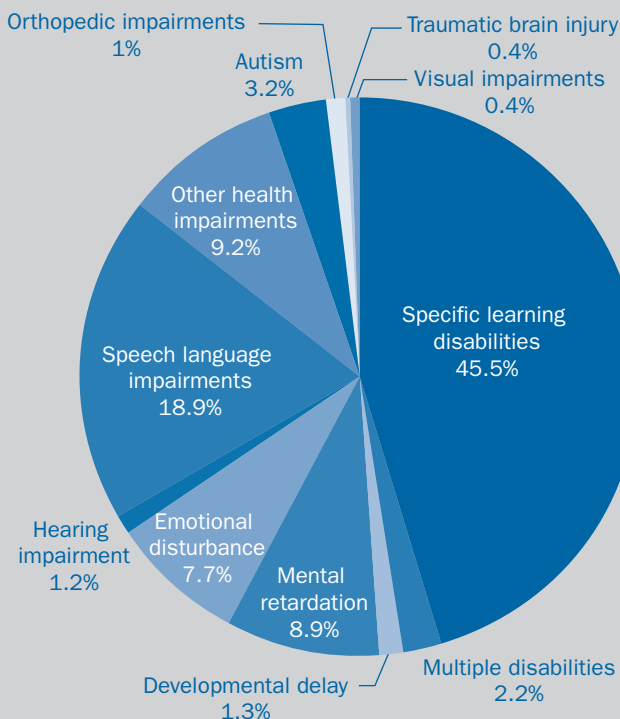
The learning characteristics of students with disabilities vary greatly. Understanding who these students are, and how their disabilities may affect their learning, is foundational to understanding how assessments can yield data to hold schools accountable for the results. See Figure 1 for a summary of categorical distribution of students with disabilities.

The impact of the spectrum of disabilities on students' ability to learn is described by disabilities expert Martha Thurlow, director of the National Center on Educational Outcomes (NCEO), as follows:

Most students with disabilities (75 percent altogether) have learning disabilities, speech/language impairments, and emotional/behavioral disabilities. These students, along with those who have physical, visual, hearing, and other health impairments (another 4–5 percent), are all students *without intellectual impairments*. When given appropriate accommodations, services, supports, and specialized instruction, these students (totaling over 80 percent of students with disabilities) can learn the grade-level content in the general education curriculum, and thus achieve proficiency on the grade-level content standards. In addition, research suggests that many of the *small* percent of students with disabilities *who have intellectual impairments* (i.e., generally includes students in categories of mental retardation, developmental delay, some with multiple disabilities, some with autism), totaling less than 2 percent of the entirety of the student population, *or less than 20*

percent of all students with disabilities, can also achieve proficiency when they receive high quality instruction in the grade-level content, appropriate services and supports, and appropriate accommodations.

Figure 1: Categorical Distribution of Students with Disabilities



Note: Percentages in this figure are based on a total number of 6.5 million students receiving special education services (www.IDEAdata.org, based on 2005 data).

Source: IDEA Part B Child Count, 2005, “Students Ages 6 Through 21 Served Under IDEA,” by disability category (Tables 1–3), www.IDEAdata.org (accessed September 2008).

This last point—that students with intellectual impairments can also achieve proficiency—is supported by the research of Dr. Kevin McGrew, a coauthor of the Woodcock-Johnson III, one of the most widely used instruments for assessing both cognitive abilities and achievement in children and adolescents. Using student testing data from the Woodcock-

Johnson III development processes, McGrew studied whether measured intelligence quotients, or IQs, can predict eventual academic achievement. He found that “it is not possible to predict which children will be in the upper half of the achievement distribution based on any given level of general intelligence. For most children with cognitive disabilities (those with below average IQ scores), it is not possible to predict individual levels of expected achievement with the degree of accuracy that would be required to deny a child the right to high standards/expectations.”¹

Still, it is common to hear educators, members of the public, or policymakers say, “Well, of course these students don’t do well on the tests, they have disabilities!” Once those low expectations are entrenched, they play out in very destructive ways. The literature on the effects of teacher expectations on student achievement is compelling, demonstrating conclusively that what we expect in student learning is typically what we get, regardless of student ability. (See McGrew and Evans 2004 for a summary of the literature.) This is alarming, given that so many educators seem to believe that students with disabilities cannot learn the content expected for other students—or, as the earlier excerpt from *Voice of San Diego* suggests, that they have too many learning needs to warrant the effort required for them to learn it. The next section will address what content is expected for all students, including high school students with disabilities.

Standards-Based Reform, Expectations, and Student and System Accountability

Federal policy has played a significant role in improving how the public education system serves students with disabilities. By the end of the 1970s, federal policy (in the form of PL 94-142) guaranteed that students with disabilities had access to school buildings. Over the course of the 1990s, federal laws funding both special education² and education for the disadvantaged³ as part of the standards-based-reform national agenda not only defined the right of students with disabilities to the same goals and standards as all other students, but also required the full inclusion of every student in assessments designed to provide data on how well all students were being taught. Most recently, the reauthorization of the Elementary and Secondary Education Act—known as the No Child Left Behind (NCLB) of 2001 (PL 107-110)⁴—and the passage of the Individuals with Disabilities

Education Improvement Act (IDEA) of 2004 (PL 108-446)⁵ more closely aligned the two major education laws with common accountability for results. Although the focus in IDEA is on individual accountability and the focus in NCLB is on systems accountability, both laws are built on the goal of raising academic achievement through high expectations and high-quality education programs, to improve outcomes for all students, including those with disabilities.

Regardless of these laws, decisions about what every student should know and be able to do in a standards-based system are made at the state and local levels and not at the federal level. Since the late 1990s, policymakers and citizens in every state have grappled with the fundamental question of “What is a well-prepared student?” and each state has answered that by defining content standards (what) and achievement standards (how well) which identify essential skills and knowledge for students to master at each grade level. Cumulatively, these standards define what a high school graduate is expected to know and be able to do.

States receiving federal funding under either IDEA or Title I of NCLB are required to develop such standards, and NCLB requires standards to apply to all students in all public schools. The state-developed grade-level content and achievement standards are the foundation on which states build an assessment system. NCLB requires states to assess all students once annually in grades three through eight and at least once in grades ten through twelve in mathematics and reading, and once annually in each of three grade bands in science. IDEA clarifies that students who receive special education services are to have access to and make progress in the general curriculum based on these same standards, and reinforces the requirement of full inclusion of all students with disabilities in the NCLB-required assessments (as well as in all other assessments administered by the schools). IDEA further specifies that states and districts must provide appropriate options for all students with disabilities to participate in these assessments, including requirements for universal design of assessments, accommodations, and alternate assessments.

These assessments are used for a variety of purposes:

- NCLB requires states to use the assessments to hold schools accountable for the achievement of these standards by all students. A standards-based accountability system that requires the school system to demonstrate that all students can meet state or locally defined standards is meant to ensure that all students are prepared for a successful future. This is called **system stakes**, or **system-level/school-level accountability**. This means that there are consequences when public schools do not ensure that all students have mastered the essential content. Inclusive large-scale assessments used for system accountability are meant to shine a light on whether schools are teaching all students well.
- The use of standards-based assessments to hold students accountable for learning is called **high-stakes testing for students**, or **student accountability**. Typically, student accountability is meant to assure future educators or employers that this student knows and can do what the state has determined to be essential for future success.⁶ The stakes for students may include some type of grade promotion or retention; in some states, assessments are used to decide whether or not a student may be granted a regular diploma.

It is possible that student accountability and system accountability working together may have powerful effects on student achievement, but there are unintended consequences when students are held accountable within a system that is not achieving the expectation that all students be successfully taught. Too often in the past, when some students or groups of students were not achieving, the easy answer was to lower individual or group expectations through strategies like those evidenced in the opening story of this chapter—“focus[ing] on skills that will help them live independently instead of prepping for college or beyond, studying shopping lists and sales tax instead of calculus or Cervantes ... funneled into easier classes instead of making ordinary classes accessible through technology or other aids.” As this experience shows, the disconnect can result in negative outcomes for some students, and raises questions about whether the students who have a right to a free appropriate public education (as required by IDEA) have in fact been provided with that opportunity.

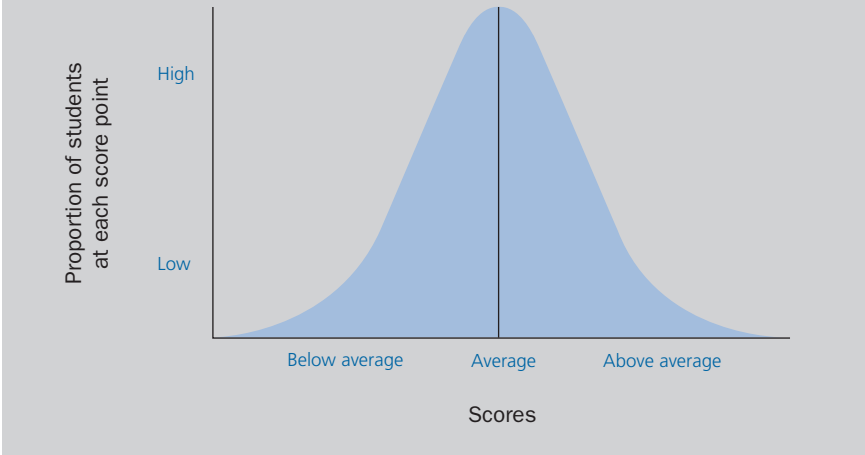
When some students, or groups of students, are not achieving to the levels expected for all students, the *expectations* should not change. Instead, *the services, supports, and specialized instruction* that ensure all students achieve the standards should change; in many cases, this requires increased training and support to the educators who work with these students. Through its standards and assessments requirements, NCLB has initiated a positive shift from the promotion of a separate and less rich or challenging curriculum for students with disabilities to an expectation of grade-level academic achievement as defined for all students. IDEA and NCLB jointly support this shift, but much work needs to be done to overcome decades of low expectations and deeply engrained beliefs among some stakeholders that nothing can be done to improve the achievement of students who have disabilities. As policy choices are made now and in the future, this history and set of beliefs needs to be articulated so that the implications of choices are clearly understood, and the consequences carefully monitored.

Understanding the Purposes of Assessments

Parents, policymakers, and the public can make informed decisions for individual students and about the policies that support improved student achievement by understanding the varying purposes and uses of common types of large-scale assessments. These varying purposes and uses define how students with disabilities should participate in each testing option. These fundamental differences are commonly misunderstood by many stakeholders in discussions of NCLB.

Norm-referenced tests (NRTs): One common use of tests has been to compare students' performance and rank-order them accordingly. NRTs are well suited for this, as these tests provide percentile ranks that tell us the percentage of a norm group—that is, the scores obtained by other students—that a given student performed as well as or better than. A graph called a “normal curve” or a “bell-shaped curve” similar to the one in Figure 2 is sometimes used to show that often most students perform about average and fewer students score much higher or much lower than the average. A student who performs as well as the average student will be equal to or better than 50 percent of the norm group and equal to or lower than 50 percent of the norm group. Hence, the average student is smack in the middle of this “normal” curve.

Figure 2: Norm-Referenced Test Results



NRTs are good for comparing students with respect to very general domains of performance, as well as for sorting out large populations for specific purposes (like army personnel assignments, or admission to college). However, they are not as well suited for evaluating students' performance with respect to curricula such as those defined in state-developed curriculum frameworks. They are also not meant to provide diagnostic information with respect to specific skill areas. Rather, they help to indicate strengths or weaknesses relative to others—that is, to the specific norm group that was available when the test was developed. Tests that are designed to measure what groups of students know compared to well-defined content and achievement standards require a different type of assessment called criterion-referenced testing (see next page).

More than a century of norm-referenced testing designed to distribute students along a normal curve has affected the perceptions of teachers, parents, policymakers, and the public—many are familiar with the NRTs they have themselves taken throughout their own academic career. This has resulted in a popular belief that on any skill taught, half of the students will perform “below average.” In this context, there is a temptation to predict or assume *which* students will end up on the bottom. Although data from many states suggest that more than half of the students performing “below

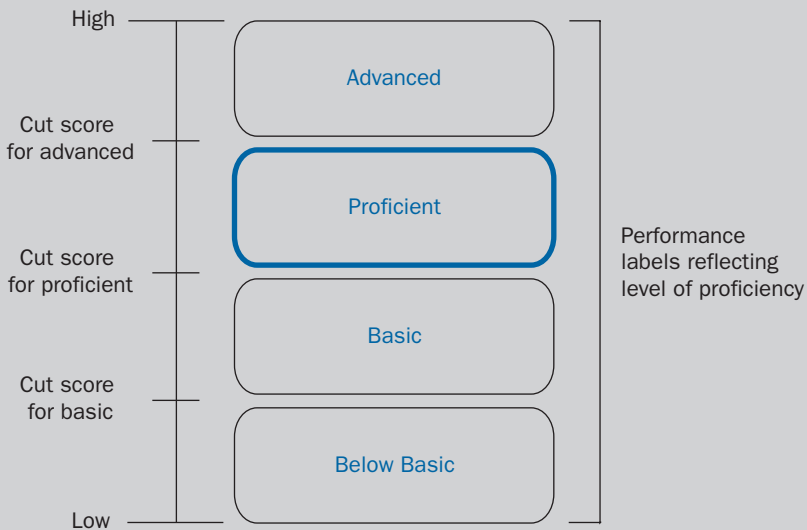
average” do not have disabilities and are disproportionately poor and minority students,⁷ there is a negative and pervasive public perception that students with disabilities are always—by definition—the lowest-performing students.

Another drawback to norm-referenced testing is that it cannot indicate whether a student’s test performance is satisfactory in and of itself. Scoring “above average” or “below average” means little if the “average” is very low or very high.

Criterion-referenced tests (CRTs): By contrast, criterion-referenced tests are designed to measure and provide information regarding how well students have mastered specific knowledge and skill areas they have been taught. For example, they can be used to answer questions regarding whether a student’s performance on a test signifies “proficiency” in a subject area. Such achievement-level classifications are usually set on a test by establishing one or more “cut scores” to reflect standards determined by carefully selected groups of educators. Examples of these achievement standards include the performance standards established by states to meet the requirements of NCLB (e.g., below basic, basic, proficient, or advanced). Students’ test results are compared to these cut scores to determine whether their performance is satisfactory. In CRTs, it is irrelevant whether a student’s performance is at, above, or below the average of some norm group. Students who have been taught well should be able to demonstrate proficiency on the standards. Figure 3 depicts how CRT results reflect the levels of achievement of knowledge and skills defined by the academic standards.

In the past, when, after receiving the same instruction provided to all students, students with disabilities did not perform well compared to their peers, common practice was to provide a lower or slower curriculum than that taught to their peers. This inevitably resulted in an increasingly large gap in performance, year after year. By advancing the use of criterion-referenced assessments, NCLB promotes a shift in the teaching of students with disabilities. The altered goal contains three parts: to teach all students a rich curriculum based on the same content; to tailor services, supports, and specialized instruction for some students to be sure they master the content; and to test to ensure that they have indeed learned it.

Figure 3: Criterion-Referenced Test Results



Cut scores required for each category of performance are determined through systematic panel review processes by stakeholders with expertise in the content being assessed and experience with the students being tested.

Out-of-level testing: Sometimes, when comparing a low-performing student's test results to a norm group's performance (through NRTs) or to predefined achievement standards (through CRTs), information about the student's learning is limited. For this reason, educators have sought additional ways to measure the performance of some students—in particular, students who achieve at very low levels on the assessments designed for their enrolled grade. One strategy has been to use out-of-level testing, assessments based on below-grade-level skills and knowledge, instead of the skills and knowledge for a student's enrolled grade level. Prior to NCLB's requirements that states use CRTs, out-of-level testing was performed by many local schools, districts, and even states as part of accountability systems, and results were used in developing students' Individualized Education Plans (IEPs). This strategy was often employed in the name of shielding some students from difficult testing situations and protecting their sense of self-worth. However, research suggests that with these lower expectations came lower achievement.⁸ A student with

disabilities who is consistently tested out of level may never be able to meet the requirements for high school graduation—a consequence that permanently affects their dreams, aspirations, self-worth, and postsecondary options. Special educators report that, too often, teachers, parents, and students do not think about these long-term consequences.⁹

Although out-of-level testing may be appropriate for some purposes, it is not designed to measure how well students are achieving against common expectations in a standards-based accountability system.

Adaptive testing: A related strategy for assessing students with disabilities has been the use of a type of adaptive testing. These tests typically include a computerized administration of test items that, based on how a student responds to earlier items (and often the relative difficulty of items), automatically selects additional items for each student. The use of such assessments results in individualized tests for each student. Advocates for the use of adaptive tests suggest that the individualized set of items will generate data that better matches what each student actually can do, as opposed to what they cannot do. Given the endless variation of what is tested in each situation, the results of these assessments also reflect varied and often unspecified content expectations, essentially measuring different standards for each student. In other words, putting a bank of test items on a computer that selects items for each student by difficulty does not generate useful information about how well the student has been taught the grade-level content. These adaptive assessment tools meet the requirements—and higher expectations—of standards-based reform only when they are designed to ensure that they are measuring student performance relative to grade-level content and achievement standards for the student's enrolled grade.

Options for Including All Students in Standards-Based Assessments

To effectively measure standards-based learning by all students, assessment systems must include options that ensure all students can demonstrate what they know. Currently, NCLB—both the legislation and related regulations—permit an array of assessment options for students with disabilities. These options are described below and summarized in Table 1.

General assessment, with or without accommodations: The vast majority of students with disabilities are given the general assessment that is taken by their peers without disabilities, with or without accommodations. These tests are developed with design elements—often called “universal design”—that make tests more accessible and result in more accurate scores that reflect actual student knowledge and skills, not extraneous factors. For example, most standards-based academic assessments used for school accountability are *not* intended to measure student characteristics and skills such as visual acuity, hand-eye coordination, or the ability to find isolated facts within a puzzle of distracting information. During the development of assessment items, universal design characteristics are considered in order to remove the effects of these kinds of extraneous and confounding factors. The result is an assessment that measures student abilities in the skills and knowledge intended to be assessed by the test, and not the unrelated effects of their disabilities.

Universal design does not mean that accommodations are no longer necessary. Accommodations are *changes in testing administration or materials* that enable students to participate in assessments in a way that allows their abilities—rather than the effects of their disabilities—to be assessed. Without accommodations, the assessment may not accurately measure a student’s knowledge and skills. Even with assessments that are universally designed, some students may still need accommodations related to

- presentation (e.g., repeat directions, read aloud, large print, Braille, etc.);
- equipment and material (e.g., calculator, amplification equipment, manipulatives, etc.);
- response (e.g., mark answers in book, scribe records response, point, etc.);
- setting (e.g., study carrel, student’s home, separate room, etc.); or
- timing/scheduling (e.g., extended time, frequent breaks, etc.).

Use of accommodations on a standards-based assessment assumes that careful consideration is given to whether the grade-level content and achievement standards being measured remain constant despite the use of the accommodation. The foundation for the assessment—the academic

content and achievement standards—remains the foundation even when accommodations are thoughtfully provided, thus maintaining high expectations for student achievement.

The collective knowledge base on the effects of accommodations on the content being measured is growing, but there are considerable complexities in the case of the most challenging content and student combinations. While it is widely accepted that most accommodations do not change what the assessment is measuring, more substantive changes in test administration and materials sometimes alter both what is measured and the meaning of the results. These are considered *modifications*. For example, a common focus in the lower grades is to measure a student's ability to decode printed text (as in the third-grade reading assessments). If the tester reads the test passages aloud to a student, this fundamentally changes the nature of the test, and measures reading comprehension, perhaps, but not the decoding of printed material. In this case, a read-aloud strategy would be considered a modification—a change that alters what is being measured—and not an accommodation.

However, for a small number of students, modifications are the only way they can interact with portions of the test. As noted by special education legal advocates Kathleen Boundy and Joanne Karger,

Many students with specific learning/reading disabilities struggle greatly with decoding text, yet have strong comprehension skills when access to information is provided through alternative modes that include: auditory, tactile, visual, and a combination of auditory and visual modalities. Similarly, individuals who are visually impaired may not be able to decode text and participate in the State assessment without an accommodation that allows them access to the information and questions in the text on which their comprehension is being assessed. The failure to differentiate between decoding skills and the broader comprehension of information and range of literary knowledge that are within the scope of the academic content standards embedded in a language arts curriculum denies the meaningful and effective participation of students with specific learning disabilities, who are otherwise unable to participate.¹⁰

States are working to develop testing strategies that capture student knowledge without lowering standards or losing the integrity in measurement and the meaning of the test. A few states have made and defended sometimes controversial decisions on these issues, like those described above by Boundy and Karger, but these states also require close monitoring and accountability for schools where these accommodations are selected for students. There is a delicate balance between access and inadvertently lowering standards and, thus, expectations and outcomes.

Alternate assessments: There is a small group of students with disabilities who cannot demonstrate what they know on the regular assessments, even with the use of accommodations. For these students, NCLB permits the development of alternate assessments to measure how well students know the skills defined in their enrolled-grade content standards. Over the past decade, there has been a rapid evolution of our understanding of how students build competence, especially with regard to students who need alternate assessments, but many misconceptions and erroneous assumptions around this issue remain. It is important to build a common understanding of the distinctions so the goals of reform are not derailed by these misunderstandings. Alternate assessments provide a critical role in ensuring that we obtain truly accurate measures of the knowledge and skills of all students with disabilities. NCLB allows for three alternate assessment options, described below and summarized in Table 1.

Alternate assessments based on grade-level achievement standards (AA-GLAS) are meant to assess the *same* content with the same definition of “how well and how much” as is measured by the general assessment. The format of the test is the key variable—for example, a portfolio of student work and a panel of content experts/teachers who review each portfolio instead of a multiple-choice or constructed-response test. There are very few such alternate assessments in place in states now, in part because of the very difficult challenge of showing comparability to the general assessment. This challenge is in part a policy decision that requires high comparability as a control on potentially lowered standards being hidden by such alternate assessments. Still, there are policy options that would avoid lowering standards and also allow for multiple ways to demonstrate grade-level achievement. The goal is to ensure that students who have disabilities that

interfere with good measurement of the skills and knowledge they do in fact have are included in accountability on the equivalent expectation of grade-level achievement standards. The challenges faced in designing the strong accommodations policies described by Boundy and Karger are directly related to these potential policy shifts. By reexamining the opportunities of the AA-GLAS, policymakers should find ways to avoid unnecessary use of less rigorous testing options like the remaining two alternate assessments, discussed below.

Alternate assessments based on modified achievement standards (AA-MAS) are meant to assess the *same* content with a possibility of a slightly less difficult definition of “how well and how much” than is used for the general assessment. It is unclear from NCLB regulatory language and guidance what exactly that means, and not every state is opting to develop an AA-MAS. If a state chooses to develop an AA-MAS, it is required to define the standard of “how well and how much” through a documented and validated standard-setting process, involving stakeholders who know the student and the content and who in theory have the best interests of the students in mind. Given the history of low expectations and deeply held beliefs among some stakeholders that many students with disabilities cannot learn the same content as their peers, these definitions are of concern to many advocates, who fear that lowered expectations will be reinforced by provision of this testing option.

The governing regulations limit the number of IDEA-eligible students who can be determined proficient for purposes of determining whether a school, district, or state has made AYP, based on the modified assessment—up to 2 percent of the general school population, or about 20 percent of special education students. To date, no state has received full approval through the federal peer-review process for its AA-MAS.

To complicate the identification of students who can take the AA-MAS, the data from multiple states indicate that the students who are the lowest-performing 2 percent on regular assessments are a blend of students with and without disabilities, and who predominantly represent student groups who have historically been on the low side of the achievement gap.¹¹ However, the current regulation allows states to implement the modified

assessment only for students with disabilities. This data, combined with evidence that many of these students have not been taught the challenging curriculum expected for all students, suggest a need for thoughtful and data-based processes to understand what a modified achievement standard should represent.

Over the next several years, states will need to work in partnership with researchers and experts to better understand and identify the appropriate students for participating in the AA-MAS. A key question in this process is whether these students have been provided the services, supports, and specialized instruction they need to succeed. The goal of this work must be to understand how these students can build and demonstrate the skills and knowledge they need to earn a regular diploma and to succeed in adult life. States must focus on the development of modified achievement standards and modified assessments that raise expectations for all students and close persistent achievement gaps.

Some have argued that this testing option has resulted in unexpected positive consequences—specifically, that there is attention placed on students who previously were ignored. Even so, it has the potential to create unintended but negative effects that may perpetuate low expectations, sustain achievement gaps, and limit students’ access to graduation with a regular diploma and college and work opportunities.

Alternate assessments based on alternate achievement standards (AA-AAS) are meant to assess the *same* content with less depth, breadth, and complexity than the regular assessment, and with a *different* definition of “how well and how much.” Just as is the case with the AA-MAS, states must define these standards using a documented and validated standard-setting process, and there are concerns about whether these standards reflect an appropriate raising of the bar of expectations that will yield increased achievement. There is, however, less controversy among advocates about whether this type of alternate assessment is a necessary option. Instead, there is debate within special education about what content should be assessed—given the flexibility states have in defining the depth, breadth, and complexity of the content to be assessed—and which students should be eligible for this testing option.

The AA-AAS are intended to be used with students with significant cognitive disabilities, typically defined as those with the most severe intellectual disabilities and multiple disabilities—children who represent fewer than 1 percent of all students, or less than 10 percent of all students who have disabilities. When the regulation permitting the AA-AAS was proposed, stakeholders debated which students should be included. Estimates of how many students have the most severe intellectual and multiple disabilities ranged from less than 0.5 percent of the total student population to as high as 3 percent. The lowest percentage (0.5 percent) was supported by data in states that report moderate and severe mental retardation as separate from all students with mental retardation and from Centers for Disease Control data on incidence of correlated disability diagnoses.¹² The higher estimates generally included many students with mild disabilities from all disability categories, many of whom do not have intellectual disabilities but who were performing at low levels. The eventual selection of 1 percent as the cap on the percent of students whose scores can be treated as proficient for purposes of school accountability was a compromise. Certainly, more students can participate in the AA-AAS than 1 percent, but from a policy perspective the cap on how the scores are used in accountability was intended to prevent inappropriate inclusion of many students in a lower achievement expectation than evidence suggests is warranted.

Inclusive assessment systems have been the cornerstone of policies based on an assumption that by including all students in assessments used to determine how well schools have taught all students, educators will be motivated to ensure that students who have not had access to the general curriculum in the past will be taught. Studies from multiple states show that students with significant cognitive disabilities have benefited from a noticeable increase in their access to the general curriculum because of the NCLB requirements for these assessments. The consequential validity studies of the AA-AAS document the benefits of the policy push of including all students in standards-based assessment and accountability systems. There have been reports of dramatic increases in other valued outcomes for these students, such as increased use of assistive technology, which leads to an increased level of independence, an increased implementation of inclusive settings, and an increased interaction with typical peers.¹³

The achievement of these students on grade-level content is very different from their general education classroom peers, but the evidence of their work is compelling. These students are able to learn academic content with reduced complexity, breadth, and depth clearly linked to the same grade-level content as their peers. (The federally produced publication *Learning Opportunities for Your Child Through Alternate Assessments* provides specific examples of what that can look like.)¹⁴

Researchers and practitioners are working side-by-side to capture the nature of the linkages to the grade-level content, but the evidence of this improvement in student learning is startling, given that schools have not given these students access to this content in the past.¹⁵

Examining the Effect of Inclusive Testing Practices

Current test results show us that, generally, students with disabilities are not performing as well as typical peers. As states have examined this gap, many have found that the students with and without disabilities who currently score low on tests often have not been taught the tested content.¹⁶ These investigations have served to raise awareness about improving instruction for students with disabilities. The combination of the pressure to test all students and the focus on improving instruction has increased the pressure on schools to learn what works to ensure successful outcomes for all students.

To supplement anecdotal evidence, many states have instituted formal procedures to use assessment and accountability data to identify schools where reforms are yielding very high achievement for students with disabilities. There are also a few well-designed studies focused on what is occurring in schools where test scores are higher for students with disabilities. These studies consistently identify common characteristics among schools where students with disabilities achieve at high levels. As summarized in one study, the schools have

- a pervasive emphasis on the curriculum and alignment with the standards;
- effective systems to support curriculum alignment;
- an emphasis on inclusion and access to the curriculum;

- a culture and practices that support high standards and student achievement;
- a well-disciplined academic and social environment;
- continuous use of student data to inform decisionmaking;
- unified practices supported by targeted professional development;
- access to resources to support key initiatives;
- effective staff recruitment, retention, and deployment;
- flexible leaders and staff who work effectively in a dynamic environment; and
- effective leadership.¹⁷

In the past decade, NCEO's surveys of states have recorded state staff perceptions of changes occurring in their districts and schools. Survey respondents speak of improvements in the performance of their students, attributing those improvements to clear assessment participation policies, alignment of student Individualized Education Plans (IEPs) with standards, improved professional development, development and provision of accommodation guidelines and training, increased access to standards-based instruction, and improved data collection.¹⁸ Analyses of publicly reported assessment data since 2000–01 show improvements in the transparency of data for students with disabilities, for both participation and performance.¹⁹ For example, according to NCEO's research, the number of states with clear reporting to the public about students with disabilities' participation in testing options increased from only five states in 2000–01 to twenty states in 2004–05. These data also showed large increases in the percentage of students with disabilities who participate in the assessment system across time for most states.

A Vision for a Principled Approach to Accountability Assessments for Students with Disabilities

Building on research and practice, NCEO has identified the principles and characteristics that underlie inclusive assessment and accountability systems.²⁰ The vision for a principled approach to accountability assessments for students with disabilities includes the following six core principles:

Principle 1. All students are included in ways that hold schools accountable for their learning.

Principle 2. Assessments allow all students to show their knowledge and skills on the same challenging content.

Principle 3. High-quality decisionmaking determines how students participate.

Principle 4. Public reporting includes the assessment results of all students.

Principle 5. Accountability determinations are affected in the same way by all students.

Principle 6. Continuous improvement, monitoring, and training ensure the quality of the overall system.

These principles reflect the belief that all students with disabilities can and should have meaningful access to the same education as their peers without disabilities. For that to occur, they need to be taught well in the same curriculum and with the same expectations as their same-age classmates. Teachers should use systematic standards-based assessments both in the classroom as they are teaching and at the end of instruction to know how well the students were taught. These assessments will give schools the information they need to design instruction and supports so these students achieve in spite of barriers related to disabilities.

Although standards-based assessments are essential tools in the drive to ensure all students achieve at high levels, assessments can simply point out where teaching and opportunities to learn need to be improved. For improvement to occur, teachers need and deserve high-quality training, coaching, and professional support so they can be successful teaching all students. Staff development support to teachers is a necessary but sometimes neglected component of standards-based reform. Parents and the public can join forces to ensure that all teachers have the skills they need to do this important and challenging work. Our schools must be structured to allow students with disabilities to avoid the barriers that their disabilities create when accessing the curriculum and when demonstrating what they know and can do on assessments. Success in doing so is a critical step on the path toward lifelong success.

Policy Conclusions

National and state efforts in standards-based reform should include all students in standards, assessment, and accountability systems and

processes, including all students with disabilities. Any attempt to exempt schools from being accountable for a group of students undermines the entire reform effort. Previous assessment and accountability policies supported by federal funding for the disadvantaged (e.g., allowing the use of NRTs and out-of-level testing) and for students with disabilities (e.g., reliance on individual student accountability provisions connected to IEPs) have contributed to the well-documented achievement gaps. Federal policy related to standards-based reform, including standards, assessments, and accountability systems and processes, must adhere to this principle to diminish the negative effects of applying different expectations to different groups of students.

Standards-based assessment systems should include strategies that permit all students to show what they know and can do on the academic content standards defined for typical peers of the same age and grade level, despite the barriers of disability. This includes a continued emphasis on universal design of assessments, the development and implementation of high-quality accommodations policies, and provision of alternate assessments that allow different ways of demonstrating what a student knows and can do, including expanded options for alternate assessments based on grade-level achievement standards (AA-GLAS).

Any change in academic achievement standards for a group of students, including those already defined in regulatory language, should be reviewed to ensure that these options raise the bar of academic expectations, and thus increase system accountability for the outcomes of students who may participate in the option. Each option should be able to withstand scrutiny by external experts on whether the underlying assumptions are grounded in current research or practice that supports improved academic achievement for students who may participate in these options. Although there is evidence emerging from the use of the AA-AAS to that effect, there does not appear to be a similar body of evidence from the use of the AA-MAS thus far, or in the research base cited in the regulation.

The consequences of participation in any testing option that changes the achievement standards (including those options in place now and those proposed in the future) should be closely monitored. Higher achievement should be evidenced independently of state-set proficiency levels, which

could be artificially elevated by state-defined modified or alternate achievement-standard-setting practices. Careful study of performance-level descriptors and achievement-standard-setting practices can inform conclusions of whether low expectations are reinforced by these options. If so, any option that does not withstand scrutiny as a high expectation standard should be discontinued so that schools, districts, and states are held accountable for educating all students to high standards.

Afterword: Story from San Diego

Excerpt from E. Alpert, “Deterred from Diplomas for Better or Worse,” Voiceofsandiego.org, October 2, 2008:

As graduation rates have grown, brain research has shown the risks of underestimating children with disabilities, even those with severe conditions that prevent them from speaking, said Anne M. Donnellan, director of the Autism Institute at University of San Diego’s School of Leadership and Education Sciences. Donnellan has seen a number of nonverbal students such as Peyton Goddard overcome diagnoses of mental retardation and graduate from college.

Hehir [Tom Hehir, Harvard researcher and former assistant secretary of education in the Clinton administration] likewise noted that diagnoses are sometimes wrong and students should be given the benefit of the doubt. For instance, conventional wisdom that students with Down syndrome couldn’t learn to read has been shattered as many prove themselves capable of reading and writing as well.

“I’m not interested in predicting what people can do,” Donnellan said. “We’ve made some terrible mistakes with that.”

Lance Rogers believes he was a victim of those mistakes. Now a sophomore at the Marcy School, a San Diego Unified center that combines classes and counseling, Rogers said he’s taking algebra, chemistry and history to earn the diploma he once was blocked from.

By pursuing a diploma, “I did something they didn’t think I was going to accomplish,” he said. “They didn’t say it like that. But that’s what it comes down to.”

The views expressed in this chapter are those of the author and do not necessarily represent those of the Alliance for Excellent Education.

About the Author

Rachel Quenemoen is a senior research fellow at the University of Minnesota and the technical assistance team leader for the National Center on Educational Outcomes. She is coprincipal investigator of NCEO's federally funded technical assistance center. Ms. Quenemoen has worked for thirty years as an educational sociologist focused on research-to-practice efforts. She has been a multidistrict cooperative administrator in both general and special education, and for the last fifteen years has worked at the state and national levels on educational change processes and reform efforts related to standards-based reform and students with disabilities, building consensus and capacity among practitioners and policymakers. Her current research and technical assistance priorities include alternate assessment of students with significant disabilities and research focused on the causes of and solutions for the achievement gap between students with disabilities and their typical peers. She is the author of numerous chapters, articles, presentations, and papers related to inclusive assessment of students with disabilities, and has coauthored a book on alternate assessment. She is the proud parent of an adult daughter who has Down syndrome.

¹ K. S. McGrew and J. Evans, *Expectations for Students with Cognitive Disabilities: Is the Cup Half Empty or Half Full? Can the Cup Flow Over?* (Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes, 2003).

² Individuals with Disabilities Education Act Amendments (IDEA) of 1997, Public Law 105-17, 105th Cong., 1st sess.

³ Improving America's Schools Act of 1994, Public Law 103-382, 103rd Cong., 2nd sess.

⁴ No Child Left Behind Act of 2001, Public Law 107-110, 107th Cong., 1st sess.

⁵ Individuals with Disabilities Education Improvement Act (IDEA) of 2004, Public Law 108-446, 108th Cong., 1st sess.

⁶ M. Johnson, L. Thurlow, and K. E. Stout, *Revisiting Graduation Requirements and Diploma Options for Youth with Disabilities: A National Study* (Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes, 2007).

⁷ M. Perie, *Understanding the AA-MAS: How Does It Fit into a State Assessment and Accountability System?* presentation to CCSSO SCASS meeting, February 4, 2009, http://nceia.org/cgi-bin/pubspage.cgi?sortby=pub_date (accessed March 17, 2009).

⁸ J. Minnema, M. Thurlow, and S. H. Warren, *Understanding Out-of-Level Testing in Local Schools: A First Case Study of Policy Implementation and Effects* (Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes, 2004).

⁹ Ibid.

¹⁰ K. Boundy and J. Karger, "Special Issues Affecting Inclusion of Students with Dyslexia in Statewide Assessments and Their Implications," *Perspectives on Language and Literacy International Dyslexia Association* 34, no. 4 (2008): 36–40.

¹¹ Perie, *Understanding the AA-MAS*.

¹² U.S. Department of Education, *Raising Achievement: Alternate Assessments for Students with Disabilities*, <http://www.ed.gov/print/policy/elsec/guid/raising/alt-assess-long.html> (accessed March 8, 2009).

¹³ H. Kleinert, S. Kennedy, and J. Kearns, "Impact of Alternate Assessments: A Statewide Teacher Survey," *Journal of Special Education* 33, no. 2 (1999); M. Turner, L. Baldwin, H. Kleinert, and J. Kearns, "An Examination of the Concurrent Validity of Kentucky's Alternate Assessment System," *Journal of Special Education* 34, no. 2 (2000).

¹⁴ U.S. Department of Education, Office of Special Education and Rehabilitative Services, *Learning Opportunities for Your Child Through Alternate Assessments* (Washington, DC: U.S. Department of Education, 2007).

¹⁵ D. M. Browder, S. L. Gibbs, L. Ahlgrim-Delzell, G. Courtade, A. Mraz, and C. Flowers, "Literacy for Students with Significant Cognitive Disabilities: What Should We Teach and What Should We Hope to Achieve?" *Remedial and Special Education*, in press; C. Flowers, S. Wakeman, D. Browder, and M. Karvonen, *Links for Academic Learning: An Alignment Protocol for Alternate Assessments Based on Alternate Achievement Standards* (Charlotte, NC: University of North Carolina at Charlotte, 2007); see <http://www.nceo.info>; C. Flowers, S. Wakeman, D. Browder, and M. Karvonen, "An Alignment Protocol for Alternate Assessments Based on Alternate Achievement Standards," *Educational Measurements: Issues and Practice*, in press.

¹⁶ Perie, *Understanding the AA-MAS*.

¹⁷ Donahue Institute, *A Study of MCAS Achievement and Promising Practices in Urban Special Education: Report of Field Research Findings (Case Studies and Cross-Case Analysis of Promising Practices in Selected Urban Public School Districts in Massachusetts)* (Hadley, MA: University of Massachusetts, Donahue Institute, Research and Evaluation Group, October 2004), <http://www.donahue.umassp.edu> (accessed March 16, 2009).

¹⁸ S. J. Thompson, C. J. Johnstone, M. L. Thurlow, and J. R. Altman, *2005 State Special Education Outcomes: Steps Forward in a Decade of Change* (Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes, 2005).

¹⁹ M. L. Thurlow, R. F. Quenemoen, S. S. Lazarus, R. E. Moen, C. J. Johnstone, K. K. Liu, L. L. Christensen, D. A. Albus, and J. Altman, *A Principled Approach to Accountability Assessments for Students with Disabilities* (Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes, 2008).

²⁰ M. Thurlow, R. Quenemoen, S. Thompson, and C. Lehr, *Principles and Characteristics of Inclusive Assessment and Accountability Systems* (Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes, 2001); M. Thurlow, R. Quenemoen, J. Altman, and M. Cuthbert, *Trends in the Participation and Performance of Students with Disabilities* (Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes, 2007).