
Considerations for the Alternate Assessment based on Modified Achievement Standards

Section I

Identifying and Understanding the Population and Their Curriculum

**Chapter 2: Identifying Students and Considering Why and
Whether to Assess them with an Alternate
Assessment Based on Modified Achievement
Standards**

**Chapter 3: Developing Standards-Based IEPs that Promote
Effective Instruction**

**Chapter 4: The Challenges of Conceptualizing what Low
Achievers Know and How to Assess Their
Competence**

The contents of this publication were developed under cooperative agreement S283B050019 with the U. S. Department of Education. However, the contents do not necessarily represent the policy of the Department of Education, and you should not assume endorsement by the Federal Government.

For the full version of this white paper, visit:

http://nycomprehensivecenter.org/initiatives/inits_sp_assessment



University of the
State of New York
State Education
Department

SECTION I

IDENTIFYING AND UNDERSTANDING THE POPULATION AND THEIR CURRICULUM

The first challenge is to determine who the students are who are in need of a new assessment. The focus is on students with disabilities who are not achieving proficiency on grade-level standards and who do not appear to be making significant progress towards achieving that proficiency. But beyond that, it is important to explore various aspects of these students, including the nature of their disability and why it might hinder learning. And, we would be remiss not to explore the issues of curriculum and instruction to see whether opportunity to learn is having a larger impact on performance than the nature of the disability.

These three chapters tie together these ideas to describe the population in terms of who they are, necessary elements of their instruction, and how they learn. More specifically, each chapter delves into different theories involving the fluidity of this population. Chapter 2, by Rachel Quenemoen, focuses on the notion of the least dangerous assumption by considering exclusionary criteria. It provides a history of regulations regarding students with disabilities and discusses applications of the current regulations to the school environment.

In Chapter 3, Meagan Karvonen focuses on procedural integrity by providing an overview of standards-based individualized education programs (IEPs) and describing how to promote improved opportunities to learn the standards-based curriculum with specialized instruction, services, and supports based on individual student learning characteristics. Then, Chapter 4, by Jim Pellegrino, helps us understand explanatory constructs by discussing the broader understanding of student cognition, describing possible sources of differences among students that have implications for learning, instruction, and assessment. It provides information on student cognition and explores issues related to barriers to learning for low achievers.

Understanding who the students are, what and how they are taught, and identifying any barriers to learning is a first step towards understanding how best to assess what they know and can do. This section provides the backbone for the later sections on test development and technical issues.

This section is stronger because of the insightful comments of the expert panel members who reviewed these chapters. In particular, comments from Claudia Flowers, Gerald Tindal, Brian Gong, and Suzanne Lane were incorporated into these chapters.

CHAPTER 2

IDENTIFYING STUDENTS AND CONSIDERING WHY AND WHETHER TO ASSESS THEM WITH AN ALTERNATE ASSESSMENT BASED ON MODIFIED ACHIEVEMENT STANDARDS

Rachel Quenemoen

This chapter presents some of the complex issues that need to be considered when identifying students who may benefit from participating in an alternate assessment based on modified achievement standards (AA-MAS). The chapter starts with a historical perspective of the regulatory language creating AA-MAS. It informs readers of the initial rationale for creating AA-MAS and the concerns of advocacy groups. The preliminary requirements for identifying students who may be eligible for participation in AA-MAS are also discussed. Research findings about low-performing students are introduced to provide the readers with an understanding that low-performing students are students with and without disabilities. These findings also illustrate that identifying students who will benefit from participating in AA-MAS requires much more than knowing the students' previous large-scale test performance or the students' disability category. While student characteristics are important, they are only part of the consideration for determining AA-MAS eligibility. A discussion of teacher perceptions of student characteristics and opportunity-to-learn issues is followed by potential best-practice interventions and instructional practices (i.e., Response to Intervention (RtI), progress monitoring) to provide readers with information on strategies that may benefit all students. It is followed by an examination of policy assumptions about instructional and curricular strategies used with students and how they relate to assessment choices. Ultimately states will make policy decisions that will define students who may participate in AA-MAS. These policy decisions are framed with a discussion of social justice, guiding philosophy, and coherence of the overall instruction, curriculum, and assessment systems. The chapter ends with a set of questions for

states to answer as they consider their options and potentially develop and implement an AA-MAS.

The chapter is written in the context of the imperatives of a system accountability model since the AA-MAS was conceptualized initially as an option within such a model. That is, this chapter assumes that an AA-MAS is a key component of a policy that is designed to improve student achievement and to narrow achievement gaps that have affected certain groups of students differentially over time. The underlying policy assumes that poor performance by students on the state assessment will result in consequences for schools and districts that will motivate educators to provide better services to students, services that will enable them to learn and achieve to proficiency. The path to these improvements is through improved instruction and curriculum, although that implication seems to be missing in many discussions about assessments used for system accountability. Because of this essential but sometimes neglected component of system accountability, examples of standards-based instruction and curriculum strategies and interventions are included in this chapter (as well as the next, Karvonen, Chapter 3, this volume) to augment this volume's focus on the assessment component of the policy imperative. Possible validity-related questions regarding the relationship of high quality standards-based instruction and curriculum to achievement of students with disabilities on standards-based assessments are posed in the concluding section, and examples of studies that have uncovered these relationships are cited (e.g., Barr, Telfer, & DiMuzio, 2009; Cortiella & Burnette, 2007; Donahue Institute, 2005).

History of Assessment Options Related to AA-MAS

In order to understand which students may meet the requirements for participation in an AA-MAS, it is important to begin with a review of the policy discussion that framed the initial regulation. There was immediate and intense debate surrounding the announcement of the

proposed regulation, primarily focused on the research USED cited as the rationale. This debate continues as states study whether or how they will implement these assessments.

In April 2005, addressing a group of chief state school officers and other officials, Secretary of Education Margaret Spellings announced new flexibility in assessing students with disabilities under No Child Left Behind (NCLB) regulations. Secretary Spellings called it a "workable, sensible approach that was based on scientific research," permitting states to develop and use modified assessments for students with "persistent academic disabilities." These students were defined as those "who need more time and instruction to make substantial progress toward grade-level achievement." The research base cited was summarized and sent to all chief state school officers. These materials began with a reference to the earlier (2003) NCLB regulation permitting alternate achievement standards for students with significant cognitive disabilities, defined first in a notice of proposed rules in 2002, and finalized in 2003.

The 0.5% cap originally included in the August 2002 proposed regulation [1%] was based on data outlining the prevalence rates of students with the most significant cognitive disabilities. It was tied to a definition of such students which: 1) excluded students with mild mental retardation and other students who were two or fewer standard deviations below the mean, and 2) included students with intellectual functioning and adaptive behavior three or more standard deviations below the mean. When this rule was finalized, the Department expanded the cap to 1.0% to allow States and districts more flexibility in its implementation and removed the definition from the regulation. However, research conducted and reviewed by Reid Lyon at National Institute for Child Health and Human Development and Jack Fletcher at the University of Texas indicates that the 1.0% cap is, in fact, too low, if the Department follows the definition currently provided in the December 2003 regulation's preamble (a student in one of 13 disability categories who cannot reach grade-level standards, even with the best instruction possible) (USED, 2005).

The USED provided a summary of research that supported this increase in students who could participate in alternate assessments against less challenging achievement standards. In the research summary, Lyon and Fletcher found that “the best-designed instructional interventions achieved a range of success from a low of 50% to a high of 90% of participating students reaching grade-level reading standards.” They concluded, that the “totality of this research suggests that there are about 1.8% to 2.5% of children who are not able to reach grade-level standards, even with the best instruction” (USED, 2005).

Advocates for students with disabilities responded to the proposed new flexibility with concern. Central to their concern was the fear that students who participate in an assessment based on a lower standard will also receive instruction in a lower curriculum. The implication was that the option of modified achievement standards would limit struggling students’ access to academic instruction and needed research-based interventions to accelerate learning and, over time, preclude their attainment of a standard diploma. One advocacy group, the National Council for Learning Disabilities (NCLD), critiqued the research base summarized by Lyon and Fletcher for USED, noting that although the research on effective reading interventions cited was important for remediation and for new methods for identification of learning disabilities, these reading intervention studies did not support the federal policy changes proposed in the new regulation (Wendorf, 2005). More recently, NCLD has concluded: “The studies that were originally used by the U.S. Department of Education in 2005 to justify the 20 percent number were flawed. In fact, in one of the major studies cited to justify the new policy, only 11 percent of the students were special education students and the additional studies cited did not include any special education students” (Kaloi, 2007).

A number of the requirements in the final regulation reflect compromises that resulted from this vigorous debate (e.g., students must have standards-based IEP goals; students are not precluded from earning a standard high school diploma), but the controversy related to use of modified achievement standards—and implications of use for student achievement—remains.

Given the controversy, state policymakers need to grapple with whether they believe this is a distinct group, separate from both the group of students defined in the “1%” regulation as having significant cognitive disabilities and from other students with disabilities. State policymakers, educators, and advocates had far less difficulty in coming to consensus on the appropriateness of alternate assessments based on alternate achievement standards (AA-AAS) as a pathway to higher expectations and achievement. In many ways, the students referenced in the 1% regulation were, by and large, unarguably a distinct group, albeit a heterogeneous group of students with unique characteristics, many with complex disabilities. However, in implementation, some students may be inappropriately included in AA-AAS instead of a more challenging assessment. Historical low expectations affect decision-making, as do past performance patterns of students who have not been taught the content to be assessed. Students who are inappropriately included in an AA-AAS may be harmed by assumptions that they cannot learn the full range of grade-level academic content when the result is that they will never be taught that content, leading to a self-fulfilling prophecy.

As quoted from the USED (2005) 2% materials, the earlier 1% figure was established as a compromise; the estimates of how many students may have the most severe intellectual and multiple disabilities initially (.5%) 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. Thus, the 1% cap on inclusion of scores from AA-AAS as proficient for AYP calculations incorporates some flexibility already, but from a policy perspective to balance the flexibility, the cap was intended to prevent inappropriate inclusion of too many students in a different achievement expectation. The controversies of AA-AAS tend to be about the nature of the content being taught and assessed and the technical issues related to test design, but not whether different achievement standard(s) could be an appropriately high expectation for students with significant cognitive disabilities.

In contrast, there is limited consensus on students referenced in the 2% regulation. Some policymakers reference students who are “just above” the students who participate in AA-AAS, with an achievement expectation far below the grade-level expectation, perhaps adding to the students already included in AA-AAS through the flexibility of the compromise 1% cap. Others suggest that these students are those who “just miss” the proficiency determination on the general assessment, and the modified achievement standard should be “just below” the grade-level achievement standard. Yet another interpretation is that students with disabilities perform on a continuum with no defined borders; thus, additional achievement standards may be necessary in order to count more students as proficient against multiple standards on a sliding scale set. The regulatory language and research base referenced in the regulation are not clear about the target population for AA-MAS.

Ultimately, state-defined modified achievement standards should be a policy statement of what is an appropriately high expectation for some state-defined group of students, an expectation that should improve their achievement and outcomes in order to be consistent with the letter and the spirit of NCLB and IDEA. It is essential that state policymakers articulate who the target students are and how they build competence in the academic domains tested prior to deciding whether and how to develop an assessment based on modified achievement standards. Then, decisions about the design of the assessment itself can adhere to the policy imperatives, instead of the assessment choices inadvertently shaping the policy outcomes.

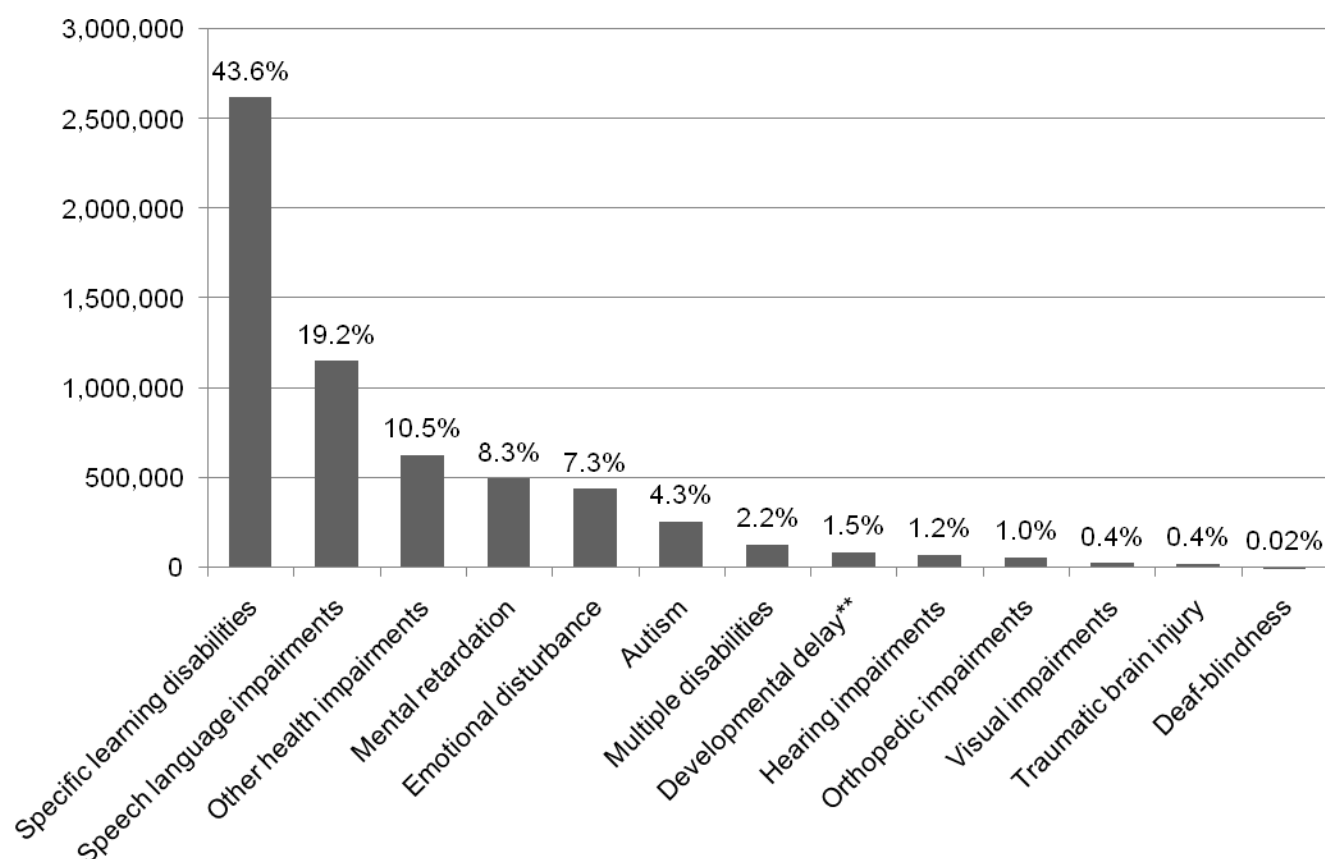
Who Are the Students? The Complexity of Regulatory Requirements

The regulation specifies two primary requirements for participation in an AA-MAS. The student must be identified as having a disability that precludes attainment of grade-level achievement standards within the current year; and the student must have an IEP that references grade-level content.

Disability Categories Overview: Limitations of Categorical Designations to Predict Attainment of Standards

As cited in the research summary underlying the regulation, students with disabilities are defined as having a primary disability label under 13 categories. The learning characteristics of students who are assigned to these categories vary greatly both among and within the categories. See Figure 2-1 for a summary of categorical distribution of students with disabilities, by primary disability.

Figure 2-1. Distribution of Primary Disability Categories*



* Percentages in this figure are based on a total number of 6,007,832 students receiving special education services (www.ideadata.org, 2007) counted under primary disability only.

** Developmental delay is applicable only to children ages 3 through 9.

Based on the 2007 IDEA Part B Child Count data in the United States and outlying areas (www.ideadata.org, 2007), 43.6% of students received special education services for specific learning disabilities (see Figure 2-1). The next largest disability group is speech or language

impairments totaling 19.2%, followed by the category of students with other health impairments at 10.5%, mental retardation at 8.3%, and emotional disturbance at 7.3%. Students with autism make up 4.3% of students served in special education, and students with multiple disabilities make up 2.2% of these students. Smaller categories of students in special education include students with developmental delay at 1.5% (category for ages 3-9 only), hearing impairments at 1.2%, and orthopedic impairments at 1.0%. Students with visual impairments and traumatic brain injury each make up 0.4% of students served by special education, while the remaining 0.02% of students make up the deaf-blindness category.

The criteria used to determine student eligibility for special education and related services under IDEA are defined by each state (within certain federal parameters), and thus the criteria vary from state to state. Although a few categories include criteria that are relatively objective (e.g., vision, hearing, and physical characteristics), all categories include more subjective judgment as well. Some categories (e.g., specific learning disabilities and mental retardation) have widely varying criteria, subject to interpretation in multiple ways. It is common within and across states to have the criteria yield students who have very different learning characteristics sharing the same label and students who are very similar to one another having different categorical labels. The National Association of School Psychologists (2002) position statement on categorical labels summarizes research on categorical labels thus:

- State by state and district by district variations exist in the definition and criteria for specific disability conditions, despite the common language of IDEA and its regulations;
- Particularly among the more subjective, “mild” disability categories of Specific Learning Disability, Mental Retardation, Emotional Disturbance, and Speech/Language Impairment, labeled students show significant overlap in skills and receive highly similar instruction;
- Among students with very low achievement, there are no consistent distinctions between those identified as disabled (e.g., SLD) and those who are considered “Slow Learners”;

- Regardless of instructional needs, students tend to be placed in programs based on labels; and
- Despite increased opportunities for inclusion in the general education program, students who are labeled as having a disability are less likely to have general education friends, are less likely to have instructional goals tied to the general education curriculum, are more likely to drop out of school and have lower rates of successful adult outcomes (NASP, 2002).

These complexities related to disability categorical labels and criteria for determining eligibility for special education services affect some groups of students more than others. For example, English Language Learners (ELLs) are disproportionately identified as also having disabilities in many states, and in many states, ELLs with disabilities are the lowest performing group overall. The challenges of determining whether a student is eligible for special education services in the context of limited English proficiency (or culture, ethnic, or socioeconomic status) is additional reason to use caution in assuming that special education categorical labels are useful for purposes of identifying students for assessment options. See Artiles, A. J., & Ortiz, 2002; Abedi, 2006; Abedi, 2007; Minnema, Thurlow, Anderson, & Stone, 2005 for more information on these learners.

Still, some general comparisons can be made about the nature of the disability categories. For example, the relationship of categorical label to 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, other health impairments, and emotional/ behavioral disabilities. These students, along with those who have physical, visual, and hearing 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 total student population, or *about 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. (Thurlow, 2007)

The reality that many students with disabilities currently do not achieve at proficiency raises questions of whether or not they are all receiving the required high quality instruction in the grade-level content, appropriate services and supports, and appropriate accommodations.

Beyond that type of general observation across disability categories, it is difficult to define how the specific categorical labels differentiate how students learn and demonstrate their learning, or how, as required in the regulation, this disability prevents them from attainment of grade-level achievement within the current year.

IEPs and Access to the General Curriculum

The second primary requirement for states to meet in designing participation criteria for the AA-MAS (in addition to being identified as eligible to receive special education services under IDEA) is the requirement that the student must have an IEP that references grade-level content. In some ways, this requirement is redundant of the foundational requirements of being eligible to receive special education services. That is, having access to the curriculum in order to meet the educational standards of the public agency is how special education is defined in IDEA: *“specially designed instruction, at no cost to parents, to meet the unique needs of a child*

with a disability, including instruction conducted in the classroom, in the home, in hospitals and institutions, and in other settings ...” [20 U.S.C. §1401 (29)]. Specially-designed instruction is defined as “adapting, as appropriate to the child’s needs, the content, methodology, or delivery of instruction to address the unique needs of the child that result from the child’s disability; to ensure access of the child to the general education curriculum, so that the child can meet the educational standards within the jurisdiction of the public agency that apply to all children” [34 CFR §300.39 (b)(3)].

These definitions are not new to IDEA 2004. In 1999, disability rights attorney Paul Weckstein wrote, “An IEP that sets lower goals and does not focus on these standards [that is, the educational standards of the public agency] is usually illegal. Nor is it generally legal to assign a student with disabilities to a low-track regular program that does not teach to these standards. These rights are protected by the federal IDEA and Section 504 of the Rehabilitation Act of 1973” (Weckstein, 1999, p. 314). These rights were reinforced through NCLB in 2001 and the reauthorization of IDEA in 2004, according to Karger and Boundy (2008):

These two statutes [NCLB 2001 and IDEA 2004] together have lifted expectations for learning and have underscored the legally enforceable rights of students with disabilities to be effectively taught by highly qualified teachers, to be provided an opportunity to learn to the same high standards as their peers without disabilities, and to be included in all state and district-wide assessments. Protections provided under Section 504 of the Rehabilitation Act of 1973 (Section 504) and the Americans with Disabilities Act (ADA) and grounded in the Fourteenth Amendment of the U.S. Constitution also ensure that students with disabilities are not subject to discriminatory policies and practices. Rather, these students must be provided meaningful opportunities to learn the knowledge and skills necessary to attain proficiency on their respective state standards; full and fair opportunities to demonstrate their level of mastery of state standards through participation in

appropriate assessments used to improve their instruction and learning; and equal opportunities to be counted in the publicly reported data system that is used to hold schools, districts, and states accountable for the academic performance of all students. (p. 11)

Even though IDEA has required that all students who receive special education services should be provided the services, supports, and specialized instruction so that they achieve proficiency on the state standards, many students with disabilities have not been receiving that instruction. In 1984, special education researcher Anne Donnellan wrote that “the criterion of least dangerous assumption holds that in the absence of conclusive data, educational decisions ought to be based on assumptions which, if incorrect, will have the least dangerous effect on the likelihood that students will be able to function independently as adults.” (p. 142). She concluded that barring proof to the contrary, educators need to assume that poor performance is due to instructional deficits instead of student deficits. The regulation requires that IEP teams examine objective evidence demonstrating that the student’s disability has precluded the student from achieving proficiency and the guidance suggests that “Such evidence may include the student’s performance on State assessments or other assessments that can validly document academic achievement” (USED Non-regulatory Guidance, 2007, p. 17). IEP teams will have to determine that such evidence is sufficient to ensure that instructional deficits are not the cause of low performance, as opposed to the student’s disability. Using data from large-scale assessments of content that the student has not been taught seems to result in circular logic for this purpose. That is, documenting student academic achievement on content that has not been taught by an assessment of that content tells us nothing about whether or how the student’s disability precluded their achievement, and only tells us what the student knows and can do prior to instruction. Later in this chapter and in the next chapter (Karvonen), there are discussions of methods for documenting the effectiveness of instructional and curricular

strategies that could be used to ensure that instructional deficits are not the cause of low performance.

The specification in the regulation that students who participate in an AA-MAS must have standards-based IEPs was meant to assuage the concerns raised by advocates that a less challenging achievement standard would result in further inappropriate instruction in a lower level track. Whether or not the stipulation was in the regulation, states have an obligation to make sure that all students with disabilities are receiving the services, supports, and specialized instruction necessary for them to make progress in—and achieve proficiency in—the curriculum based on the state standards defined for all students. Unless there is assurance that has occurred, the least dangerous assumption is that poor performance relates to poor opportunities to learn. See Karvonen, chapter 3, this volume for more information on how the IEP process can be used to improve opportunities to learn.

Operationalizing the Regulatory Language: Identifying Who May Be Eligible

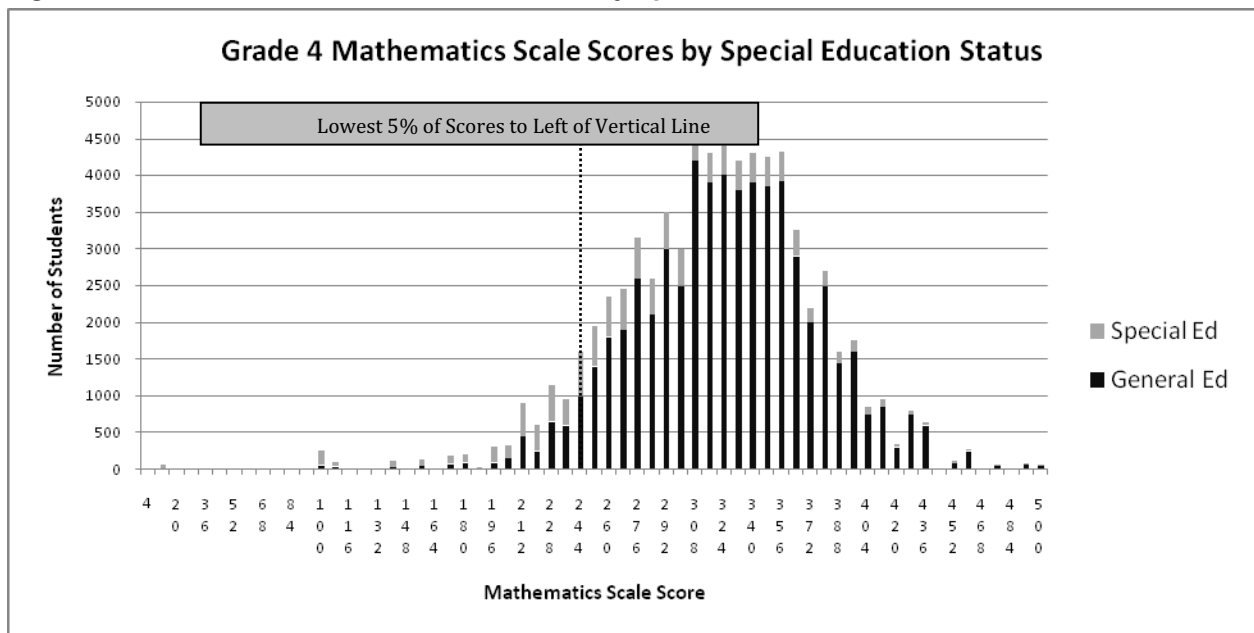
Since the AA-MAS regulation was finalized, states have struggled to identify the students who are low performing and might be eligible for this assessment. Additionally, a few researchers have attempted to understand what opportunities to learn the low-performing students have had. Given the initial controversies about the research base, the limited utility of categorical labels, the necessity of ensuring access to the general curriculum, and the lack of agreement in the field of who these students are, these studies have been challenging. There are debates about whether states should identify the students based on a percentage (i.e., the 2% of students with disabilities with the lowest scores), or whether studies should be based on the characteristics of the population and their instruction in the content defined in the regulation (i.e., those taught the curriculum but not likely to be proficient that year). Most studies have taken the former approach, mining state assessment data for students who perform at the lowest end, although data from opportunity-to-learn investigations are increasingly challenging

the assumptions that all low- performing students have indeed been taught the standards-based curriculum.

Studies of Low-Performing Students

States have tried to operationalize who the eligible students are in varying ways (Fincher, 2007; HB Study Group from Colorado, 2005; Marion, Gong, & Simpson, 2006; New England Compact, 2007). One of the first investigations looking at how students with disabilities currently perform on large-scale assessments under NCLB was done by researchers at the National Center for the Improvement of Educational Assessment (NCIEA). As indicated in Figure 2-2, the scores of students with disabilities occur at all scale scores in the distribution and the scores of students without disabilities also occur at all scale scores in the distribution (Marion, Gong, & Simpson, 2006).

Figure 2-2: Grade 4 Mathematics Scale Scores by Special Education Status



Reprinted by Permission

This study foreshadowed results of studies in multiple states: the lowest performing students on state assessments under NCLB are not all students with disabilities. For example, Perie (2009) summarized data mining approaches in two states, Georgia and South Carolina.

Georgia mined data from three years of the state test, identifying persistent low performers in grades 5 and 8 as students with three years of data scoring in the lowest of three achievement levels. South Carolina looked at grades 4 and 7, identifying students with two years of data scoring in the lowest of four levels. In Georgia, the percentage of persistently low-performing students who have documented disabilities ranged from 40% to 55%; in South Carolina, the percentages of students with disabilities among the lowest performers ranged from 39% to 49%. (Perie, 2009). These data will vary based on the nature of states' proficiency standards and depending on the numbers of years for which data are available. That is, two years of data may show different patterns than four years of data. The studies vary in methodology and findings, and many of the research reports are included in the accompanying resources list. Chapter 10 (Domaleski, this volume) addresses ways states can learn from these efforts to design their own data-mining study.

Student Characteristics and Opportunity-to-Learn Investigations

Several states have attempted to understand more about the educational characteristics of these low-performing students. Perie (2009) summarized results of teacher surveys and focus groups in several states that captured teacher perceptions about the nature of these students' learning. These findings of teacher perceptions indicate: "that the core academic curriculum is significantly modified or specifically designed for the student; the student is making fairly consistent progress but not at expected (or targeted) level; there is a gap between actual performance and targeted level of performance which is evident over a period of time (at least 2 consecutive years); the gap continues to widen or remains the same; despite the provision of "good" interventions, the student is not progressing at the rate expected for grade level; accommodations alone do not allow the student to fully demonstrate knowledge; and all appropriate accommodations have been exhausted."

Georgia has conducted a curriculum implementation survey of teachers that asked whether low-performing students were receiving instruction in the grade-level curriculum and to what depth and degree. Teachers self-reported their instructional practices and curriculum choices; the investigators suggest that teacher interviews and direct classroom observations would be preferable but not feasible with resources available. Although the results of the study are not yet published, their initial findings suggest that at fifth and eighth grade in mathematics and at eighth grade in reading, general education teachers have higher expectations (deeper levels of understanding) for students than do special education teachers, but fifth grade reading responses showed higher expectations among special education teachers (Fincher, 2009). Qualitative data are being used to help illuminate these differences, but results are not yet available.

The New England Compact study of the gap included interviews with teachers, and substantial direct quotations from teachers. Compare and contrast the following teacher observations of similar students (Parker & Saxon, 2007):

Teacher 1: They tend to be slow learners. They tend to be “shady 80s,” that is what I call them. Seventy to 75 makes you mentally retarded or learning impaired. If you are in the 90s, then you are okay. These are shady 80s. They show up every day for school, and they sit down and crank out their little homework. They don’t have a clue what the homework means, but they have it done. They always have a notebook. They always have a sharpened pencil . . . These guys are good students in the classroom. They have their notebook and their pencil. Clueless. They have no mechanism to practice it . . . They never move from that very pretend area of teaching. They do it fairly well. These kids are going to get 70s. These kids will get it right, but they don’t have a clue how they got it right. It never becomes theirs. (p. 7)

Teacher 2: My teaching practice just in the past year has changed dramatically. My thrust now is to really concentrate on eighth grade GLEs [grade-level expectations],

even though most kids I have in my resource room are third grade level for math, maybe fourth grade . . . I found math strategies presented on the third, fourth grade level . . . they were exposed to strategy on their level, so then we worked through problems up to the eighth grade level. So that's a new direction for me, because I'm not sure I've always had the expectation that they could do eighth grade math. So my expectation has changed, and my teaching practice as a result of that has changed. (p. 8)

Clearly, as evidenced by the first teacher, not all educators are implementing the least dangerous assumption related to their expectations for these students. In some cases, refocusing attention on the needs of struggling learners changes teacher behavior (Teacher 2), but that is not always the case (Teacher 1).

At the request of the Colorado State Legislature, a Colorado study group reviewed their reading and math data (grades 3–10) and found that not all of the students performing in the lowest one-third on the state tests were students receiving special education; accommodations were not consistently provided to all eligible students; and they saw “substantial longitudinal growth” toward grade level achievement for the majority of the students over time. When the study group initiated actual observations of instructional opportunities, they found that students (with and without disabilities) who were making the greatest gains toward grade-level achievement were those attending schools that provided “intensive, targeted, research-based instruction” (HB Study Group, 2005). The next section elaborates on possible strategies to improve opportunities to learn and to identify students who may benefit from participation in AA-MAS.

Strategies to Improve Opportunities to Learn and Identify Students Who May Benefit from AA-MAS

There are studies focused on what is occurring in schools where students with disabilities are performing well (e.g., Barr, Telfer, & DiMuzio, 2009; Cortiella & Burnette, 2007;

Donahue Institute, 2005). These studies consistently identify common characteristics among schools where students with disabilities achieve at high levels. As summarized in one study, the schools have: (1) a pervasive emphasis on the curriculum and alignment with the standards, (2) effective systems to support curriculum alignment, (3) an emphasis on inclusion and access to the curriculum, (4) a culture and practices that support high standards and student achievement, (5) a well-disciplined academic and social environment, (6) continuous use of student data to inform decision making, (7) unified practices supported by targeted professional development, (8) access to resources to support key initiatives, (9) effective staff recruitment, retention, and deployment, (10) flexible leaders and staff that work effectively in a dynamic environment, and (11) effective leadership (Donahue Institute, 2004).

A recommendation from the Colorado study of students in the gap was that the state should implement sound “data-driven recommendations that focus on student learning and on valid measurement of that learning.” The implementation of Response to Intervention (RtI) strategies in many states is meant to ensure that research-based early literacy screening and early intervention processes are used to help identify struggling learners as soon as possible. According to the National Center on Response to Intervention, “Response to intervention integrates assessment and intervention within a multi-level prevention system to maximize student achievement and to reduce behavior problems. With RtI, schools identify students at risk for poor learning outcomes, monitor student progress, provide evidence-based interventions and adjust the intensity and nature of those interventions depending on a student’s responsiveness, and identify students with learning disabilities.” (See <http://www.rti4success.org>). In addition, RtI can help identify students with learning disabilities sooner so they can achieve more. This approach is not limited to students with disabilities and can assist states like Colorado that have a commitment to improve instruction and outcomes for all students identified as low-performing, with and without disabilities. As such, RtI holds

promise for states that choose to emphasize intervention on instruction and curriculum as opposed to relying solely on large-scale assessments to improve achievement and outcomes.

There is strong research and best practices documentation for RtI in reading, especially in the primary grades. The Institute of Education Sciences (IES) has published a practice guide based on currently available evidence, explicating five recommendations for implementing RtI to identify students in need of intervention. It further describes how to carry out each recommendation and identifies potential roadblocks to implementation. (Gersten, Compton, Connor, Dimino, Santoro, Linan-Thompson, & Tilly, 2008). The authors note that while multi-tier efforts like RtI can prevent learners from falling behind through early implementation of interventions, they also note that “some aspects of RtI, however, (such as tier 1 instruction) are still poorly defined, and there is little evidence that some practices of targeted instruction will be effective” (p. 8). Still, they suggest that a coordinated multi-tier program can prevent beginning readers from becoming struggling readers in the later grades, and possibly prevent referrals to special education. They provide an exhaustive references list to support their recommendations, and categorize the recommendations based on the strength of the evidence in the research base. (Gersten et al., 2008).

Fuchs and Fuchs (2006) also provide helpful guidance to states and districts considering use of RtI related to reading. They document differences among educators about the appropriate use of RtI, noting “The first group views RtI mostly in terms of providing prevention and advocates for more tiers. The second group regards RtI mostly as identification and classification procedures and argues for fewer tiers” (p. 94). In addition, they suggest that practitioners and researchers vary in their preference for application of RtI, with practitioners viewing RtI as a problem-solving approach while researchers favor the use of standard treatment protocols. They identify numerous unanswered questions and unresolved issues, including the challenge of false positive identification for special education services in use with a problem solving approach and false negative identification in use with a standard treatment

protocol approach (Fuchs & Fuchs, 2006). They question which error is worse, but do not answer the question. Ultimately, that becomes a critical policy decision to be made at the local and state levels in implementation, and something that must be monitored closely.

Fletcher (2008) summarizes issues for consideration that may serve to inform efforts in states to grapple with defining students who meet the regulatory requirements for participation in an AA-MAS. He notes that Rtl is not appropriate solely as a special education initiative or method to meet criteria for identification of learning disabilities; it ultimately is a regular education initiative for all students. Once students are identified as meeting criteria for special education services based on not responding well to interventions, Fletcher suggests that does not give us the information needed to understand why they are not responding or how to teach or assess them. Echoes of advocates' concerns about the strength of the research base under the regulation seem to be borne out in his conclusion that "more research is needed on the characteristics of students who do not respond well to intervention since we have not really had the opportunity to study this subgroup from cognitive, interventional, and neurobiological perspectives" (p. 9). The regular education basis for Rtl also has potential benefit for states that feel an obligation to intervene on behalf of all low-performing students who do not have disabilities. Rtl processes have the potential to improve outcomes for all struggling students, not just those with disabilities.

Other progress monitoring approaches also may contribute to improved achievement for all students. Curriculum-based measurement (CBM) is of particular interest given the strong research base for many CBM methods. Recently, Fuchs, Seethaler, Fuchs, and Hamlett (2008) proposed CBMs as meeting the requirements for determining eligibility under the 2% regulation: "That is, CBM progress monitoring can be used to provide the necessary database on (a) whether grade-level proficiency is expected, (b) whether appropriate instruction has been provided, and (c) whether progress in response to that instruction is appropriate. Moreover, as it satisfies the three-pronged requirement for evidence related to identifying the 2% population,

CBM progress monitoring simultaneously provides the added advantage of helping schools enhance special education outcomes” (p. 160). Given that CBMs often are used within an RtI framework to monitor progress following an intervention, these authors suggest that efficiencies of scale will emerge that produce data to determine eligibility for an AA-MAS.

Placing these approaches within the context of standards set for the grade level, as required in current standards-based systems, is essential. Deno, Fuchs, Marston, and Shin (2001) discuss the implications of a normative approach to setting achievement expectations that assume that the current typically observed growth rates for students with disabilities are reasonable and predict future growth. These assumptions lead to the conclusion that students with learning disabilities will learn at a slower rate than typical peers. The researchers speculate that this kind of reasoning reflects the "well-accepted fact that special education, as typically practiced in this country, fails to regularly incorporate demonstrably effective methods" (Deno et al., 2001, p. 515). If this is true, then systemic interventions on the system of special education in this country to correct these deficits should be a priority in every state, as opposed to accepting lower rates of student learning.

Response to intervention and other progress monitoring approaches hold much promise for improved outcomes and higher expectations. Still, critical contextual challenges must be addressed. These challenges affect the implementation of effective progress monitoring for students with disabilities. They include historical limited access to challenging standards-based curriculum, instruction, and assessment; concerns about the target of measurement, that is, whether only basic skills or a full range of rich and challenging grade-level content should be measured; and limited practitioner understanding about use of data for effective provision of instructional strategies, interventions, and supports in a standards-based system (Quenemoen, Thurlow, Moen, Thompson, & Morse, 2003). In most schools, the path to assurance of the least dangerous assumption requires guideposts of continuing staff development, support, and oversight.

Even with the identification (and improvement) of opportunities to learn through instructional quality and curriculum access, the validity of assessments hinges on whether all students who have learned the grade-level content can show what they know on the assessments. The National Accessible Reading Assessment Projects (NARAP) are conducting a program of research and development designed to make large-scale assessments of reading proficiency more accessible for students who have disabilities that affect reading. They suggest that creating “accessible reading assessments based on accepted definitions of reading and proficiencies of reading requires knowledge of the issues specific to each disability and how they affect reading and the assessment of reading.” They have prepared a series of papers by categorical label to serve as discussion guides for partners working on test development, summarized in Thurlow, Moen, Liu, Scullin, Hausmann, and Shyyan (2009). This may serve as a resource to state stakeholders and consultants as they grapple with how to first teach and then assess students with varying characteristics. Still, there are limits to what discussions based on categorical labels will yield, given the well-documented subjectivity of state-defined eligibility criteria used to determine categorical labels and the heterogeneity of students within each category.

Regulatory Options, Policy Prerogatives, and Implications

State policymakers have a great deal of flexibility in decisions about whether and how to implement AA-MAS. They also have a responsibility to articulate thoughtfully the philosophy and beliefs that these decisions reflect. This thoughtful decision-making yields two important tests of whether the choices made are sustainable. The first test is whether students who have historically been underserved and ill-prepared for adult life will see improved outcomes. The second is whether the technical defensibility of an AA-MAS rests in validity arguments that begin in these decisions and play out in each step of assessment design. This second test is discussed throughout this volume in all the chapters; the first test begins with the choices on

who participates in AA-MAS and the effect of that participation on their access to high quality standards-based instruction and curriculum, and ultimately, their academic achievement.

Key Political and Social Justice Issues

The test of whether students will see improved outcomes must be considered as the initial decisions are made. Public articulation and discussion of these decisions can ensure that historically low expectations and opportunities to learn are not reinforced. In the spirit and the letter of both NCLB and IDEA, implementation of AA-MAS should expressly raise expectations and result in higher achievement for students who participate in the option. In the current standards-based accountability-driven reform model, any option that encourages or rewards less challenging standards for any student who could achieve at grade level (assuming they have access to the curriculum and are instructed effectively) undermines the entire system of school reform. For most students, with and without disabilities, we cannot predict with any accuracy whether they will achieve at grade level when instructed effectively. Thus, the least dangerous assumption requires that all students receive that instruction. Careful monitoring of consequences of the system is essential to ensure the intended positive consequences are occurring and unintended negative consequences are not.

Data-mining of current student performance such as that described in Chapter 10 (Domaleski, this volume) can shape the dialogue among stakeholders. Key questions to consider include: What evidence exists to suggest that students with disabilities who are low-performing differ from minority students or poor students who are low-performing? What evidence exists to support the policy assumption that some of these students cannot achieve at grade level even if their opportunity to learn (OTL) is appropriate? What do direct observations of student instructional and curricular opportunities tell us, and how does that compare and contrast to teacher perceptions? What evidence exists to support the notion that OTL is appropriate for all students in the subgroup? States need to articulate the belief systems that underlie their policy decisions and justify these beliefs based on data. There is a fundamental

tension within the regulatory options. If the consequences of participating in an AA-MAS are positive, how can we deny the same opportunities to other students with similar achievement profiles other than disability status? If the consequences are negative, how can we justify the use of the options for students with disabilities?

Varying State Assumptions, Philosophies and Beliefs, and Their Implications

As states articulate their philosophy and beliefs related to students who may participate in AA-MAS, they are constructing the foundation of their assessment choices. Examples of underlying philosophies and beliefs, which yield very different definitions of eligible students, assessment options, and eventual student achievement are below. These are very different positions; most states participating in public discussions can categorize their philosophy as more or less similar to one of these three options, at least one of which does not seem to match the AA-MAS regulatory requirements. Each of these philosophical positions has infinite variations, many of them controversial, but the dominant profile of each is presented here for general consideration.

Guiding Philosophy 1: Student performance based on past test results and instructional opportunities predict appropriate expectations for future performance. There is a group of students with disabilities who are being instructed in a modified curriculum, one linked to the general curriculum as required by IDEA but different from their enrolled-grade peers. Thus, the state needs to provide a modified assessment to match the curriculum. Given that many of these students will never catch up to their enrolled-grade peers, this will result in very different long-term outcomes. Thus, policy choices focus on modifying the assessments to match a modified curriculum. The target population for inclusion in these modified assessments includes students who may be participating in a modified curriculum.

This philosophy does not appear to match the requirements of the AA-MAS, but may

result in an additional assessment option under AA-AAS flexibility in some states. As such, it is a “nonexample” of a philosophical foundation for AA-MAS.

Implications of Guiding Philosophy 1: The assumption that there is a group of students with disabilities who participate in a modified curriculum different from their typical peers or development of an assessment option that does not result in a standard diploma would raise peer review and advocacy concerns under the requirements of the 2% regulation. Instead, this assumption better fits a very small group of students who may now participate in AA-AAS but for whom the current alternate achievement standards are not an appropriately high expectation. In other words, these are students who currently may be topping out on the AA-AAS or performing at the very lowest levels of the general assessment. There generally are very few students in most states who fit into this group, once opportunity-to-learn barriers are taken into account. An assessment built under this assumption would better match the requirements for a new, more challenging AA-AAS than the requirements of an AA-MAS. The 1% regulation permits states to set more than one alternate achievement standard. Some states do not have a full 1% of students participating in the AA-AAS, while other states find that although they have 1% or higher participation in AA-AAS, they recognize that for some students who participate, it does not reflect a sufficiently high expectation. Given the assumptions of separate track curriculum, assessment, and outcomes assumed in this first philosophy, development of a second AA-AAS with higher performance expectations than on the existing AA-AAS may be warranted.

Guiding Philosophy 2: Student achievement needs to be considered in the context of systematic opportunities to learn (services, supports, specialized instruction, etc.) the general curriculum based on grade-level content and achievement standards. For students who do not as yet have these opportunities, intervention on their opportunities to learn is the first priority in order to accelerate their learning. Given that it will take some time for these students to regain any lost ground from previous limited OTL, short term AA-MAS options are an appropriate interim measure to hold

schools accountable for their learning. These AA-MAS should cover substantially the same content as the general assessment, but may have more content coverage at the lower end of the grade-level content expectations. Still, the state policy is based on an assumption that eventually, with strong interventions and evidence-based practices in student services, supports, and specialized instruction, these students will achieve the same outcomes as typical peers.

Implications of Guiding Philosophy 2: This philosophy seems to match the language of the 2% regulation, although it is still challenging to conceptualize how to use AA-MAS as interim accountability measures with the expectation that students will catch up over time. (See Domaleski, Chapter 10, this volume for examples of how this may affect design of the assessment system.) Those students who are identified through state data-mining efforts as making gains each year but who are still below proficiency over several test administrations may be the target population under this theory. Response-to-Intervention and progress monitoring tools should be in place to get better data to understand their needs and to plan improved services supports, and specialized instruction to ensure that they do, indeed, catch up over time. A state working under this philosophy may consider requiring data from multiple sources (e.g., Rtl, progress monitoring, interim assessments, etc.) to document a decision about a student's participation in AA-MAS each year, and also require documentation of the evidence-based practices that have been implemented for the student based on these data. This evidence of opportunities could be included in requirements under the accountability system in addition to the assessment data themselves. They may also decide to implement an AA-MAS only at selected grades, and instead promote and support through training and other resources district/school efforts like Rtl and progress monitoring to prevent students from falling behind. States and districts may need to design intensive training and coaching efforts to ensure teachers have the skills necessary to effectively teach these students. Alternatively, a state may require that students exit from the AA-MAS in a set number of years, or sunset the AA-MAS

completely, depending on what data reveal about whether the combined practice of increased OTL and selected use of AA-MAS support improved achievement for individuals and for the subgroup or not over the longer term.

Guiding Philosophy 3: Student characteristics (related to disability, ethnicity, poverty, and other demographic categories) have been associated with a history of low expectations and limited opportunities to learn. Until we intervene to change learning opportunities, we cannot use past performance to predict which students could achieve to proficiency. This is true for students with disabilities and for those without disabilities. Designing a new assessment option now jeopardizes an opportunity for reform to ensure all students are taught well. The state policy must ensure students are being taught first, then see who is left achieving at low levels once all are taught well before building any assessment options that may risk perpetuating lowered expectations and outcomes.

Implications of Guiding Philosophy 3: This approach assumes that the state will provide leadership and resources to intervene on historical lack of opportunities to learn for all the students described as the low performers in state data sets discussed earlier in this chapter. Certainly systematic interventions like RtI and progress monitoring, new interim or formative assessment options, and continued efforts in states to implement programs like Positive Behavioral Supports (PBS) contribute to this effort. These efforts are difficult to take to scale, but are laudable and necessary. They are not, however, sufficient to ensure that all students who historically have had limited opportunities to learn and continued low expectations for their performance achieve at higher levels. States that voice this philosophy will look for ways to intervene on attitudes and beliefs of educators and the public who still find it appropriate to expect less of students simply because they have a disability label, or who are poor or of minority status. They will support these educators with increased training and coaching on

evidence-based methods to effectively instruct all students, with and without disabilities, in the challenging standards-based curriculum.

This philosophy results in a choice not to develop an AA-MAS at this time. Yet, the state systematically is addressing a key component of the accountability system, that of leveraging improvements in standards-based instruction and curricular access for all students, especially those students who have historically been underserved, including those with disabilities. As such, this philosophy highlights the choices states have to make on how to most effectively use limited resources to ensure the best possible outcomes. The costs associated with these decisions include opportunity costs of those choices not made.

Another philosophical stance, outside the parameters of the current accountability system: There is another philosophy underlying assessment choices in states that is not included above, since it does not assume that participation in these assessments should lead to systematic improvement in instruction and curriculum for low-performing students and ultimately to improved student achievement, as is assumed in the current accountability model. That set of beliefs is focused around the need to provide what is often described as “relief” to districts and schools from the consequences of system accountability for students with disabilities. If this philosophy dominates state discussions, then much of the guidance in this volume will be overdesigned for state purposes in developing such an assessment. A more straightforward and less expensive method to support this option is to exempt up to 2% of students from the assessment and accountability, but then the state leaders would be responsible for supporting the implications of that decision, both under Federal compliance requirements and from advocacy groups interested in protecting the rights of students with disabilities.

Relationship of AA-MAS to the State Options on Accommodations and Alternate Assessments Based on Grade-level Achievement Standards

Regardless of the underlying philosophy, states will have to define how the AA-MAS contributes to an overall system of assessment for accountability purposes. That is, there

should be a coherent educational logic or relationship between the AA-MAS and other alternate assessments (based on AAS or GLAS) as well as with the general assessment. Universal design, accommodations policies, decision-making guidelines, training, and monitoring should support the validity of the assessment system so that all students are included in ways that support use of assessment results in a standards-based accountability system.

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 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. This can result in students who have learned the skills and knowledge assessed on the test scoring below proficiency. Some students with disabilities may have barriers to showing what they know on the state general assessment, even with accommodations. These barriers may be related to several different disability characteristics (e.g., processing, sensory, physical, or emotional barriers). In order to ensure the validity of inferences of the assessment system for these students, states may need to consider an alternate assessment based on grade-level achievement standards (AA-GLAS).

Perie (2009) posed key questions that states will have to answer related to the relationship among the state assessment options: “Do we expect to see smooth transitions from one assessment to the next? How do the performance expectations relate? Is Proficient on the AA-MAS similar in nature to Proficient on the general assessment? Is it closer to Basic? Or is it somewhere in between? Is there an expectation that the AA-MAS may provide a stepping stone for students to reach Proficient on the general assessment? Or, is the expectation that students taking the AA-MAS are a unique population that will always need the modifications provided? Is a student who scores Advanced on the AA-MAS prepared to take the general assessment or an AA-GLAS or are they simply exceeding the criterion on their own assessment?”

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. However, any change in academic achievement standards for a group of students 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. The foundation for decisions about assessment options must rest and be defended based on a publicly articulated set of beliefs about teaching, learning, and eventual outcomes for all students.

Questions for States as They Choose, Build, and Defend the Validity of Their Approach

States must build an argument to defend the validity of their approach to AA-MAS that begins with articulated definitions of who the students are who will benefit from the AA-MAS, and why. The types of questions states must answer in their validity argument need to be framed while they are making the decision to develop an AA-MAS, during development of an AA-MAS, and in their continuous improvement and consequential validity studies as they implement their AA-MAS. Related to identification of the students, these include questions about the students who may participate; the instruction, support, and resources provided to these students at the local level and the quality of IEP decisions made about their participation; and about how the AA-MAS is appropriate for its articulated purposes and uses—including those of improving achievement for students who participate.

Later chapters in this volume (especially Marion, Chapter 9; Domaleski, Chapter 10) focus on the validity argument and on practical implications of state choices. It is important to include questions about how a state can validate that the appropriate students are identified and participate in an AA-MAS. This chapter is written under the assumption that implementation of AA-MAS should expressly raise expectations and result in higher achievement for students who participate in the option. That is a testable assumption, and should be considered as a state

makes choices and implements an AA-MAS. Questions about the appropriateness of AA-MAS for improving student outcomes include (adapted from Marion, 2007):

- How does this assessment provide a more accurate measure of the knowledge and skills of the participants compared with the general assessment?
- How does development of an AA-MAS yield better inferences about the students than other assessment approaches, such as improved general assessment design, appropriate accommodations, or development of AA-GLAS?
- What are the potential costs and benefits of competing uses of resources, including targeted staff development on instructional and curricular interventions for teachers of struggling learners instead of assessment development and implementation?
- How will the inclusion of the AA-MAS as part of the state's assessment system lead to better instructional and curricular opportunities for these participating students?
- Other questions identified by policymakers and stakeholders.

Marion (2007) has identified potential sources of data for many of these questions in the form of a workbook for AA-MAS development and documentation. Design of validity studies should begin while planning for any assessment option, and tools such as this volume and earlier work done by Marion and others can guide those designs.

In conclusion, in order to make decisions about whether and how to design an AA-MAS, states need to articulate a guiding philosophy that defines which students will benefit from participating in the assessment, and how they will benefit. States then can design the assessment based on the specific learning characteristics and opportunities to learn of students who may participate. Over the next several years, states will need to work in partnership with researchers and experts in curriculum, instruction, assessment, and disability issues to better understand and identify the appropriate students for participation in the AA-MAS. Ultimately, 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 standard diploma and to succeed in adult life.

References

- Abedi, J. (2007). English language learners with disabilities. In Cahlan-Laitusis, C. & Cook, L. *Accommodating student with disabilities on state assessments: What works?* (Ed.) Arlington, Council for Exceptional Children.
- Abedi, J. (2006). Psychometric Issues in the ELL Assessment and Special Education Eligibility. *Teacher's College Record*, Vol. 108, No. 11, 2282-2303.
- Artiles, A. J., & Ortiz, A. (Eds.). (2002). *English language learners with special needs: Identification, placement, and instruction*. Washington, DC: Center for Applied Linguistics.
- Barr, S., Telfer, D., & DiMuzio, M. (2009). *The Ohio Improvement Process (OIP) as a strategy for creating a viable SSOS for all Ohio districts and schools*. Ohio Department of Education presentation to USED Intradepartmental Work Group, February 4, 2009.
- Cortiella, C., & Burnette, J. (2007). *Challenging change: How schools and districts are improving the performance of special education students*. New York: National Council for Learning Disabilities.
- Deno, S., Fuchs, L., Marston, D., & Shin, J. (2001). Using curriculum-based measurement to establish growth standards for students with disabilities. *School Psychology Review*, 30(4), 466-472.
- Donahue Institute. (2004, Oct). *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. Retrieved March, 2006, from http://www.donahue.umassp.edu/docs/?item_id=12699
- Donnellan, A. (1984). The criterion of the least dangerous assumption. *Behavior Disorders*, 9, 141-150.
- Fincher, M. (2007). "Investigating the academic achievement of persistently low performing students" in the session on *Assessing (and Teaching) Students at Risk for Failure: A Partnership for Success* at the Council of Chief State School Officers Large Scale Assessment Conference, Nashville TN, June 17-20, 2007. Retrieved August, 2007, from <http://www.ccsso.org/content/PDFs/12%2DMelissa%20Fincher%20Paul%20Ban%20Paam%20Rogers%20Rachel%20Quenemoen.pdf>.
- Fincher, M. (2009). Personal communication with the author on April 16, 2009.
- Fletcher, J. (2008). *Identifying learning disabilities in the context of response to intervention: A hybrid model*. Retrieved March 2009, from http://www.rtinetwork.org/index2.php?option=com_content&task=view&id=331&pop=1&page=0&Itemid=45
- Fuchs, D., & Fuchs, L. S. (2006). Introduction to Response to Intervention: What, why, and how valid is it? *Reading Research Quarterly*, January- March. 93-99.

- Fuchs, L., Seethaler P. M., Fuchs, D., & Hamlett, C. L. (2008). Using curriculum-based measurement to identify the 2% population. *Journal of Disability Policy Studies*. 19(3), 151-161.
- Gersten, R., Compton, D., Connor, C. M., Dimino, J., Santoro, L., Linan-Thompson, S., & Tilly, W. D. (2008). *Assisting students struggling with reading: Response to Intervention and multi-tier intervention for reading in the primary grades. A practice guide*. (NCEE 2009-4045). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. Retrieved April, 2009, from http://ies.ed.gov/ncee/wwc/publications/practice_guides/ .
- HB 05-1246 Study Committee (2005, December 31). *Assessing "students in the gap" in Colorado*. Retrieved May, 2006, from <http://education.umn.edu/nceo/Teleconferences/tele11/ColoradoStudy.pdf>.
- IDEA Part B Child Count. (2007). *Students ages 6 through 21 served under IDEA, Part B, by disability category and state: Fall 2007* (Tables 1-3). Retrieved September 2008, from www.IDEAdata.org
- Kaloi, L. (2007). *A Misunderstood Policy with the Potential to Harm Millions of Kids*. Retrieved March 8, 2009, from <http://www.nclb.org/content/view/1279/311/>
- Karger, J., & Boundy, K. (2008). Including Students with Dyslexia in the State Accountability System: The Basic Legal Framework, *Perspectives on Language and Literacy International Dyslexia Association*, 34(4), Fall, 2008.
- Marion, S. (2007). *A technical design and documentation workbook for assessments based on modified achievement standards*. Minneapolis: National Center on Educational Outcomes. Retrieved April, 2009, from <http://cehd.umn.edu/nceo/Teleconferences/AAMASteleconferences/AAMASworkbook.pdf>
- Marion, S., Gong, B., & Simpson, M. A. (2006, Feb. 6). Mining achievement data to guide policies and practices on assessment options. *Teleconference on Making Good Decisions on NCLB Flexibility Options*. Minneapolis: National Center on Educational Outcomes. Retrieved April, 2009, from <http://education.umn.edu/nceo/Teleconferences/tele11/default.html>
- Minnema, J., Thurlow, M., Anderson, M., & Stone, K. (2005). *English language learners with disabilities and large-scale assessments: What the literature can tell us* (ELLs with Disabilities Report 6). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.
- National Association of School Psychologists (NASP). (2002). *Position statement: Rights without labels*. Original statement adopted by NASP Delegate Assembly in 1986. Revision adopted by NASP Delegate Assembly, July 14, 2002. Retrieved March 10, 2009, from http://www.nasponline.org/about_nasp/pospaper_rwl.aspx
- National Center on Educational Outcomes. (2008). *Previous studies that examine who the students are who may qualify to participate in an AA-MAS*. White paper for Expert Panel

Meeting Multi-State GSEG Consortium toward a Defensible AA-MAS, Minneapolis, MN: March 10, 2008.

- New England Compact. (2007). *Reaching students in the gaps; A study of assessment gaps, students, and alternatives*. Newton MA: CAST, The Education Alliance, EDC, INTASC, Measured Progress.
- Parker, C. E., & Saxon, S. (2007). *Teacher views of students in the gaps*. New England Compact Enhanced Assessment Grant. Retrieved March 9, 2009, from http://www.necomact.org/Teacher_views_of_students.pdf
- Perie, M. (2009). *Understanding the AA-MAS: How does it fit into a state assessment and accountability system?* Presentation to CCSSO SCASS meeting, February 4, 2009. Available at www.nciea.org.
- Quenemoen, R., Thurlow, M., Moen, R., Thompson, S., & Morse, A. B. (2003). *Progress monitoring in an inclusive standards-based assessment and accountability system* (Synthesis Report 53). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.
- Thurlow, M. (2007). The challenge of special populations to accountability for all. In D. Clark (Ed.), *No child left behind: A five year review* (Congressional Program, Vol. 22, No. 1) (pp. 39-44). Washington, DC: The Aspen Institute.
- Thurlow, M. L., Moen, R. E. Liu, K. K., Scullin, S., Hausmann, K. E., & Shyyan, V. (2009). *Disabilities and reading: Understanding the effects of disabilities and their relationship to reading instruction and assessment*. University of Minnesota: Partnership for Accessible Reading Assessment.
- United States Department of Education. (2005). *Raising achievement: Alternate assessments for students with disabilities*. Retrieved March 8, 2009, from <http://www.ed.gov/print/policy/elsec/guid/raising/alt-assess-long.html>.
- USED. (2007). Nonregulatory Guidance Modified Academic Achievement Standards.
- Weckstein, P. (1999). School reform and enforceable rights to quality education. In J. Heubert (Ed.), *Law and school reform: Six strategies for promoting educational equity*. (pp. 306-389). New Haven: Yale University Press.
- Wendorf, J. (2005). Open letter to Secretary of Education Margaret Spellings, March 27, 2005. Retrieved March 8, 2009, from <http://www.nclld.org/content/view/289/>

CHAPTER 3

DEVELOPING STANDARDS-BASED IEPs THAT PROMOTE EFFECTIVE INSTRUCTION

Meagan Karvonen

States have the option to create Alternate Assessments based on Modified Achievement Standards (AA-MAS) for students with disabilities who perform persistently and significantly below grade level. This population is heterogeneous within states, and may also be defined differently across states (Perie, 2009; Quenemoen, chapter 2, this volume). Regardless of their characteristics and needs, this population of students requires extensive supports and effective instruction in order to meet high expectations and transition back into eligibility for grade-level assessment.

The importance of effective instruction for this target population of students was recognized in the final regulations on the AA-MAS [U.S. Department of Education [USED], 2007a, 34 CFR 200.1(f)(iii)]. There are three alternate assessment options under NCLB: (a) those based on grade-level achievement standards (AA-GLAS), (b) those based on alternate achievement standards (AA-AAS), and (c) AA-MAS. While access to instruction based on grade-level academic content standards is recognized for all students under IDEA, the regulations for AA-MAS represent the first explicit assumption under NCLB that instruction has afforded the student maximum opportunity to learn what is assessed. Opportunity to learn requires a curriculum that is well-aligned to state standards and assessment so students can show what they know and can do. For the population of AA-MAS-eligible students, a well-aligned curriculum is a critical foundation. Since these students have not been successful with previous classroom instruction, they will also need for this curriculum to be designed and delivered using the best, evidence-based instructional practices available.

The nonregulatory guidance on AA-MAS (USED, 2007b) reinforces the LEA's responsibility to design a highly effective curriculum and instruction and requires a standards-based individualized education program, or IEP:

The primary reason for requiring IEP goals based on grade-level academic content standards is to ensure that students who participate in an assessment based on modified academic achievement standards receive instruction in grade-level content so that they can make progress towards meeting grade-level proficiency (p. 28).

In other words, the IEP is a way of driving the student's academic curriculum toward the goal of transitioning back into grade-level assessments. The IEP also provides the evidence for how the instructional program will incorporate supports to address the student's characteristics stemming from the disability.

While states may differ in how they plan to approach AA-MAS or educate their students who may be eligible for this assessment, there are federal requirements related to the contents of IEPs that are universal across states. The IEP cannot completely capture the entire academic curriculum for a student eligible to take AA-MAS. However, planning and writing the IEP can help educators think systematically about how to design high-quality instruction for this population. The purpose of this chapter is to describe how a standards-based IEP can support a well-designed educational program that ensures access to grade-level curriculum using effective instructional practices. After a description of the IEP as a document, some principles are offered for effective instruction for this population of students. Next is a description of how the IEP can promote good instruction, followed by suggestions for how states can provide guidance to IEP teams that are responsible for creating standards-based IEPs for this population of students. Requirements for state-level monitoring of IEP systems are also reviewed. The chapter ends with a section on validity evidence related to curriculum and instruction, and general conclusions. While most of this chapter is written with the intent to promote ideal and potential best practices, there are still some areas in which current, realistic

practices have been challenged to reach the optimal, “best practice.” The conclusion section of this chapter acknowledges some of these challenges. Along with the other two chapters in this section, which discuss possible characteristics of the target population including how they learn, this chapter sets the stage for the remaining chapters on designing and implementing an AA-MAS.

An Overview of the IEP

The Individualized Education Program, or IEP, is written at least annually for each student with a documented disability. Although there are several required components, the general idea is to consider the student’s present levels of performance and documented needs and strengths in order to create a comprehensive plan for the student’s priorities that year. Those priorities do include academics, but also reflect other supports that are essential to provide the student with meaningful instruction given the features of his or her disability that make access more challenging. By specifying these supports, students can more fully participate and be successful in the pursuit of their educational goals. The IEP is written and its contents agreed upon by a team that includes one or more special education teachers, a general education teacher, other educational professionals (e.g., speech/language therapists, counselors), the student’s parents or guardians, and in some cases the student as well.

IEPs have been part of special education services since 1975. However, they have not always played a central role in describing the academic curriculum. Karger (2004) reviewed the historical literature on IEPs and noted a variety of problems including a disconnect between IEPs and the curriculum, poor congruence across sections within the IEP (e.g., between documented needs and annual goals), and special educator perceptions that the IEP was irrelevant to instruction. In the early years, IEPs documented special education services that ran parallel to general education (Ahearn, 2006). In the 1990s, IEP teams began determining that students would spend more time in general education settings, often for nonacademic activities

(e.g., music classes, lunchtime in the cafeteria). This trend toward inclusion in general education settings gave students with disabilities more access to the school building, but did not give full access to the academic curriculum that was taught to students without disabilities. What special educators now refer to as “access to the general curriculum” was mandated in the purpose of special education as written in the Individuals with Disabilities Education Improvement Act (IDEA) of 1997:

To address the unique needs of the child that result from the child's disability; and to ensure access of the child to the general curriculum, so that he or she can meet the educational standards within the jurisdiction of the public agency that apply to all children (34 C.F.R. § 300.26(b)(3))

With the enactment of IDEA 1997, IEPs were required to address the student’s present levels of performance, include annual goals and short-term objectives to help the student progress in the general curriculum, and document program modifications and supports the student would need in order to progress in the general curriculum. Thus, IEP teams were first required to address general curriculum access just over a decade ago.

The next reauthorization of IDEA came after NCLB, in 2004. Under IDEA 2004, IEPs must now contain the following elements:

- A statement of the child's present levels of academic achievement and functional performance
- A statement of measurable annual goals, including academic and functional goals designed to (a) meet the child's needs that result from the child's disability to enable the child to be involved in and make progress in the general education curriculum; and (b) meet each of the child's other educational needs that result from the child's disability;

- A statement of the special education and related services and supplementary aids and services, based on peer-reviewed research to the extent practicable, to be provided to the child, or on behalf of the child;
- A statement of any individual appropriate accommodations that are necessary to measure the academic achievement and functional performance of the child on State and districtwide assessments consistent with section 612(a)(16) of the Act; and if the IEP Team determines that the child must take an alternate assessment instead of a particular regular State or districtwide assessment of student achievement, a statement of why the child cannot participate in the general assessment and why the particular alternate assessment selected is appropriate for the child....(34 CFR §§ 300.320 - 300.324)

In designing the IEP, teams must consider the student's strengths, parents' concerns, results of the most recent evaluation, and "academic, developmental, and functional needs" of the child [34 CFR §300.324(a)(1)(i-iv)]. In addition, teams must decide whether certain special factors must be considered in planning the educational program for the student:

1. In the case of a child whose behavior impedes the child's learning or that of others, consider the use of positive behavioral interventions and supports, and other strategies, to address that behavior;
2. In the case of a child with limited English proficiency, consider the language needs of the child as those needs relate to the child's IEP;
3. In the case of a child who is blind or visually impaired, provide for instruction in Braille and the use of Braille unless the IEP Team determines, after an evaluation of the child's reading and writing skills, needs, and appropriate reading and writing media (including an evaluation of the child's future needs for instruction in Braille or the use of Braille), that instruction in Braille or the use of Braille is not appropriate for the child;

4. Consider the communication needs of the child, and in the case of a child who is deaf or hard of hearing, consider the child's language and communication needs, opportunities for direct communications with peers and professional personnel in the child's language and communication mode, academic level, and full range of needs, including opportunities for direct instruction in the child's language and communication mode; and
5. Consider whether the child needs assistive technology devices and services. [34 CFR §300.324(a)(2)]

There is language throughout IDEA 2004 [see §300.321(a)(1-7), §300.321(b)(1-3)] that clearly emphasizes the academic curriculum and the link between assessment and instruction. The importance of academics is even recognized in guidance that general educators should be members of the IEP team. Relative to academic instruction, the IEP is now to reflect how the student will access the general curriculum and what the academic priorities are. New guidance on AA-MAS also calls for monitoring progress toward academic goals.

What Does it Mean for an IEP to be “Standards-Based”?

In interviews with representatives from 18 states, Ahearn (2006) found some confusion over the term “standards-based” as it applied to IEPs. Part of that confusion came from the different types of standards (i.e., content and achievement). IDEA 2004 requires “(i) A statement of measurable annual goals, including academic and functional goals designed to—(A) Meet the child's needs that result from the child's disability to enable the child to be involved in and make progress in the general education curriculum; ...” [34 CFR §300.320(a)(2)(i)]. IEPs that meet the IDEA requirements include a broad range of information to explain how the student will access the general curriculum.

The two key elements specific to students who take AA-MAS are that there must be annual goals based on grade-level academic content standards and that there must be

mechanisms in place to measure progress toward achieving those goals. The specific language from the final AA-MAS regulations is as follows:

- a. “The student’s IEP must include goals that are **based on the academic content standards for the grade in which the student is enrolled**” ...That is, while students may have their performance evaluated against modified achievement standards, they must be taught academic content based on grade-level content standards.
- b. ...“and be designed to **monitor the student’s progress in achieving the standards-based goals**”. [U.S. Department of Education, 2007a, 200.1(e)(2)(iii), (f)(2)(i), emphasis added]

Although access to a curriculum based on state content standards has been guaranteed since 1997 and was reinforced in IDEA 2004, the requirement that state content standards be reflected in IEP goals is new with AA-MAS (Thurlow, 2008). States have taken various approaches to interpreting the requirement for “standards-based” IEPs (Ahearn, 2006). Students’ present levels of academic performance are evaluated based on their mastery of state content standards and areas in which they have not yet mastered those standards. Some states then require the IEP team to write goals that emphasize the skills the student will need in order to make progress in those content standards that have not yet been mastered. Other states require IEP teams to consider academic content standards broadly when evaluating present levels of performance and setting goals, but do not require teams to base those decisions on specific grade-level standards (Ahearn).

With the standards-based approach to IEPs, it is clear these documents now play a much different role in educational planning in 2009 than they did in 1975. As states think about how to address the regulatory requirements for AA-MAS, it is important to keep in mind that the IEP cannot be a map of the entire academic curriculum for a student. It cannot contain documentation of all instructional strategies used with a student in a given year. It can, however, drive a purposeful planning process and ensure that good decisions about educational goals,

grounded in a clear understanding of student needs, are established. Thus, before addressing IEPs directly, this chapter first considers implications for the planning, delivery, evaluation, and adjustment of instruction for this population of students.

Effective Curriculum and Instruction

One common criterion for determining eligibility for the AA-MAS is that the student will not attain proficiency on grade level assessment *in the current year*, despite having received appropriate instruction all year. In reality, some students who are eligible for AA-MAS have probably not been proficient on grade-level assessments for *multiple years*—they are what some states call “persistently low-performing.” Thus, the target population may vary in their patterns of past performance, depending on how states define eligibility for the AA-MAS. Eligibility decisions based on past performance will be tied to states’ guiding philosophies and theories of action about the AA-MAS (see Quenemoen, Chapter 2, and Marion, Chapter 9, this volume). Regardless of past performance patterns, the assumption is that these students are unlikely to achieve proficiency on the general assessment *despite having had instruction in grade-level academic content*. Based on surveys and focus groups from teachers in several states, Perie (2009) synthesized findings from several focus groups on students who may be eligible for AA-MAS and provided a potential list of their characteristics:

- Require intensive, specially designed instruction and individualized supports
- Instruction repeated many times in many ways to make progress
- Passive learners, non-risk takers
- Learning and meta-cognitive deficits (e.g., poor generalization, difficulty transitioning between topics, applying learned strategies) [slides 19–21]

While this list is not exhaustive and may not represent the characteristics of AA-MAS-eligible students in all states, these types of characteristics reinforce the notion that highly effective instruction is the lynchpin for this population. Without meaningful access to the general

curriculum, low-performing students will have little hope of working toward grade level expectations. What is essential in designing an instructional program that will help students meet these high expectations for growth? The Access Center (2006) offered the following as characteristics of educational programs that provide meaningful access to the general curriculum:

- The general education curriculum includes appropriate, standards-based instructional and learning goals for individual students with disabilities, as well as reflects an appropriate scope and sequence.
- Materials and media being used are appropriate, research-based, and documented as being effective in helping students with disabilities learn general education content and skills.
- Appropriate, research-based instructional methods and practices that have a track record for helping students with disabilities learn general education content and skills are being used.
- Research-based supports and accommodations that have a track record of helping students with disabilities learn general education content and skills are being used.
- Appropriate tools and procedures for assessing and documenting whether students with disabilities are meeting high standards and achieving their instructional goals are being used (pp. 4, 6).

This section of the chapter describes several characteristics of highly effective curriculum and instruction for students who present with the challenges described above.

Suggested resources are provided in Appendix B of this report.

1. Students are given access to grade-level content

As noted earlier, ensuring access to grade level content standards is a fundamental expectation in the Federal Regulations on AA-MAS [USED, 2007b, §200.1(f)(2)(iii)]. State

content standards are often organized around content knowledge and processes reflected in standards or frameworks from national groups (e.g., National Council of Teachers of Mathematics, 2000; National Reading Panel, 2000). States differ in the grain size (i.e., level of specificity) and sequences of content area skills and knowledge when articulating their content standards. (See Pugalee and Rickelman, Chapter 5, this volume, for additional discussion on state content standards.)

By law, students eligible for AA-MAS are to be taught a curriculum that is based on chronologically-appropriate, grade-level standards. Unlike AA-AAS, where students receive instruction in content that “links” to grade-level content standards, students who take AA-MAS are expected to be working “in” grade-level content standards. That does not mean they are taught the state standards directly; instead, they are provided with a curriculum that is aligned to state standards. Pugalee and Rickelman (Chapter 5, this volume) define curriculum as, “a set of planned instructional activities that are designed to allow students to document achievement of their knowledge and skills” (p. 155).

One of the challenges in targeting academic goals for this target population of students is that they may lack foundational skills from earlier grade levels, upon which the current grade-level content standards are based. Teachers will walk a fine line between helping the student master skills from earlier grades while also being taught a curriculum based on current grade-level standards. Ahearn (2006) offered the following for how students might work toward IEP goals based on content standards from an earlier grade:

Such students are expected to make more than one year of progress through standards-based instruction because the needed skills are targeted by the teacher. Teachers scaffold instruction (i.e., provide supports as necessary) and prerequisite skills are used to work toward the grade-level standards. For example, a student who cannot read 6th grade materials may work toward a grade-level standard that calls for analyzing written materials. The cognitive processes associated with that higher level

reading skill can still be taught while the student accesses the grade-level materials in a different way. (p. 8).

The caution here is that, according to the AA-MAS regulations, IEP goals must be based on the current (chronologically-appropriate) grade-level content standards. In order to plan an effective curriculum based on remediation of missing skills and progress in current skills, teachers will have to be knowledgeable about how topics within a strand (e.g., algebra, writing) relate to each other within a grade level and how they build across grades. Teachers also will need to make choices about how best to sample content from within a standard. Clear alignment between assessment and instruction is essential if student assessment scores are to allow for valid inferences about proficiency against a set of standards (American Educational Research Association [AERA], American Psychological Association [APA], & National Council on Measurement in Education [NCME], 1999). General education teachers with expertise in content areas may play a critical role in helping special educators select content and design instruction for this population. These content experts will be fluent in their state's grade-level content standards and national standards for the discipline; know how instruction is designed to promote learning in that skill; and understand how the components relate to and build upon each other (see Pugalee & Rickelman, Chapter 5, this volume).

Curriculum materials can make it easier (or more difficult) for teachers to deliver grade-level curriculum. Adaptations of grade-level materials may be increasingly necessary the farther away from grade-level proficiency a student is working, and the more unique or complex the student's learning needs. The accessibility issue related to curriculum materials becomes increasingly important as students enter upper grade levels. In middle schools, textbooks are more like reference books and contain vocabulary that is above grade level; factual knowledge is emphasized over procedural knowledge (Hill & Erwin, 1984; Jitendra, Nolet, Xin, Gomez, Renouf, Iskold, et al., 2001). There may be a tendency for special educators working with AA-

MAS-eligible students to significantly adapt materials, potentially jeopardizing access to grade-level content if the adaptations stretch too far.

2. Instruction consists of proven practices that allow teachers to set a trajectory toward performance that can be evaluated against grade-level achievement standards

Meaningful progress toward grade-level achievement requires accurate assessment of the student's current level of performance — using multiple, high-quality assessments that are appropriate for the student and provide valid inferences about strengths and areas for growth (34 CFR § 200.1(e)(2)(ii)(B)). When states determine eligibility for AA-MAS based on a pattern of low performance across years (vs. a single year), baseline performance may also be investigated retrospectively to help teachers determine how firmly established a pattern of limited or no progress has been. Armed with this information, teachers would then also draw on their knowledge about how students learn in order to plan instruction. (See Pellegrino, Chapter 4, this volume, for an extensive description of student cognition).

Once present levels of performance are known, teachers will then adopt research-based strategies — those that work with the type of student whose IEP they are designing. The What Works Clearinghouse (<http://ies.ed.gov/ncee/wwc/>) is one source of information about which interventions have evidence of effectiveness. However, one pitfall is that the strict evidence standards set by the WWC means that when the evidence base is sparse or based on less rigorous designs (e.g., those without a comparison group), emerging evidence is not synthesized. For example, in a March 2009 search of the middle school math interventions listed by WWC, only 12 of 50 interventions had reports available. (Two more reports were pending.) The rest were not synthesized because of a lack of studies that met the criteria. Even when the WWC offers helpful syntheses, teachers will need to consider whether the interventions “work” for students who are eligible for AA-MAS based on their state's criteria. When WWC evidence is lacking, the field may look to other sources of evidence for state-of-the-

art practice (see the Spring 2009 issue of *Exceptional Children*, for example, or literature syntheses and meta-analyses published in journals such as *Review of Educational Research*).

Selecting research-based strategies and applying them indiscriminately will not automatically boost student learning. Teachers also use their knowledge of content and how skills link and build upon one another toward a goal for performance that year. Concepts such as learning progressions and learning maps will be discussed in detail later in this volume (Chapters 4, 5, 7, and 9), but it is worth mentioning them briefly here. Where research exists to support cognitive models of learning, teachers may use those models to plan for a sequence of instruction. Where such evidence does not yet exist, familiarity with typical content sequences may guide planning. Since all students do not learn in the same sequence, awareness of when skills really are (or are not) prerequisite or foundational concepts will also be important. Regardless of whether the teacher thinks in terms of progressions, maps, or sequences, identifying baseline knowledge and skills at a fine-grained level (e.g., a student's specific content knowledge within strands and process skills within a subject area) rather than coarse-grained level (e.g., the student scored below basic in English language arts the previous year), the foundation will be set for a meaningful path for teaching and learning.

If teachers have accurately identified present levels of performance and effective instructional strategies for that skill and that student, and have a solid understanding of how to build skills in that area, what should be the target for that student? One option would be to consider the state's modified achievement level descriptors and evaluate how much progress the student would need to make in order to move from his or her current AA-MAS achievement level to the next highest one (see Perie, Chapter 7, this volume, for a discussion of modified achievement standards). Over a period of years, long-term planning for the student could include annual goals for certain skills with a multi-year goal of transitioning back to eligibility for assessment options based on grade-level achievement standards. IEP teams will make decisions about how to set interim achievement targets at lower levels while also maintaining

the highest possible performance expectations. General educators with deep content expertise may be helpful in setting reasonable interim targets. The concepts of learning progressions or learning maps may also guide decisions about interim targets. IEP teams may wish to guard against setting lower targets for several consecutive years, as there may be a point at which the student is unintentionally tracked into a level of performance that precludes his or her later participation in assessments based on grade-level achievement standards.

Assuming a growth trajectory has been set and the student makes progress during the academic year, there is a risk that the student will lose momentum—and perhaps even lose ground that was gained—during long breaks (i.e., summer vacation on a 9-month calendar or intersession on a year-round calendar). IEP teams have the option to include Extended School Year (ESY) services as part of the student’s plan in order to prevent loss of the skills and knowledge that the student built during the academic year.

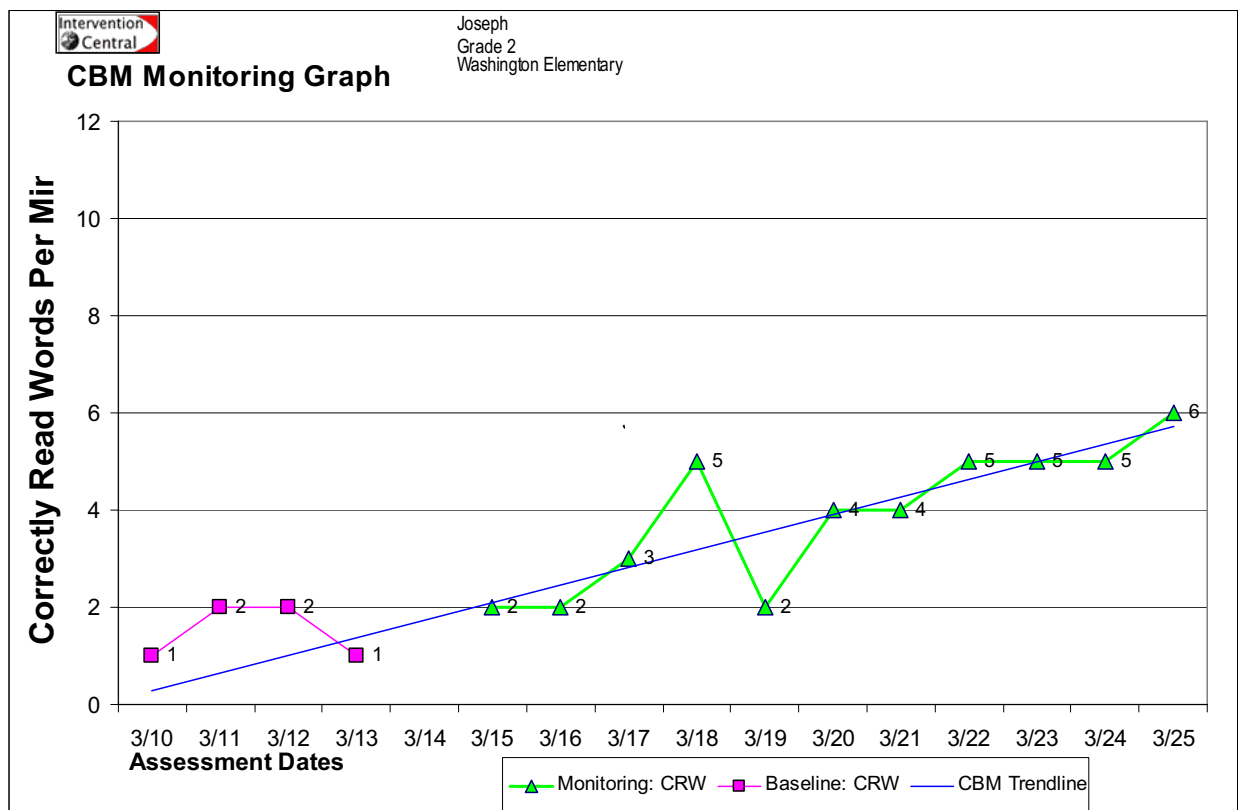
3. Instruction is flexible and responsive to student progress

Even with a well-designed, long-term plan for instruction, teachers cannot just assume the student will progress according to that plan. They will need to monitor progress closely and know when to make decisions to adjust instruction. A vague plan to monitor progress using “teacher observation” will not be sufficient. Teachers need to know how to design or identify, and use, effective formative assessment methods.

Established techniques such as curriculum-based measurement (CBM) and progress monitoring can help teachers track student progress. CBM offers a way to periodically assess student progress toward long-term goals using assessments that are technically sound but easy to use in everyday instruction (Deno, 2003; Stecker, Fuchs, & Fuchs, 2005). Stecker (n.d.) provides a concise overview of how to use CBM for progress monitoring, including how to use CBM data to set IEP goals and short-term objectives. One option for monitoring progress is for teachers to chart performance on a graph with a superimposed trend line (see an example in

Figure 3-1). Establishing the trend line between present level of performance and target level of performance helps the teacher monitor interim progress. There are numerous resources available on the internet for teachers to customize and create graphs for tracking student progress using Excel or Web-based forms (see Appendix B). A synthesis report published by the National Center on Educational Outcomes (NCEO; Quenemoen, Thurlow, Moen, Thompson, & Morse, 2004) places short-term CBM in the broader context of progress monitoring linked to summative assessments and improvement across years.

Figure 3-1. Example of Graph for Tracking Student Progress against a Trajectory



As the literature base on CBM and progress monitoring grows, new studies emerge that may help teachers refine their practice to maximize benefit while minimizing effort. For example, Jenkins, Graff, & Miglioretti (2009) determined that reading growth (defined as words read correctly) could be determined with less frequent measurement but that it was important to

obtain sufficient baseline data and an appropriate number of measurements at each data collection period.

Teachers may effectively set a goal, collect data to measure progress using technically sound measures, and recognize when students are not progressing as they had hoped. The other critical skill is how to determine *why* that progress has not been made. If present levels of performance and student needs were correctly identified in the planning stages, the teacher may be reasonably satisfied that there were no unidentified deficits in prerequisite concepts. However, the teacher may find weaknesses in student knowledge of certain components of the target skill, or may be able to uncover the source of the problem through error analysis and identification of the student's faulty thinking. Although the adjustment of instruction is dealt with extensively in the progress monitoring and Response to Intervention (RtI) literature, Nitko (2004) offers several general recommendations for how to conduct diagnostic assessments of students' problems in learning targeted content. The ability to monitor and adjust instruction requires creativity and persistence by teachers who design and deliver the instructional program.

4. The instructional program minimizes barriers and provides the full range of supports necessary to promote growth

To provide effective instruction, teachers need to determine what aspect of student performance represents the student's actual mastery level and skill deficits, and what aspects may be related to the disability or other barriers to learning. IEP teams are responsible for determining what types of accommodations are appropriate in instruction and assessment, based on how the student's disability impedes his or her ability to learn in the general curriculum. (See the section titled "State Guidance to IEP Teams" later in this chapter for a full discussion of accommodations.) If the student has characteristics considered under IDEA to be special factors that interfere with learning (behavior, limited English proficiency, visual impairment, hearing impairment, or the need for assistive technologies), those areas must also

be addressed. As the options for assistive technologies expand, teachers will need good resources for identifying appropriate devices that minimize barriers stemming from specific student needs (see Appendix B). Teachers may also want to consider how environmental variables, such as instructional setting and grouping, may influence students' access to the general curriculum (Soukup, Wehmeyer, Bashinski, & Bovaird, 2007).

Given the diversity of students who may be eligible for AA-MAS, not all students in a class will need the same learning supports. It is possible that AA-MAS-eligible students can receive learning supports as a matter of standard practice for *all* students. Differentiated instruction allows teachers to identify how each AA-MAS-eligible student will access grade-level content; respond to student progress and adjust instruction accordingly; and incorporate learning supports needed due to student disabilities. Differentiation is more than just grouping—it involves variation in the curriculum (i.e., different materials) and how it is taught (i.e., individual work, small group, whole group; Gibson & Hasbrouck, 2008). However, differentiation for the target population *cannot* involve teaching to lower grade-level standards.

There are two more promising options for minimizing barriers and providing supports that are worth mentioning here. Both are applicable to broader ranges of students and may be useful beyond the target population for AA-MAS. Universal design for learning (UDL) offers a framework for reducing barriers by allowing for multiple means of representation, expression, and engagement (CAST, 2008). UDL is offered as a way to build flexibility and differentiation into instruction from the outset, thereby avoiding the need to adapt curriculum after the fact. Response to Intervention (RtI) focuses on prevention of long-term failure through early assessment and intervention, with frequent progress monitoring and adjustment (National Center on Response to Intervention, n.d.). Although described in IDEA 2004 as a means of identifying and serving students with specific learning disabilities, RtI can be applied to an entire student body for multi-tiered interventions rather than a single sub-population with identified

disabilities. Rtl is discussed in more detail in Chapter 2 of this volume (Quenemoen) and in resources listed in Appendix B.

This section provides only an overview of principles that may guide effective instruction. States and LEAs will determine how best instructional practices translate locally, given state policies and local norms. However, the practices described here offer alternatives to help students who need to demonstrate substantial progress if they hope to reach grade-level proficiency. In many ways, the principles described here are suggestions for operationalizing the IDEA requirement for specifically designed instruction [20 U.S.C. § 1401(29)].

Using the IEP to Promote Quality Instruction

The federal guidance that requires the IEP to contain goals based on grade-level content standards implies that what gets documented is what gets taught. However, the IEP cannot reflect a student's entire academic curriculum in detail. How does the IEP meet the letter of the law while also providing a blueprint for the student's broader educational needs? This section offers suggestions for how to use the IEP to guide high-quality academic instruction, by identifying needs, writing goals, and monitoring progress. It concludes with a discussion on how IEP teams can discuss and select optimal supports based on individual needs.

Reviewing Present Levels of Performance and Identifying Need

When planning an IEP, teams carefully consider present levels of performance and prioritize needs in order to support growth toward grade-level achievement. Present levels of performance may be described using recent assessment data (collected using sound measures) that are specific about student performance and areas for growth. For example, an IEP team will be better able to target goals in areas of need with detailed description (e.g., "Joey can read fifth grade texts with 70% accuracy and 85 WPM but has difficulty inferring meaning of unfamiliar words using context clues.") than generalities (e.g., "according to the Woodcock-Johnson, Joey is reading at grade level 4.2"). CBM may be useful in determining present levels

of performance, as a series of baseline measures can be used to describe student performance in specific ways that promote good goal-setting (Stecker, n.d.).

Selecting and Writing Academic Goals

The final regulations for AA-MAS indicate that the IEP must contain goals “based on the academic content standards for the grade in which the student is enrolled” [USED, 2007a, p. 9; 200.1(e)(2)(iii)]. Since the AA-MAS must cover the same range of content as the typical grade-level assessment (USED, 2007b, p. 21), IEP teams will need to consider the full range of grade-level content standards for each subject to determine what strands or objectives represent priorities for instruction. It is important to remember that if a student takes an AA-MAS in one subject but is eligible for assessment against grade-level achievement standards in other subjects, the IEP is *not* required to include those areas assessed against grade-level standards. In other words, IEP goals are only required for areas in which the student’s disability is interfering with participation in the grade-level achievement test. However, the IEP team may choose to write goals for subjects in which the student is assessed against grade-level achievement standards.

In subjects the IEP team has determined will be assessed using the AA-MAS, the target for instruction is the student’s chronologically age-appropriate, grade-level content standards. Even if a state has designed its content standards as vertical progressions that build across grade levels, it is not acceptable to base IEP goals on academic standards at a lower grade level (NCEO, 2007).

Annual IEP goals represent priorities in the student’s curriculum that year. That does not mean that the content of a goal statement must match the wording of a state content standard. IEP teams may identify essential skills that support general curriculum access in that area. Indeed, teams may identify pivotal skills—those that cross academic subjects and promote growth across the curriculum (Browder, Spooner, Wakeman, Trela, & Baker, 2006). For

example, the team may write a goal for the student to learn certain strategies for monitoring comprehension of a reading passage. This self-evaluation skill may not be part of a reading content standard in that grade level, but a student who can monitor comprehension may make better progress in reading as well as other content areas. As Pellegrino (Chapter 4, this volume) notes, metacognitive skills are most effective if taught in specific content areas rather than outside the context of a content area. A team may determine that a specific content goal is the best choice to support the student's growth toward grade-level proficiency, but the team should be discouraged from treating the content standards as an à la carte menu.

Once the priorities for IEP goal contents have been determined, the team then follows best practices for writing effective IEP goals. Good IEP goals are measurable, which means they cannot be vaguely written and need to reflect an observable behavior. They also specify the conditions under which the behavior will be demonstrated and the criterion for performance. There are numerous guides for writing effective IEP goals. Two excellent examples are Bateman & Herr (2006) and Courtade-Little & Browder (2005).

Monitoring Goal Attainment

With IDEA 2004, IEP teams are no longer required to write short-term objectives (STOs) for each goal [except for students who will be assessed using AA-AAS; see USED, 2007b, p. 29, IDEA 614(d)(1)(A)(I)(cc)]. Without this requirement, IEP teams will need to be diligent in thinking through and planning for progress monitoring. Progress must be reported at least as frequently as it is for students without disabilities (which typically translates to a report card cycle). It is hard to imagine that AA-MAS-eligible students can successfully accelerate their learning if teachers are doing the bare minimum to meet the letter of the law for monitoring progress. Bateman and Herr (2006) offer a convincing argument that teams should write STOs to operationalize the broader goal, even though that step is no longer federally required. Teams

may also develop and document other methods in accordance with their state's interpretation of the federal requirement to monitor progress.

Regardless of whether teams write STOs for these students, it is important that the criterion reflected in the goal be based on projected growth from present level of performance, in a given period of time, with a target that sets the student on a path toward grade-level achievement. Ideally, methods for monitoring goal attainment would follow some of the CBM or progress monitoring approaches described earlier in this chapter. Again, these methods would be based on technically sound, appropriate assessment instruments, administered at a frequency that allows teachers to monitor and report progress and adjust instruction without becoming a burdensome process that detracts from instruction.

Choosing and Designing Supports

As mentioned previously, IEP teams will design educational programs for AA-MAS-eligible students that allow for achievement based on modified rather than grade-level achievement standards, while also building toward eventual participation in assessments based on grade-level achievement standards. IEP teams will decide how to design appropriate supports for learning, without setting the standard so low that the student is not held to the highest expectations possible that year. The IEP can provide clear markers about what supports are needed, in which parts of the curriculum, and under what conditions. If the team instead provides blanket coverage of certain supports across *all* subject areas and settings (including accommodations, positive behavior support plans, and ESY services), the team may end up overcompensating for the student's disability and providing supports that are not needed or even serve to limit student growth.

One decision support tool IEP teams can use to plan for academic instruction is provided in Table 3-1 (Quenemoen, 2009). Teams can ask themselves a series of questions to consider individual student strengths and needs within the context of his or her educational system.

These questions would be asked and answered differently for each student, since each student may have different priorities for the academic year.

Table 3-1. Question Guide to Support IEP Team Decision-Making Process

SYSTEM OPPORTUNITIES AND CHALLENGES—representatives from school and district must provide this information and support	STUDENT STRENGTHS AND NEEDS—IEP team grapples with data-based decision-making to identify services, supports, and specialized instruction so that the student will be successful
1. What is the required content in the next grade level?	1. What are the student’s current strengths and needs in the academic content areas? What data do we have to make that determination? What accelerated or remedial services and supports are necessary to ensure success in the content for the next grade level?
2. Where and when during the school year is the required content taught in Math? ELA? Science? Social Studies? Other? How is it taught?	2. What adaptations and accommodations can the student use to access the grade-level content regardless of specific deficit basic skills in reading or mathematics or English language? What data do we have to support these choices? How will we determine if their use is effective or needs changing?
3. What instructional and curricular options are currently available in this school to allow all students to achieve proficiency in the goals and standards set for all students?	3. What specific instructional strategies work well for this student? What types of curricular materials work well for this student? If the needs of this student and the options available don’t align well, what aids, services, supports, and instruction does the student need to be successful in spite of gaps? How can the current options be changed?
4. What array of services does the school provide to meet the students’ other needs?	4. What specific nonacademic needs does this student have? What goals and objectives will address those needs? How do these relate to the student’s academic success?
5. What is the curricular map into the future, what are essential understandings every student needs to achieve this year?	5. How can we set priorities to ensure the essential understandings are mastered by this student, but still allow the TIME the school and the student needs to address all needs.
6. How does all this go together, with professional development, support, continuous improvement, community linkages . . . so that ALL children are successful?	6. How do we align curriculum, instruction, supports, services, and needs so that THIS child is successful at grade level?

Reprinted with permission from Quenemoen (2009).

It will be important for IEP teams to carefully consider the implication of choosing a particular type of adaptation for the student’s educational program and progress. If the content is significantly modified, teachers run the risk of unintentionally lowering expectations for the student. If the student is eligible for the AA-MAS *because* his or her disability precluded success in assessments based on grade-level achievement standards [as required in 34 CFR § 200.1(e)(2)(i)], designing instruction that changes the construct may lead teachers to continue a pattern of working around the disability (i.e., modification) rather than eliminating its influence

through effective supports (i.e., accommodation). Again, the educators on IEP teams will need to be very skilled in choosing adaptations and evaluating their impact on the student.

State Guidance to IEP Teams

Just as teachers design supports for student learning, states offer supports to IEP teams so educators can effectively meet the lofty goals described earlier. As a matter of routine, states already provide guidance to IEP teams on how to manage the “paperwork” aspect of IEPs (i.e., adhering to procedural aspects of IDEA, appropriate documentation of required IEP components). However, now that the IEP is a living document rather than a bureaucratic formality (Sopko, 2003), states have opportunities to provide a wide array of guidance to help teams design effective instructional programs as well. Recognizing that districts and individual schools act as a filter to interpret state guidance, this section highlights supports that states may wish to provide IEP teams planning for this target population of students.

Decisions about Participation in AA-MAS

Federal regulations to guide eligibility for AA-MAS are detailed in the final AA-MAS regulations and elsewhere in this volume (see Perie, Chapter 1, this volume). At a minimum, states are responsible for clearly conveying to the IEP team:

- that a student from any disability category may be eligible for AA-MAS [34 CFR §200.1(f)(1)(ii)]
- various options for students with disabilities to participate in statewide assessment (including assessments based on grade level, modified, or alternate achievement standards), including the impact of assessment choice on the student’s educational options according to State or local policy [34 CFR §200.1(f)(1)(iii)]
- clear and appropriate guidelines to determine when AA-MAS is the appropriate assessment option [34 CFR §200.1(f)(1)(i)]

- that students may be assessed against modified achievement standards in one or more subjects for which assessments are administered under §200.2 [34 CFR §200.1(f)(2)(i)]

Teams must also use a pattern of data to determine the appropriate assessment participation option (USED, 2007b, p. 18). States may also wish to give IEP teams guidance on when it would be appropriate for a student to resume participation in state assessments based on grade-level achievement standards.

Regardless of states' eligibility criteria, IEP teams will need clear guidance that allows them to differentiate in rather nuanced situations. Specific criteria and supporting rationales will help IEP teams understand how the criteria apply to each student. States may find that a flowchart, checklist, or other decision support tool may help teams correctly apply the criteria to arrive at appropriate eligibility determinations. Lazarus, Rogers, Cormier, & Thurlow (2008) reviewed states' AA-MAS participation guidelines and found 15 categories of criteria across nine states that had an operational AA-MAS. Most frequently occurring were criteria based on the regulatory language (e.g., must have an IEP; decision not made based on a specific disability label not progressing at a rate at which the student would be expected to reach grade-level proficiency within the year; the use of multiple measures to determine previous performance; student cannot demonstrate knowledge on grade-level assessment even with appropriate accommodations). Some states also provided exclusion criteria, such as not being based on placement setting, not being eligible for AA-AAS, and not having a pattern of past performance attributable to excessive absences or other non-instructional factors (e.g., cultural, language, economic).

Lazarus et al. (2008) also compiled the materials they reviewed from the nine states into their synthesis report. As states think about how to provide clear guidance to IEP teams about appropriate assessment participation decisions, they may wish to adapt some of these published examples for their own use. For example, Maryland provides an IEP Team Decision-Making Process Eligibility Tool that teams can use to compile past assessment data, past

instruction, and clear evidence of specific content-area interventions and relevant IEP goals in the past years. That type of tool could be helpful to teams in determining that the student really had received effective instruction in prior years. As another example, North Dakota provides an 8-item checklist with an accompanying flowchart to guide teams to the most appropriate large-scale assessment option for the student. As more states move closer to deciding whether and how to provide AA-MAS options, additional examples will likely be forthcoming.

Decisions about Accommodations

IEP teams make choices about what accommodations are appropriate under which circumstances. Accommodations remove the influence of a disability on a student's ability to show what she or he knows and can do. Accommodations should not be confused with modifications, which are adaptations that change the construct being taught or assessed. IEP teams determine appropriate accommodations based on how the disability influences the individual student's participation in the educational program. Accommodation needs may differ by subject area or type of learning activity, and are determined separately for instruction, classroom assessment, and statewide assessment. Accommodations typically fall into one of four categories: presentation format, response format, timing, and setting (AERA, APA, & NCME, 1999). IEP teams are to choose the right kinds of accommodations for the right student in order to reduce construct-irrelevant variance and allow for valid inferences about student performance on assessments. The chosen accommodations should also provide sufficient support for the student to fully participate in instruction. Accommodations are discussed in more detail in Abedi (Chapter 9, this volume).

Unfortunately, there is a history of IEP teams having difficulty correctly interpreting accommodations (Byrnes, 2008) and choosing and applying accommodations (Shriner & Destefano, 2003) using best-practice principles. There may be confusion on the team about issues such as when to use accommodations, and the difference between an accommodation

and a modification. When choosing accommodations for AA-MAS, IEP teams are required to avoid accommodations that would, if used on the grade-level assessment, invalidate the score (USED, 2007b, p. 27).

IEP teams will also consider the match between accommodations given for instructional purposes and in large-scale assessments. There is extensive literature supporting the correspondence between these two, but IDEA gives IEP teams latitude to allow instructional accommodations that would be considered necessary “for a student to advance toward attaining his or her annual goals, to be involved in and make progress in the general curriculum, and to be educated alongside his or her nondisabled peers” [USED, 2007b, pp. 32-33; see also 34 CFR 300.320(a)(4)(i-iii)]. In other words, there is no legislation that precludes the use of accommodations in instruction that would invalidate results if used on a statewide assessment. Conversely, states may prohibit the use of accommodations on an assessment if they have not also been given in instruction. Given this potential mismatch, IEP teams will need to be deliberate in their choice of instructional accommodations beyond those that will be allowable for the assessment the student will take.

States can support the appropriate choice of accommodations in several ways. They can provide full descriptions of the types of accommodations, with examples, organized by category and with indications about where their use is appropriate (see New York’s guidelines, for example: <http://www.vesid.nysed.gov/specialed/publications/policy/testaccess/policyguide.htm>). States may also develop decision support materials or even computer-based tools, such as those described by Kopriva, Koran & Hedgspeth (2007).

Regardless of which accommodations a state approves, the decisions should be made using the best data available. NCEO offers an online accommodation bibliography (<http://www2.cehd.umn.edu/NCEO/accommodations/>), which is a searchable database with annotated bibliographies that allow for quick scanning of outcomes across studies. As states deliberate which accommodations to allow under which conditions, they will need a process to

review those potential accommodations. One such process, using a panel review, is described in Almond & Karvonen (2007). Finally, Thurlow, Christensen, & Lail (2008) summarized a review of reviewer comments made about accommodations practices described in states' peer review materials. Lessons learned from past peer reviews may be informative to states that implement new accommodations policies or options for AA-MAS-eligible students.

Other Guidance

State and local educational agencies coordinate and sponsor professional development for teachers on a wide array of topics. They may not be able to cover every topic described in this chapter, but an effective needs assessment could reveal areas to prioritize in planning professional development on topics related to AA-MAS-eligible students. Special education bureaus may be able to sponsor professional development on topics such as determining academic priorities, designing learning progressions, differentiating and adjusting instruction, CBM, and progress monitoring. As new technologies or teaching strategies become available, professional development plans may be adjusted to incorporate promising practices. Job-embedded professional development (e.g., professional learning communities, instructional coaches) may be the best approach to support teachers as they develop these instructional skills (Joyce & Showers, 2002).

Also, since the instructional implications associated with AA-MAS are relatively new, states should not overlook the importance of addressing teacher beliefs about what this population of students can do and how to hold them to high expectations (see Quenemoen, Chapter 2, this volume). States may wish to provide guidance to IEP teams on how to set interim achievement targets that are lower without precluding students from being eligible for assessments based on grade-level achievement standards by the time they would need to take assessments linked to high school graduation requirements. The Access Center (2004) provides several other suggestions for how states can offer training on IEP writing.

States may also be able to warehouse on their Web sites a wide variety of sample materials, including well-structured IEP goals and objectives, links to information on research-based curricula, assistive technologies, and data collection materials for formative assessment. One example is the site developed by the Georgia Department of Education for teachers of students who take the state's AA-AAS (<http://gadoe.georgiastandards.org/impairment.aspx>). These Web sites may also offer a way for teachers to exchange their own created materials through a community working toward the same goal—figuring out how to do what nobody has yet done for these students. State-level assessment departments could coordinate resources related to large-scale assessment. Assessment and Special Education divisions could collaborate to provide guidance on how to link formative and summative assessments.

Finally, states and LEAs may provide guidance and professional development to strengthen collaboration among IEP team members. Successful instruction for AA-MAS-eligible students will require effective collaboration between general educators and special educators. In schools where the responsibility for educating this population still lays solely with the special education staff, administrators may want to guide a shift toward the philosophy that the whole school is responsible. The principal, as instructional leader for *all* students, may need to send a strong message in order to prompt this shift.

One other aspect of collaboration should not be neglected. Parents are partners on the IEP team and must be equipped to be meaningful contributors to the planning that will take place. They will need to understand what the AA-MAS is, and what participation means for their children [§200.1(f)(1)(iv)]. If the instructional program under AA-MAS represents a fork in the road leading away from past practices, parents will also need to understand what to expect differently in terms of progress monitoring, reporting progress to families, and mid-year instructional changes. One excellent resource for this information is a parent guide published by the National Center on Educational Outcomes (Cortiella, 2007).

Systems for Monitoring IEPs

According to IDEA 2004 regulations, the state's primary responsibility for monitoring IEPs includes "(1) Improving educational results and functional outcomes for all children with disabilities; and (2) Ensuring that public agencies meet the program requirements under Part B of the Act, with a particular emphasis on those requirements that are most closely related to improving educational results for children with disabilities" [300.600(b)(1-2)]. While IEP monitoring tends to focus on procedural compliance, states may need to consider how they will monitor the IEP's role in ensuring that AA-MAS-eligible students receive highly effective instruction in grade-level content.

Review for substance rather than procedural compliance will be a significant challenge and states may find such reviews to be resource-intensive. One option is to implement an educational benefit review, such as the one designed by New York State Department of Education (2008). This review, conducted with a small sample of IEPs from local education agencies selected annually on a rotating cycle, focuses on whether the IEP was "reasonably calculated for the student to receive educational benefit." (p. 1). The process starts when the review team extracts information from the IEP about the student's educational program, including services, supports, needs, annual goals, accommodations, and modifications. The team then analyzes relationships among components of the IEP to determine whether needs, goals, and services were aligned to promote progress within a 3-year cycle. The comparison includes relationships across three consecutive, annual IEPs to determine if there is evidence of educational benefit such as progress toward goal attainment and increased complexity in goals. Finally, the review team evaluates the results for evidence of best practices and areas for improvement, and makes a final determination about whether the program was designed to result in educational benefit for the student.

Greater standardization of IEP content and format will lighten the burden of conducting substantive IEP reviews. Some states have implemented centralized IEP warehouses with

Web-based interfaces. These systems ease the monitoring process in two ways. IEP contents are more standardized (or at least organized the same way across districts), and the central location allows for easier access to facilitate reviews.

Some guidance is emerging on how states might monitor compliance with specific elements of the regulations. For example, Burdette (2009) conducted surveys and follow-up interviews with state education agencies regarding their procedures for monitoring IEPs for evidence of student progress toward annual goals. Another option would be to adapt a framework like the one Karger (2004) proposed for evaluating how IEPs reflect a student's access to the general curriculum. If a state chose this type of systematic review of the substantive elements, a consistently-applied content analysis procedure would be established. Rater training and calibration would be needed to support the quality of judgments for high-inference ratings.

Validity Evidence

This chapter has addressed standards-based IEPs as vehicles for designing effective curriculum and instruction in order to help AA-MAS-eligible students access grade-level content and make progress toward grade-level proficiency. This section describes how instructional issues contribute to the validity argument for AA-MAS.

The clearest link between instruction and assessment validity is opportunity to learn (OTL). If scores on the AA-MAS are intended to reflect modified achievement of grade-level content standards, one premise is that students have had the opportunity to learn the content that is assessed. Marion (chapter 9, this volume) offers two sample theories of action that illustrate how states might articulate the role of instruction in their validity arguments. One premise in his first example is that “teachers provide instruction that is aligned with these high academic expectations and ensure that students get the supports necessary to allow them to succeed with grade-level content” (pp. 320–321). If this statement was supported by evidence,

low scores would reflect limited achievement in the target domain, rather than lack of instruction in that domain.

For students who take AA-MAS, OTL requires instruction in the targeted content domain, with supports that removes barriers related to the disability. OTL is a matter of degree rather than an absolute dichotomy, and can be difficult to measure (AERA, APA, & NCME, 1999, Standard 13.5). Alignment studies provide one way of assessing correspondence between curriculum and assessment. Teacher-reported curriculum measures such as Surveys of Enacted Curriculum (Porter & Smithson, 2001) or the Curriculum Indicators Survey (Karvonen, Wakeman, Flowers, & Browder, 2007) are a way of incorporating curricular data in alignment studies.

IEPs also offer a source of evidence related to OTL, and can be used for other validity investigations as well. Content analysis can yield data such as:

- Curricular priorities reflected in the IEP, which can then be evaluated against grade-level content standards to document instruction in the intended content. Curricular priorities may also be compared with AA-MAS content, although that type of comparison does not yield strong “alignment” evidence.
- The quality of the instruction program for providing general curriculum access through:
 - Links between present levels of performance and annual goals
 - Correspondence between accommodations provided for instruction & assessment
 - Evaluative judgments about criteria for performance specified in the academic goals, and whether the expected growth from present level of performance to end of year goal reflects reasonable (and high) expectations
 - Appropriate use of other learning supports to promote meaningful access and remove barriers (e.g., ESY services, behavior intervention plans)
- Evidence of appropriate decisions made about student participation in AA-MAS.

Especially using retrospective analysis of multiple IEPs, or the IEP decision support tools

described in the eligibility section earlier in this chapter, investigators may determine whether the student's past pattern of instruction really does point to eligibility based on poor performance despite effective instruction.

Two investigations of this nature are being conducted at the time of this writing, although the studies have not yet reached the data analysis phase (Karvonen, 2009).

Finally, evidence of the instructional program may be linked to student test scores and desired student outcomes (AERA, APA, & NCME, 1999, Standard 13.9). While this type of investigation addresses validity evidence in cases where the score leads to high-stakes decisions for the individual student (e.g., promotion, graduation), one could argue that the stakes for AA-MAS-eligible students are high every year if they are to reach grade-level proficiency in the future. The comment on Standard 13.9 relative to special education is as follows:

...when test scores are used in the development of specific educational objectives and instructional strategies, evidence is needed to show that the prescribed instruction enhances students' learning. When there is limited evidence about the relationship among test results, instructional plans, and student achievement outcomes, test developers and users should stress the tentative nature of test-based recommendations...(p. 147)

Thus, if a state chooses a model in which performance on an AA-MAS leads automatically to a decision about provision of different services or a change in test eligibility, it would be important to know whether the educational program was designed to promote learning in the content that was assessed.

Conclusion

The federal guidance on AA-MAS places a greater responsibility for results on the quality of instruction compared with other types of alternate assessment. Fortunately, the role of

the IEP has become more central and is no longer seen as a compliance document (Sopko, 2003). A good, standards-based IEP can guide educators to support meaningful learning so students can work toward grade-level proficiency. States have noted benefits to using a standards-based approach to IEPs including higher levels of student achievement, integration of special and general education, and benefits to parents who participate in the IEP process (Ahearn, 2006). Aside from representing the concept of opportunity to learn, there is some evidence that having IEPs aligned with state academic content standards makes a difference for instruction (McLaughlin, Nolet, Rhim, & Henderson, 1999) and achievement (Karvonen & Huynh, 2007).

Challenges and Caveats

This chapter has presented suggestions for IEP teams to set curricular priorities, build services and supports into the educational program, and guide teachers in developing effective instruction. The chapter also offers suggestions to states on providing guidance and support to IEP teams in order to promote highly-effective practices for students who are not achieving at grade level. These recommendations are intended to help the field move forward toward optimal practices that help students meet high expectations. In some areas, educators may be prepared to enact these strategies immediately. For example, there is an extensive body of practice on how to design and implement formative assessment. In other areas, there is still a significant gap between current, realistic practice and the ideal. These gaps exist at the state, local, and school/teacher levels.

States have many competing priorities and will need to decide what is pragmatic in the systems and supports they establish. State and local agencies may begin to set paths toward ideal practice by evaluating where they are and how to reach the next logical step. For example, existing IEP monitoring systems could be reviewed in light of the AA-MAS regulations to evaluate what elements of the system would need to be changed in order to meet federal

requirements. In order to move the system forward, state stakeholders could define “ideal” practice as it relates to IEP content. They could then develop and pilot that system on a limited basis, using early data for formative purposes before scaling up.

The largest challenges at the school level lie in the lack of research—both on cognitive models of learning and content progressions for this population, and on effective instructional practices. Where there is a longer history of research (e.g., CBM practices), it may be challenging to distill and disseminate these strategies to teachers in ways that promote change in their practices. When resources are limited, instructional leaders at the state and local level may have less time to analyze promising new strategies that are worth disseminating to their teachers. Even when strong evidence exists for particular strategies, and that evidence is shared with teachers, there is still the conundrum of interpreting the AA-MAS requirement that students be taught in grade-level content standards even when they may not have mastered prerequisite or foundational content. Local agencies may wish to weigh the potential benefits of focusing their professional development on a broad view for designing and using IEPs in effective instructional planning versus professional development on separate topics related to good instruction (e.g., detailed support on translating academic content, how to collect formative data, how to choose accommodations). Predominant needs may vary by district or school.

As teachers develop IEPs that are consistent with federal mandates for AA-MAS, it is important to remember that academics are part of the broader educational program. Instructional time is finite, and there are many competing priorities for these students (Thurlow, 2008). Where possible, teachers will need to think creatively about how to capitalize on relationships with other parts of the program (e.g., academics embedded in therapeutic or transition goals; a focus on learning strategies or academic goals that support growth across the curriculum) and provide instruction that combines academics with other values for the population. Without these combinations, the risk is that instruction will become more

fragmented. This fragmentation may be detrimental to the goal of moving the student toward being prepared for assessment against grade-level achievement standards.

The final AA-MAS regulations emphasize the need for this target population of students to access grade-level content standards rather than modified content standards in order to maintain high expectations for student learning (USED, 2007b, p. 17,755). There are still many questions left unanswered about how students can access grade-level content, make more progress than they have made previously, and have their achievement measured against modified standards using assessments that allow for valid inferences about what they know and can do. If students who are eligible for AA-MAS remain identified as persistently low-performing over several years, what are the consequences for those students in the future? Are they permanently disadvantaged? Although participation in AA-MAS is not supposed to preclude participation in requirements for a high school diploma [34 CFR 200.1(f)(2)(iv)], what are the long-term consequences if the educational program does not help students make up for earlier learning deficits? Aside from the social justice implications, what are the economic costs for failing to help students be adequately prepared for successful completion of high school graduation requirements (Levin, 2009)? As described by Quenemoen (Chapter 2, this volume), the population of low-performing students also includes many students *without* disabilities. If educators successfully design and implement effective instructional programs for struggling students with disabilities, they also have the potential to lead the way on instruction for all students who are persistently low-performing — not just those with disabilities. Individual growth or learning plans would allow teachers to translate the ideas behind the IEP as a means of instructional planning to the rest of the population of low-performing students, in order to promote success for *all* struggling students.

References

- Access Center. (2004). *Aligning IEPs with state standards and accountability systems*. Washington, DC: American Institutes for Research. Retrieved April 22, 2009 from http://www.k8accesscenter.org/training_resources/aligningieps.asp
- Access Center. (2006). *Teaching matters: The link between access to the general education curriculum and performance on state assessments*. Washington, DC: American Institutes for Research. Retrieved April 22, 2009 from www.k8accesscenter.org/documents/TeachingMattersBrief_001.pdf
- Ahearn, E. (2006, May). *Standards-based IEPs: Implementation in Selected States*. Alexandria: Project Forum, National Association of State Directors of Special Education (NASDE). Retrieved April 21, 2009 from <http://www.projectforum.org/docs/Standards-BasedIEPs-ImplementationinSelectedStates.pdf>
- Almond, P., & Karvonen, M. (2007). *Accommodations for a K to 12 standardized assessment: Practical implications for policy*. In C. Cahalan Laitusis & L. L. Cook (Eds.), *Large-scale assessment and accommodations: What works?* (pp. 117-136). Arlington, VA: Council for Exceptional Children.
- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (1999). *Standards for educational and psychological testing*. Washington, DC: AERA.
- Bateman, B. D., & Herr, C. M. (2006). *Writing measurable IEP goals and objectives* (2nd ed.). Verona, WI: Attainment Company.
- Browder, D. M., Spooner, F., Wakeman, S., Trela, K., & Baker, J. N. (2006). Aligning instruction with academic content standards: Finding the link. *Research and Practice for Persons with Severe Disabilities*, 31(4), 309-321.
- Burdette, P. (2009, February). *State tracking to measure student progress toward IEP goals*. Alexandria, VA: National Association of State Directors of Special Education. Retrieved March 30, 2009 from <http://www.projectforum.org/docs/StateTrackingtoMeasureStudentProgressTowardIEPGoals.pdf>
- Byrnes, M. (2008). Educators' interpretations of ambiguous accommodations. *Remedial and Special Education*, 29(5), 306-315.
- California Department of Education (2008, December 18). Educational benefit activity. Sacramento, CA: Author.
- CAST (2008). *Universal design for learning (UDL) guidelines - Version 1.0*. Wakefield, MA: Author. Retrieved April 21, 2009 from <http://www.cast.org/publications/UDLguidelines/version1.html>
- Cortiella, C. (2007). *Learning opportunities for your child through alternate assessments: Alternate assessments based on modified achievement standards*. Minneapolis, MN:

University of Minnesota, National Center on Educational Outcomes. Retrieved April 21, 2009 from <http://cehd.umn.edu/nceo/OnlinePubs/AAMASParentGuide.pdf>

- Courtade-Little, G., & Browder, D. (2005). *Aligning IEPs to academic standards for students with moderate and severe disabilities*. Verona, WI: IEP Resources.
- Deno, S. L. (2003). Developments in curriculum-based measurement. *Journal of Special Education, 37*(3), 184-192.
- Gibson, V. & Hasbrouck, J. (2008). *Differentiated instruction: Grouping for success*. Boston: McGraw-Hill.
- Hardman, M. L. & Dawson, S. (2008). The impact of federal public policy on curriculum and instruction for students with disabilities in the general classroom. *Preventing School Failure, 52*(2), 5 – 11.
- Hess, K. (2008, June). *Developing and using learning progressions as a schema for measuring progress*. Paper presented at 2008 CCSSO National Conference on Student Assessment, Orlando, FL. Retrieved March 31, 2009 from http://www.nciea.org/cgi-bin/pubspage.cgi?sortby=pub_date
- Hill, W., & Erwin, R. (1984). The readability of content textbooks used in middle and junior high schools. *Reading Psychology, 5*(1), 105-117.
- Individuals with Disabilities Education Improvement Act of 2004, P. L. No. 108-446, 20 U.S.C. section 611-614.
- Jenkins, J. R., Graff, J. J., & Miglioretti, D. L. (2009). Estimating reading growth using intermittent CBM progress monitoring. *Exceptional Children, 75*(2), 151-163.
- Jitendra, A. K., Nolet, V., Xin, Y. P., Gomez, O., Renouf, K., Iskold, L., & DaCosta, J. (2001). An analysis of middle school geography textbooks: Implications for students with learning problems. *Reading & Writing Quarterly, 17*(2), 151-173.
- Joyce, B. R., & Showers, B. (2002). *Student achievement through staff development* (3rd ed.). Alexandria, VA: Association for Supervision and Curriculum Development.
- Karger, J. (2004). *Access to the general education curriculum for students with disabilities: The role of the IEP*. Wakefield, MA: National Center on Accessing the General Curriculum. Retrieved from http://www.cast.org/publications/ncac/ncac_iep.html.
- Karvonen, M. (2009). *IEP content analysis protocols*. Cullowhee, NC: Western Carolina University.
- Karvonen, M., & Huynh, H. (2007). The relationship between IEP characteristics and test scores on alternate assessments for students with significant cognitive disabilities. *Applied Measurement in Education, 20*, 273-300.
- Karvonen, M., Wakeman, S. L., Flowers, C. P., & Browder, D. M. (2007). Measuring the enacted curriculum for students with significant cognitive disabilities: A preliminary investigation. *Assessment for Effective Intervention, 33*(1), 29-38.

- Kopriva, R., Koran, J., & Hedgspeth, C. (2007). Addressing the importance of systematically matching student needs and test accommodations. In C. Cahalan Laitusis & L. L. Cook (Eds.), *Large-scale assessment and accommodations: What works?* (pp. 145-165). Arlington, VA: Council for Exceptional Children.
- Lazarus, S. S., Rogers, C., Cormier, D., & Thurlow, M. L. (2008). *States' participation guidelines for alternate assessments based on alternate achievement standards (AA-MAS) in 2008* (Synthesis Report 71). Minneapolis, MN: National Center on Educational Outcomes.
- Levin, H. M. (2009). The economic payoff to investing in educational justice. *Educational Researcher*, 38(1), 5-20. doi: 10.3102/0013189X08331192
- McLaughlin, M. J., Nolet, V., Rhim, L. M., & Henderson, K. (1999). Integrating standards: Including all students. *Teaching Exceptional Children*, 31(3), 66-71.
- NASDE (2007). *A seven-step process for creating standards-based IEPs*. Available at: <http://cehd.umn.edu/nceo/Teleconferences/AAMASTeleconferences/SevenStepProcess.pdf>
- National Center on Educational Outcomes [NCEO]. (2007, May 17). *Teleconference notes. Standards-based IEPs and IEP goals based on grade-level standards*. Available at: <http://www.education.umn.edu/NCEO/USED2percentTele051707.pdf>.
- National Center on Response to Intervention (n.d.) *What is RTI?* Retrieved April 21, 2009 from <http://www.rti4success.org/>
- National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics*. Reston, VA: Author.
- National Reading Panel. (2000). *Report of the National Reading Panel, teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction*. Washington, DC: National Institute for Literacy.
- New York State Education Department (2008, December 23). *Special education quality assurance: IDEA effective instructional practices focused review manual*. Albany, NY: New York State Education Department, Office of Vocational & Educational Services for Individuals with Disabilities.
- Nitko, A. J. (2004) Formative evaluation using informal diagnostic assessments. In *Educational assessment of students* (4th ed., pp. 288-303). Upper Saddle River, NJ: Pearson Education.
- Perie, M. (2009, February). *Understanding the AA-MAS: How does it fit into a state assessment and accountability system?* Presentation to the SCASS groups, Orlando, FL. Retrieved March 26, 2009 from http://www.nciea.org/publications/CrossSCASS_MAP09.pdf
- Porter, A. C., & Smithson, J. L. (2001, December). *Defining, developing, and using curriculum indicators* (CPRE Research Report Series RR-048). Retrieved April 22, 2009 from http://www.cpre.org/images/stories/cpre_pdfs/rr48.pdf

- Quenemoen, R. (2009). *Success one student at a time: What the IEP team does*. Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes. Retrieved April 20, 2009 from <http://www.nceo.info/Tools/StandardsIEPtool.pdf>
- Quenemoen, R., Thurlow, M., Moen, R., Thompson, S. & Morse, A. B. (2004). *Progress monitoring in an inclusive standards-based assessment and accountability system* (Synthesis Report 53). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes. Retrieved April 20, 2009 from <http://education.umn.edu/NCEO/OnlinePubs/Synthesis53.html>
- Shriner, J., & Destefano, L. (2003). Participation and accommodation in state assessment: The role of Individualized Education Programs. *Exceptional Children*, 69(2), 147-161.
- Sopko, K. M. (2003). *The IEP: A synthesis of current literature since 1997*. Alexandria, VA: National Association of State Directors of Special Education, Project FORUM. Retrieved April 21, 2009 from <http://www.projectforum.org/docs/iep.pdf>
- Soukup, J. H., Wehmeyer, M. L., Bashinski, S., & Bovaird, J. A. (2007). Classroom variables and access to the general curriculum for students with disabilities. *Exceptional Children*, 74(1), 101-120.
- Stecker, P. M. (n.d.). *Monitoring student progress in Individualized Educational Programs using curriculum-based measurement*. Washington, DC: National Center on Student Progress Monitoring. Retrieved April 21, 2009 from http://www.studentprogress.org/library/monitoring_student_progress_in_individualized_educational_programs_using_cbm.pdf.
- Stecker, P. M., Fuchs, L. S., & Fuchs, D. (2005). Using curriculum-based measurement to improve student achievement: Review of research. *Psychology in the Schools*, 42(8), 795-819. doi: 10.1002/pitts.20113
- Thurlow, M. L. (2008). Assessment and instructional implications of the alternate assessment based on modified academic achievement standards (AA-MAS). *Journal of Disability Policy Studies*, 19(3), 132-139. doi: 10.1177/1044207308327473
- Thurlow, M., L., Christensen, L. L., & Lail, K. E. (2008). *An analysis of accommodations issues from the standards and assessments peer review* (Technical Report 51). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes. Retrieved April 20, 2009 from <http://cehd.umn.edu/NCEO/OnlinePubs/Tech51/index.htm>
- U.S. Department of Education (2007a, April 9). 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. Available at <http://www.ed.gov/admins/lead/account/saa.html#regulations>
- U.S. Department of Education (2007b, July 20), Modified Academic Achievement Standards: Non-regulatory Guidance. Washington, DC: Office of Elementary and Secondary Education, U.S. Department of Education. Available at: <http://www.ed.gov/admins/lead/account/saa.html#regulations>.

CHAPTER 4

THE CHALLENGES OF CONCEPTUALIZING WHAT LOW ACHIEVERS KNOW AND HOW TO ASSESS THEIR COMPETENCE

Jim Pellegrino

This chapter considers some of the most important issues surrounding the “what” and “how” of assessment as applied to any population of students, but especially those students who fall in the range of low academic achievement as measured by the typical achievement tests used for purposes of NCLB accountability. It is intended as a bridge between the prior two chapters, with their focus on identification of those students whose academic performance is such that assessment relative to modified achievement standards may be appropriate, and the next section of this report with its focus on issues regarding the content and design of any such assessment.

The first section of this chapter is concerned with two “big ideas” that are essential to the development, implementation and use of assessments for any group of students: (1) understanding relationships among assessment, curriculum and instruction, and (2) conceptualizing assessment as a process of reasoning from evidence that should be driven by theories and data on student cognition. The second section then elaborates on one key element from section one — those aspects of cognition that underlie student knowledge and performance in subject matter domains and that must be considered in the design and interpretation of student assessments. The third section builds from the prior two. It focuses on important aspects of student cognition and provides examples of their development in the instructional domains of reading and mathematics. The fourth section considers the implications of the preceding sections for multiple facets in the design of valid and useful student assessments. The final section concludes by considering the validity of assessment practices that might be considered for this population of students in light of the content of preceding

sections. It also suggests some things we still need to know to make progress in the design of valid assessments based on modified achievement standards.

Two Critical Issues for Conceptualizing Student Assessment

The Curriculum-Instruction-Assessment Triad

Whether we recognize it or not, assessment does not and should not stand alone. Rather, it is one of three central components in the educational enterprise — curriculum, instruction, and assessment. The three elements of this triad are linked, although the nature of their linkages and reciprocal influence is often less explicit than it should be. Furthermore, the separate pairs of connections are often inconsistent which can lead to an overall incoherence in the educational enterprise.

Curriculum consists of the knowledge and skills in subject matter areas that teachers teach and students are supposed to learn. The curriculum generally consists of a scope or breadth of content in a given subject area and a sequence for learning (Pugalee & Rickelman, Chapter 5, this volume). Content standards in each subject matter area typically outline the goals of learning, whereas curriculum sets forth the more specific means to be used to achieve those ends. *Instruction* refers to methods of teaching and the learning activities used to help students master the content and objectives specified by a curriculum. Instruction encompasses the activities of both teachers and students. It can be carried out by a variety of methods, sequences of activities, and topic orders. *Assessment* is the means used to measure the outcomes of education and the achievement of students with regard to important competencies. Assessment may include both formal methods, such as large-scale state assessments, or less formal classroom-based procedures, such as quizzes, class projects, and teacher questioning.

A precept of educational practice is the need for alignment among curriculum, instruction, and assessment (e.g., NCTM, 1995, 2000; Webb, 1997). Alignment, in this sense, means that the three functions are directed toward the same ends and reinforce each other

rather than working at cross-purposes. Ideally, an assessment should measure what students are actually being taught, and what is actually being taught should parallel the curriculum one wants students to master. If any of the functions is not well synchronized, it will disrupt the balance and skew the educational process. Assessment results will be misleading, or instruction will be ineffective. Alignment is difficult to achieve, however. Often what is lacking is a central theory about the nature of learning and knowing around which the three functions can be coordinated.

Most current approaches to curriculum, instruction, and assessment are based on theories and models that have not kept pace with modern knowledge of cognition and how people learn (e.g., NRC, 1999a, 1999b, 1999c, 2000, 2001a; Shepard, 2000). They have been designed on the basis of implicit and highly limited conceptions of cognition and learning. Those conceptions tend to be fragmented, outdated, and poorly delineated for domains of subject matter knowledge. Alignment among curriculum, instruction, and assessment could be better achieved if all three are derived from a scientifically credible and shared knowledge base about cognition and learning in subject matter domains. The model of learning would provide the central bonding principle, serving as a nucleus around which the three functions would revolve. Without such a central core, and under pressure to prepare students for the accountability tests, teachers may feel compelled to move back and forth between instruction and external assessment and teach directly to the items on a state test. This approach can result in an undesirable narrowing of the curriculum and a limiting of learning outcomes. Such problems can be ameliorated if, instead, decisions about both instruction and assessment are guided by a model of learning in the domain that represents the best available scientific understanding of how people learn (NRC, 2000).

Assessment as a Process of Reasoning from Evidence

Educators assess students to learn about what they know and can do, but assessments do not offer a direct pipeline into a student's mind. Assessing educational outcomes is not as straightforward as measuring height or weight; the attributes to be measured are mental representations and processes that are not outwardly visible. Thus, an assessment is a tool designed to observe students' behavior and produce data that can be used to draw reasonable inferences about what students know. Deciding what to assess and how to do so is not as simple as it might appear.

The process of collecting evidence to support inferences about what students know represents a chain of reasoning from evidence about student learning that characterizes all assessments, from classroom quizzes and standardized achievement tests, to computerized tutoring programs, to the conversation a student has with her teacher as they work through a math problem or discuss the meaning of a text. In the 2001 report *Knowing What Students Know: The Science and Design of Educational Assessment* issued by the National Research Council, the process of reasoning from evidence was portrayed as a triad of three interconnected elements — the *assessment triangle* (NRC, 2001a). The vertices of the assessment triangle represent the three key elements underlying any assessment: a model of student *cognition* and learning in the domain of the assessment; a set of beliefs about the kinds of *observations* that will provide evidence of students' competencies; and an *interpretation* process for making sense of the evidence. These three elements may be explicit or implicit, but an assessment cannot be designed and implemented without some consideration of each. The three are represented as vertices of a triangle because each is connected to and dependent on the other two. A major tenet of the *Knowing What Students Know* report is that for an assessment to be effective and valid, the three elements must be in synchrony. The assessment triangle provides a useful framework for analyzing the underpinnings of current

assessments to determine how well they accomplish the goals we have in mind, as well as for designing future assessments.

The *cognition* corner of the triangle refers to a theory or set of beliefs about how students represent knowledge and develop competence in a subject domain (e.g., fractions). In any particular assessment application, a theory of learning in the domain is needed to identify the set of knowledge and skills that is important to measure for the task at hand, whether that be characterizing the competencies students have acquired thus far or guiding instruction to further increase learning. A central premise is that the cognitive theory should represent the most scientifically credible understanding of typical ways in which learners represent knowledge and develop expertise in a domain. More will be said in the next section about what we know about the nature of cognition and the development of subject matter competence.

Every assessment is also based on a set of beliefs about the kinds of tasks or situations that will prompt students to say, do, or create something that demonstrates important knowledge and skills. The tasks to which students are asked to respond on an assessment are not arbitrary. They must be carefully designed to provide evidence that is linked to the cognitive model of learning and to support the kinds of inferences and decisions that will be made on the basis of the assessment results. The *observation* vertex of the assessment triangle represents a description or set of specifications for assessment tasks that will elicit illuminating responses from students. In assessment, one has the opportunity to structure some small corner of the world to make observations. The assessment designer can use this capability to maximize the value of the data collected, as seen through the lens of the underlying beliefs about how students learn in the domain.

Every assessment is also based on certain assumptions and models for interpreting the evidence collected from observations. The *interpretation* vertex of the triangle encompasses all the methods and tools used to reason from fallible observations. It expresses how the observations derived from a set of assessment tasks constitute evidence about the knowledge

and skills being assessed. In the context of large-scale assessment, the interpretation method is usually a statistical model, which is a characterization or summarization of patterns one would expect to see in the data given varying levels of student competency. In the context of classroom assessment, the interpretation is often made less formally by the teacher, and is usually based on an intuitive or qualitative model rather than a formal statistical one.

A crucial point is that each of the three elements of the assessment triangle not only must make sense on its own, but also must connect to each of the other two elements in a meaningful way to lead to an effective assessment and sound inferences. Thus to have an effective assessment, all three vertices of the triangle must work together in synchrony. Central to this entire process, however, are theories and data on how students learn and what students know as they develop competence in aspects of the curriculum.

Fundamental Components of Cognition and Some Implications for Assessment

This section begins the process of specifying some of what we know about the nature of human cognition that has implications for instruction, learning, and assessment. We begin at the level of what is general about the human cognitive system in terms of how it processes information and the types of knowledge and skill that are developed over time through instruction and practice. This “generic” description is meant to apply to the minds of virtually all individuals with the possible exception of those with the most severe cognitive impairments. The utility of such a generic description is that it provides a basis for considering some of the sources of differences in learning and performance that might be associated with persistently low levels of academic achievement. We consider such implications at the end of this section before moving to Section III where we examine some of what we know about specific aspects of cognition and learning in the domains of reading and mathematics and what that might imply for assessment.

Working Memory

One of the chief theoretical advances to emerge from cognitive research is the notion of *cognitive architecture*--the information processing system that determines the flow of information and how it is acquired, stored, represented, revised, and accessed in the mind. One of the most critical components of the cognitive architecture is *working memory*. It has been conceptualized as the system we use to process and act on information that is immediately before us and that we are consciously processing (Baddeley, 1986). Rather than viewing working memory as a "place" in the cognitive system, contemporary theoretical work has conceptualized working memory as a kind of cognitive energy level or "resource" that exists in limited amounts, with substantial individual variations. It is a well established fact that there are reliable developmental and individual differences in working memory capacity that predict a range of cognitive outcomes including scores on conventional tests of intelligence and achievement (e.g., Unsworth & Engle, 2007).

A significant aspect of the construct of working memory is that it plays a central role in virtually any cognitive activity we can imagine, determining the success or failure of many if not most of our intellectual endeavors. The range of activities that are impacted by the capacity and efficiency of one's working memory includes such things as executing a procedure like multicolumn addition or subtraction while monitoring the products of the process and the sequential steps; the act of representing and learning a new procedure like learning to "borrow across zero" in multicolumn subtraction; and the process of reading and comprehending a piece of narrative or expository text, including activities such as resolving issues of reference and making inferences (e.g., Miyake, Just, and Carpenter, 1994).

Metacognition

The term *metacognition* (literally "thinking about thinking") is commonly used to refer to the selection and monitoring processes, as well as to more general activities of reflecting on and

directing one's own thinking. Good learners have strong metacognitive skills (Hatano, 1990). They monitor their problem-solving, question limitations in their knowledge, and avoid oversimplistic interpretations of a problem. In the course of learning and problem-solving, such individuals display certain kinds of regulatory performance such as knowing when to apply a procedure or rule, predicting the correctness or outcomes of an action, planning ahead, and efficiently apportioning cognitive resources and time.

There is ample evidence that metacognition develops over the school years; for example, older children are better than younger ones at planning for tasks they are asked to do. Metacognitive skills can also be taught. For example, people can learn mental devices that help them stay on task, monitor their own progress, reflect on their strengths and weaknesses, and self-correct errors. It is important to note, however, that the teaching of metacognitive skills is often best accomplished in specific content areas since the ability to monitor one's understanding is closely tied to domain-specific knowledge and expertise.

Types of Knowledge and Processes of Acquisition

Long-term memory contains two distinct types of information — information about “the way the world is” (declarative knowledge) and procedural information about “how things are done” (procedural knowledge). It is one thing to know what it means to throw a 90-mile-per-hour fastball for a strike in baseball and quite another to be able to actually do it! Knowing about something (making a soufflé) is not the same as actually being able to do that thing. Much of what we would like students to learn in school is a combination of both declarative and procedural knowledge and for both types of knowledge we want them to access and use that knowledge in a highly fluent and relatively automatic fashion.

Unlike working memory, long-term memory is, for all practical purposes, an effectively limitless store of information. What matters most in learning situations is not the capacity of working memory—although that is often a factor in the speed and/or accuracy of processing—

but how well one can evoke the knowledge stored in long-term memory and use it to reason efficiently about information and solve problems in the present.

As part of studying the nature of knowledge in long-term memory, researchers have probed deeply the nature of competence and how people acquire large bodies of knowledge over long periods of time. Studies have revealed much about the kinds of mental structures that support problem-solving and learning in various domains; what it means to develop competence in a domain; and how the thinking of high achievers differs from that of novices (e.g., Chi, Feltovich, & Glaser, 1981). What distinguishes high from low performers is not simply general mental abilities, such as memory or fluid intelligence, or general problem-solving strategies. High performers have acquired extensive stores of knowledge and skill in a particular domain. But perhaps most significant, their minds have organized this knowledge in ways that make it more retrievable and useful. Because their knowledge has been encoded in a way that closely links it with the contexts and conditions for its use, high achievers do not have to search through the vast repertoire of everything they know when confronted with a problem. Instead, they can readily activate and retrieve the subset of their knowledge that is relevant to the task at hand (Simon, 1980; Glaser, 1992). Such findings suggest that teachers should place more emphasis on the conditions for applying the facts or procedures being taught, and that assessment should address whether students know when, where, and how to use their knowledge.

Considerable effort has also been expended on understanding the characteristics of persons and of the learning situations they encounter that foster the development of expertise. Much of what we know about the development of expertise has come from studies of children as they acquire competence in many areas of intellectual endeavor, including the learning of school subject matter. (This is further discussed in the section titled "Domain Specific Aspects of Cognition and Learning.") From a cognitive standpoint, *development* and *learning* are not the same thing. Some types of knowledge are universally acquired in the course of typical development, while other types are learned only with the intervention of deliberate teaching

(which includes teaching by any means, such as apprenticeship, formal schooling, or self-study). Infants and young children appear to be predisposed to learn rapidly and readily in some domains, including language, number, and notions of physical and biological causality. Infants who are only 3 or 4 months old, for example, have been shown to understand certain concepts about the physical world, such as the idea that inanimate objects need to be propelled in order to move (Massey & Gelman, 1988). By the time children are 3 or 4 years old, they have an implicit understanding of certain rudimentary principles for counting, adding, and subtracting cardinal numbers (Gelman, 1990; Gelman & Gallistel, 1978).

In math, the fundamentals of ordinality and cardinality appear to develop in all non-disabled human infants without instruction. In contrast, however, such concepts as mathematical notation, algebra, and Cartesian graphing representations must be taught. Similarly, the basics of speech and language comprehension emerge naturally from millions of years of evolution, whereas mastery of the alphabetic code necessary for reading typically requires explicit instruction and long periods of practice (Geary, 1995). Much of what we want to assess in educational contexts is the product of such deliberate learning.

With respect to assessment, one of the most important findings from detailed observations of children's learning behavior is that children do not move simply and directly from an erroneous to an optimal solution strategy (Kaiser, Proffitt, and McCloskey, 1985). Instead, they may exhibit several different but locally or partially correct strategies (Fay and Klahr, 1996). They also may use less advanced strategies even after demonstrating that they know more advanced ones, and the process of acquiring and consolidating robust and efficient strategies may be quite protracted, extending across many weeks and hundreds of problems (Siegler, 1998). These studies have also found, moreover, that short-term transition strategies often precede more lasting approaches and that generalization of new approaches often occurs very slowly.

The Role of Practice and Feedback

Every domain of knowledge and skill has its own body of concepts, factual content, procedures, and other items that together constitute the knowledge of that field. In many domains, including areas of mathematics and science, this knowledge is complex and multifaceted, requiring sustained effort and focused instruction to master. Developing deep knowledge of a domain such as that exhibited by high achievers, along with conditions for its use, takes time and focus and requires opportunities for practice with feedback.

Whether considering the acquisition of some highly specific piece of knowledge or skill such as the process of adding two numbers, or some larger schema for solving a mathematics or physics problem, certain laws of skill acquisition always apply. The first of these is the *power law of practice*: acquiring skill takes time, often requiring hundreds or thousands of instances of practice in retrieving a piece of information or executing a procedure. This law operates across a broad range of tasks, from typing on a keyboard to solving geometry problems (Anderson, 1981; Rosenbloom & Newell, 1987). According to the power law of practice, the speed and accuracy of performing a simple or complex cognitive operation increases in a systematic nonlinear fashion over successive attempts. This pattern is characterized by an initial rapid improvement in performance, followed by subsequent and continuous improvements that accrue at a slower and slower rate.

Practice, however, is not enough to ensure that a skill will be acquired. The conditions of practice are also important. The second major law of skill acquisition involves *knowledge of results*. Individuals acquire a skill much more rapidly if they receive feedback about the correctness of what they have done. If incorrect, they need to know the nature of their mistake (Thorndike, 1931). One of the persistent dilemmas in education is that students often spend time practicing incorrect skills with little or no feedback. Furthermore, the feedback they ultimately receive is often neither timely nor informative. Unguided practice (e.g., homework in math) can be for the less able student, practice in doing tasks incorrectly. One of the most

important roles for assessment is the provision of timely and informative feedback to students during instruction and learning so that their practice of a skill and its subsequent acquisition will be effective and efficient (Black & Wiliam, 1998; Sadler, 1989; Wiliam, 2007).

The Role of Social Context, Cultural Norms, and Student Beliefs

Much of what humans learn is acquired through discourse and interactions with others. For example, science, mathematics, and other domains are often shaped by collaborative work among peers. Through such interactions, individuals build communities of practice, test their own theories, and build on the learning of others. For example, those who are still using a naive strategy can learn by observing others who have figured out a more productive one. This situation contrasts with many school situations, in which students are often required to work independently or even competitively. Yet the display and modeling of cognitive competence through group participation and social interaction is an important mechanism for the internalization of knowledge and skill in individuals (Rogoff, 1990). Studies suggest that much of knowledge is also highly “situated”—it is embedded within systems of representation, discourse, and physical activity. A part of developing competence is learning to participate in communities of practice which in turn serve as sites for developing identity as a member of various communities and what happens in those communities as enabled by a variety of artifacts and tools (Lave, 1988).

The beliefs students hold about learning are another social dimension that can significantly affect learning and performance (e.g., Dweck & Legitt, 1988). For example, many students believe, on the basis of their typical classroom and homework assignments, that any math problem can be solved in 5 minutes or less, and if they cannot find a solution in that time, they will give up. Many young people and adults also believe that talent in mathematics and science is innate, which gives them little incentive to persist if they do not understand something

in these subjects immediately. Conversely, people who believe they are capable of making sense of unfamiliar things often succeed because they invest more sustained effort in doing so.

If mathematics is presented by a teacher as a set of rules to be applied, students may come to believe that “knowing” math means remembering which rule to apply when a question is asked (usually the rule the teacher last demonstrated), and that comprehending the concepts that underlie the question is too difficult for ordinary students. In contrast, when teachers structure math lessons so that important principles are apparent as students work through the procedures, students are more likely to develop deeper understanding and become independent and thoughtful problem-solvers (Lampert, 1986).

Some Implications for Low-Achieving/performing Students

What are some possible implications of the cognitive architecture and the nature of knowledge and its development for understanding the performance of low-achieving students? It would nice if we could provide definitive answers to such a question but in many cases we lack a research base that allows us to do so. Nevertheless we can speculate on some of the possible causes of low performance and the implications for both instruction and assessment. For example, some of the problem may be an information processing bottleneck issue, especially as regards the capacity of working memory and the management of attentional resources. Such a bottleneck has implications for the processes of learning and knowledge acquisition as well as for performance in a testing complex. It may well be the case that the ability to integrate content and to proceduralize knowledge, which are key aspects of the process of learning, are slowed or impaired by limitations in basic processing capacities. This is not to say that individuals cannot acquire the knowledge that is intended but rather that the speed and conditions needed to do so may differ. Similarly, differences in performance in a testing situation may have little to do with the availability of the appropriate knowledge but the load on working memory that taxes the person’s capacity to manage the situation within the time

demands of the testing situation. Similar issues arise regarding aspects of metacognition and the capacity to develop and/or exercise such skills in a given learning or performance situation.

Without convincing evidence that it is the architecture per se that contributes to low achievement, it is reasonable to assume that much of the problem of low achievement represents a deficit in the nature of the forms of knowledge that are demanded by different areas of the curriculum. This almost sounds tautological — low achievement by definition means lack of knowledge. But low achievement may not be associated with a lack of knowledge per se but a failure to develop the forms of knowledge that are associated with higher levels of competence and performance. If students perform poorly on tests of domain-specific achievement it is appropriate to ask how much of the problem may result from a failure of sufficient opportunity to learn the content required to attain higher levels of competence. In turn, much of that deficiency might be a function of the failure to make explicit for such students that which is often tacit in the learning situation and more readily discerned or inferred by non-disabled students. Learning is a process of constructing knowledge and such a constructive process occurs regardless of the forms of instruction—from guided discovery and hands-on experiences to collaborative learning to direct instruction to rote memorization. Because knowledge is constructed rather than delivered there is always a potential gap between what was intended in the instructional environment and what was actually understood and represented by the student.

As noted earlier, the development of knowledge is constituted within particular contexts and situations — an “interactionist” perspective of development (Newcombe & Huttenlocher, 2000). Accordingly, assessment of children’s development in contexts of schooling should include attention to the nature of classroom cultures and the practices they promote, as well as to individual variation. For example, the kinds of expectations established in a classroom for what counts as a mathematical explanation or what serves as a summarization or interpretation of a text affect the kinds of strategies and explanations that children pursue and the kinds of

responses they are likely to give in an assessment context. Because knowledge is constructed from experience, we may need to pay more attention to the nature of the experiences that low-achieving students encounter as the conditions of learning and how those experiences align with the conditions and expectations for performance in an assessment context. In essence, any assessment or testing situation is a test of transfer (Ruiz-Primo et al., 2002). What is near transfer for some students may be far transfer for others given the conditions of learning and the situated as well as socio-cultural nature of their knowledge (see, e.g., Hickey & Pellegrino, 2005; Pellegrino & Hickey, 2006).

Despite all we know about cognition, we must remind ourselves that there are many questions yet to answer about the ways in which low-achieving students differ from their regular education peers and the possible causes as well as consequences. Some of the possible answers lie in a better understanding of the nature of knowledge and skill in specific curricular domains and how that develops over time and with instruction.

Domain Specific Aspects of Cognition and Learning

Detailed models of cognition and learning in specific curricular areas can be invaluable for evaluating the progress of any individual or group, as well as for informing teaching and learning. In other words, a well-developed and empirically validated model of thinking and learning in an academic domain can be used to design and select assessment tasks that support the analysis of various kinds of student performance. Models with power highlight the main determinants of and obstacles to learning and include descriptions of students' conceptual progressions as they develop competence and expertise. Consistent with these ideas, there has been a recent spurt of interest in the topic of "learning progressions" (see NRC 2005, 2007). Learning progressions describe "successively more sophisticated ways of reasoning within a content domain that follow one another as students learn" (Smith et al., 2006).

Duncan and Hmelo-Silver (2009) have provided a description of “essential features” of learning progressions that attempts to capture an emerging consensus derived from panel discussions organized by the Center on Continuous Instructional Improvement and the Consortium for Policy Research in Education (Corcoran, Mosher, & Rogat, 2009). As described by Duncan and Hmelo-Silver (2009), there are four essential features that define something as a learning progression. First, learning progressions are focused on a few foundational and generative disciplinary ideas and practices. Several researchers have argued that it is the combined focus on content and practice which is unique to the current definition of learning progressions and central to the development of scientific literacy (Smith et al., 2006). Second, these progressions are bounded by an upper anchor describing what students are expected to know and be able to do by the end of the progression and by a lower anchor describing assumptions about the prior knowledge and skills of learners as they enter the progression. The upper anchor is informed by analyses of the domain as well as societal expectations. Third, they describe varying levels of achievements as the intermediate steps between the two anchors. These levels are derived from syntheses of existing research on student learning in the domain as well as empirical studies of the progression (such as cross sectional studies and teaching experiments). Levels of achievement are provided in the form of learning performances that can serve as evidence of students’ level of understanding and competency. Fourth, learning progressions are mediated by targeted instruction and curriculum. They are not developmentally inevitable and as such do not describe learning as it naturally develops in the absence of scaffolded curriculum and instruction.

Below we consider some of what we currently know about the components of competence and the progression of learning in the domains of mathematics and reading. We are not offering these descriptions as learning progressions that meet the criteria outlined above, but as illustrations of what we do know about how knowledge and competence develops over time and with instruction for certain aspects of the domains of reading and mathematics. In

considering the information that is provided about the sequence of learning and cognitive development we must remind ourselves that two pertinent questions, which we need to answer empirically, are whether low-performing students can be characterized as simply lagging behind in the pace of their development and whether they follow the same or different progressions. Clearly, being able to answer such questions is essential to the process of better educating these students as well as for providing valid and fair assessments that are tied to appropriately modified achievement standards that in turn have coherence within and across grade levels.

K-8 Reading

There is an unusual degree of consensus regarding the goals of early reading instruction. The consensus is captured in the National Research Council Report, *Preventing Reading Difficulties in Young Children* (NRC, 1998), and in the report of the National Reading Panel, *Teaching Children to Read* (NICHD, 2000). The goals are often expressed in terms of the competencies children should be able to demonstrate at the end of grade three: (a) read age-appropriate literature independently with pleasure and interest; (b) read age-appropriate explanatory texts with comprehension for the purpose of learning; and (c) talk and write about those texts in age-appropriate ways. Achieving these goals requires simultaneous development of an interdependent set of abilities: decoding skills, reading fluency, oral language development, vocabulary development, comprehension skills, and the ability to encode speech into writing.

The foundation for early reading lies in the earlier, informal acquisition of language. With little effort, children with intact neurological systems acquire the sounds of their language, its vocabulary, and its methods of conveying meaning (NRC, 1998). The path that children travel in acquiring language is predictable (NRC, 1998), though the age at which particular skills and abilities are mastered varies somewhat. As proficiency with language use grows, children develop the ability to *think about* language. Before that ability develops, children do not

distinguish between the word and the object to which it refers. Children can begin to develop rudimentary metalinguistic skills as early as age three. Acquiring this ability allows children to play with, analyze, and pass judgment on the correctness of language.

The trajectory of language development as described above is universal, though the richness of the environment affects the pace and extent of language development powerfully (Hart & Risley, 1995; Huttenlocher, 1998). For example, Graves and Slater (1987) found that first-graders from higher-income families had a vocabulary that was double the size of those from low-income families. The differences are highly relevant because verbal ability generally, and vocabulary development particularly, are good predictors of success in early reading.

While typical language development supports reading acquisition, other abilities required for effective reading mastery are unlikely to develop unless children receive formal instruction. With few exceptions, children need systematic instruction in the alphabetic principle to learn to decode words, and to learn how to encode words in writing (Adams et al., 1998). This instruction is what is referred to as “phonics.” But successful phonics instruction rests on a more fundamental ability: phonemic awareness. This is the awareness, for example, that the word “cat” consists of three separable sounds — c/ a/ t. The distinction is important because phonics instruction that teaches the mapping of separate sounds onto letters requires for success that a student hear those separate sounds.

Learning the alphabetic principle is prerequisite to reading. However it is not nearly sufficient to help children reach the desired third-grade competencies. Phonics instruction must be integrated with comprehension instruction, opportunities to develop fluency in reading through practice, instruction to enhance and practice oral and written language abilities, and opportunities to acquire rich vocabulary and background knowledge. The failure of any one of these will result in falling short of the third-grade goals. If fluency does not develop, little meaning is taken from a text that a student must plod through. If background knowledge is

inadequate, even a fluent reader will be unable to engage with and learn from the text. The components of successful reading are tightly intertwined.

In addition to building vocabulary, oral language instruction can extend a child's ability to understand and use academic, or literate, language. This is the decontextualized language that minimizes contextual cues and shared assumptions (e.g., by explicitly encoding referents for pronouns, actions, and locations (Olson, 1977). These extensions of discourse into the decontextualized register of academic language are what predict literacy success into middle school, controlling for home variables (Dickinson & Sprague, 2001). These relationships between preschool oral language and middle-school reading comprehension are clearly mediated by decoding instruction in the primary grades (Whitehurst & Lonigan, 2001). But the point is that language intervention that builds vocabulary and decontextualized language structures needs to occur *prior* to and *during* decoding instruction, rather than later.

Writing is at the heart of mastering the alphabetic system. Writing starts with the encoding of speech to print. The ability to phonemically segment sounds in speech and represent them in conventional writing develops over time. A complete representation of a word's spelling in memory developed through writing will enhance the speed and accuracy with which it is recognized (Ehri, 1998; Perfetti, 1992). Thus, the writing of words supports the reading of words and, over time, builds toward the writing of text, which can support the comprehension of text.

In addition to understanding the contributors to successful reading acquisition, there is also an extensive research base on the typical hurdles that children encounter (NRC, 1998; NICHD, 2000). It is now well established that a significant number of children have difficulty learning the alphabetic principle because they have not developed phonemic awareness. Among children who learn to decode words but do not comprehend well, fluency is often the culprit; if children struggle slowly through a text, their comprehension when they have finished will be poor. Fluency can suffer if children spend too little time actively engaged in effective

reading practice, or if vocabulary and background knowledge are too weak to allow the student to read with understanding (Lesgold & Perfetti, 1981).

In contrast to the above, in the area of reading comprehension much remains to be known as reflected in an assessment of research needs by the RAND Reading Study Group (RRSG; RAND, 2002a), as well as in the report of the National Reading Panel (NICHD, 2000). Those reports make clear that with regard to both student learning and teacher preparation, the research base to support practice is weak.

What Should Children Know and Be Able To Do? The answer to this question is sometimes given in terms of state or national standards for reading and language arts but such answers are often inadequate when it comes to development over time (see the discussion of standards for reading in Pugalee and Rickelman, chapter 5, this volume). An answer to this question is implied by the RRSG in its definition of reading comprehension as “the process of simultaneously extracting and constructing meaning through interaction and involvement with written language” (RAND, 2002a). To *extract* meaning requires the reader to decode the words and form a mental representation of what the text actually says, at both a local (sentences, phrases, and their interconnections) and global level (the “gist” of the text’s meaning). To *construct* meaning requires that the reader create a “situation model,” or an understanding of the intended meaning conveyed with these words that is informed not just by the text, but by the knowledge and experience that the reader brings (Kintsch, 1998). The situation model is the foundation from which inferences are drawn. Consider the sentence, “The sky was a clear, bright blue the day she first saw Charles.” The sentence does not state that it is not raining, but the reader can infer this from the bright blue sky. More importantly, it says nothing about whom Charles might be to the referenced woman, but we infer that he will be significant and memorable—not a plumber who will fix her drain then disappear.

We would be pleased if a six-year-old student could read the above sentence, and understand it semantically. But we would expect a 16-year-old student to develop a situation

model that is more complex due to greater developmental maturity, more experience with texts and text genres, and the benefits of instruction. The high school student might appreciate the expectation created by the author with two very simple phrases, and might productively reflect on how that expectation might change if the sky were dark and the wind threatened to carry away all in its path. And yet our understanding of the typical progression of student reading comprehension between ages 6 and 16 is poorly mapped, with a consequence that our instructional support for comprehension is poorly defined as well. As the RRSB argues, “without research-based benchmarks defining adequate progress in comprehension, we as a society risk aiming far too low in our expectations for student learning.”

Research in this area has far to go. Many research perspectives offer relevant insights (e.g., Graesser, Mills, & Zwaan, 1997; Pearson & Hamm, 2002), but as yet there are no integrated theories and companion models that provide a foundation for accumulating knowledge and guiding instruction. Moreover, mapping progress in reading comprehension requires that the phenomenon can be measured. Here again the knowledge base is weak. Worse, what we do know suggests that existing, commonly used measures of comprehension can be misleading. They capture meaning extraction and short-term memory, but these are not good predictors of meaning construction. Interventions that can improve short term recall can actually weaken inferencing capacity (Mannes and Kintsch, 1987). Both the mapping of progress in reading comprehension and the evaluation of instructional interventions to improve reading comprehension (e.g., Beck & McKeown, 2001; Beck et al., 1997; Pressley et al., 1989) depend on the development of assessments that can measure all its aspects, including the quality of the situation model.

K-8 Mathematics

Investment in recent decades by federal agencies and private foundations has produced a wealth of knowledge about the development of mathematical understanding, and

correspondingly has led to the development of curricula that incorporate such knowledge (e.g., Carpenter, Fennema, & Franke, 1996; Ginsburg, Greenes, & Balfanz, 2003; Griffin, Case, & Siegler, 1994). Much of contemporary research and theory is synthesized in a report on elementary mathematics (NRC, 2001b), and in the work of a RAND study group that produced a mathematics research agenda (RAND, 2002b). The National Research Council 2001 report presents a view of what elementary school children should know and be able to do in mathematics that draws on a solid research base in cognitive psychology and mathematics education, some of which is described below. It includes mastery of procedures as a critical element of mathematics competence, but places far more emphasis on understanding when and how to apply those procedures than is common in many mathematics classrooms. The latter is rooted in a deeper understanding of mathematical concepts, and a facility with mathematical reasoning.

The NRC committee summarized its view in five intertwining “strands” that constitute mathematical proficiency:

- *Conceptual understanding*—comprehension of mathematical concepts, operations, and relations;
- *Procedural fluency*—skill in carrying out procedures flexibly, accurately, efficiently, and appropriately;
- *Strategic competence*—ability to formulate, represent, and solve mathematical problems;
- *Adaptive reasoning*—capacity for logical thought, reflection, explanation, and justification;
- *Productive disposition*—habitual inclination to see mathematics as sensible, useful, and worthwhile, coupled with a belief in diligence and one’s own efficacy (NRC, 2001b).

Pugalee and Rickelman (Chapter 5 this volume) provide an excellent discussion of the mathematics content and process strands that have been articulated in the NCTM standards (NCTM, 2000) and that have in turn served as the basis for NAEP and state assessments. Much of that discussion aligns with aspects of the NRC's five areas of mathematical proficiency mentioned above. It is far beyond the scope of this chapter to try to capture what is known empirically about the multiple aspects of mathematical proficiency, including their development as a consequence of instruction. The literature on mathematical cognition and its development covers a diversity of topics, ranging from geometry problem solving to infant perception of numerosity (e.g., Greeno, 1978; Starkey & Cooper, 1980). However, it may be useful to consider some of what is known about even the most basic aspects of mathematical knowledge and competence. Accordingly, we have limited the discussion to current cognitive science accounts of performance on relatively basic aspects of mathematics, those that figure prominently in the early elementary school curriculum (see also Kalchman, Moss, & Case, 2001; Lesh & Landau, 1983; Schoenfeld, 1985).

The discussion that follows considers in some detail what we know about the basics of addition and subtraction, including computational procedures. The goal of doing so is to help those outside the research arena understand that even the "simplest" cognitive acts and instructional domains imply complicated forms of knowledge that are slowly acquired through experience and instruction. Furthermore, just as knowledge is not random, neither is performance, especially erroneous performance. This section concludes with a consideration of the potential value of all this detailed information for instruction and assessment. The reader may actually want to skim that concluding material before delving into what comes next.

Basic Addition. For many basic mathematics skills, expertise is necessarily defined in terms of the knowledge, processing activities, and performance of adults. Thus, to begin a discussion of cognitive analyses of basic mathematics we need to focus on theories of how adults do mental addition when faced with problems containing addends from 0–9 (e.g.,

Ashcraft, 1982; 1983, 1985, 1987). The theory assumes that adults have two basic types of mathematical knowledge. One type is an interrelated knowledge network containing the basic addition facts. As described earlier, such knowledge is referred to as declarative knowledge, i.e., knowledge of things that are true or false such as $2+3=5$. The facts stored in this network have different strengths that determine how long it takes to activate a piece of information. Thus, if the fact $2+3=5$ has greater associative strength than the fact $7+5=12$, it will take less time to retrieve (activate) the answer to the first of these two problems. The theory also assumes the existence of a second type of knowledge, specifically, methods that can be used to derive answers for problems lacking prestored answers, e.g., 14×36 vs. 4×6 . As described earlier, this is referred to as procedural knowledge, i.e., knowledge of "how to" do something. For single digit addition it might include procedures such as counting on from one of the addends an amount equal to the other addend. Adults actually have a variety of procedures for calculating answers, including shortcuts that make use of stored facts. An example is computing the answer to $28+25$ by retrieving the sum of $25+25$ and then adding 3 to 50.

This theory may seem to be nothing more than a restatement of what is intuitively obvious to any adult. For most of us, the "process" of adding single digit numbers is essentially the automatic retrieval of specific facts from memory. This process is rapid, automatic, effortless, and largely error free. What is less obvious is that such a theory of stored knowledge and retrieval processes provides the basis for explaining several phenomena observed in adults' time to produce or verify basic addition facts. One phenomenon is that adults produce answers very quickly, typically in less than a second (e.g., Ashcraft, 1985; Groen & Parkman, 1972). This can be attributed to the process of activating stored knowledge, a relatively rapid and automatic process, as opposed to computing answers by way of sequential procedures, a relatively slow and controlled process.

A second phenomenon is that the time to produce an answer systematically varies across problems. The slowest responses are for problems with "large" sums such as $9+8$, with

intermediate times for problems with medium sums such as $4+7$, and relatively fast responses for problems with small sums such as $2+1$, $3+2$ and for “ties” such as $4+4$, $7+7$, etc., and these problems are relatively homogeneous in time to respond (Ashcraft & Battaglia, 1978; Ashcraft & Stazyk, 1981; Groen & Parkman, 1972). As noted earlier, such differences in retrieval time are attributed to differences in the strength of specific facts. Stronger associations in the knowledge network are faster to activate.

A third phenomenon is that the time to reject a fact such as $4+3=12$ is substantially slower than the time for $4+3=10$, even though the first “answer” is actually further from the correct answer (Winkelman & Schmidt, 1974). Such effects are attributed to associative confusions between addition and multiplication facts. (See Ashcraft, 1982; 1985 for a more comprehensive summary of basic results in mental addition and multiplication.)

The aforementioned theory of expert solution of simple addition problems relies heavily on the assumption of differential associative strengths across the “basic facts” formed by the digits 0–9. An obvious question is whether this assumption is arbitrary or whether the assumed pattern of strength differences can be related to experiential phenomena. According to the law of frequency, items accrue strength through use and practice. Analyses of problem presentation frequency in children's mathematics texts indicate that those “basic facts” assumed to be stronger in the network actually appear more frequently in the texts (Ashcraft, 1985). Furthermore, analyses of multicolumn addition reveal that the frequency of adding 1,2, or 3 is greater than that of adding 7,8, or 9, consistent with strength patterns in the network.

Given that this theory is a plausible account of adult or expert performance, the question of developmental and instructional import concerns the nature of the progression from novice to expert. The acquisition of expertise in addition actually has its roots in the more general domain of number knowledge and quantitative understanding, acquisitions that are strongly tied to children's counting behavior (e.g., Gelman & Gallistel, 1978; Steffe, von Glaserfeld, Richards & Cobb, 1983). Prior to school entry most children have acquired relatively sophisticated counting

sequences for the digits 1-20 (Fuson & Hall, 1983; Gelman & Gallistel, 1978). Children also have a basic understanding of the "semantics" of addition and subtraction in terms of the combining and separating of quantities (e.g., Carpenter, 1985; Resnick, 1982, 1984). Their understanding of addition, in concert with their knowledge of counting, permits the solution of addition problems even in the absence of directly stored facts (e.g., Starkey & Gelman, 1982). Substantial evidence now exists that initial knowledge of addition consists of procedures for representing, combining and counting physical entities. Subsequently, addition can be performed as mental counting operations in the absence of physical objects. Such overt and covert operations constitute forms of procedural knowledge and processing that develop prior to and along with declarative knowledge and direct retrieval of addition facts (Fuson, 1982).

Evidence for an addition acquisition sequence of the type described above is of several types. First, young children with primitive counting skills often cannot solve simple addition problems if the objects representing one of the addends are hidden (Steffe, Thompson & Richards, 1982). Second, children are often observed counting fingers when solving addition problems (Fuson, 1982). Third, the counting procedures used by children transition from counting up to the cardinal value of the first addend and then counting on an amount equal to the second addend, to simply counting on from the first addend (Carpenter, 1985; Fuson, 1982; Houlihan & Ginsburg, 1981). Fourth, the time to do addition problems is closely related to counting rates for young children but not for older children (Ashcraft, Fierman, & Bartolotta, 1984). Fifth, systematic differences in the time to answer problems are consistent with models that minimize the number of counts, i.e., use of a procedure of counting on from the larger addend (Groen & Parkman, 1972; Svenson, 1975). Sixth, even for young children, there are some "facts" that are directly retrieved such as ties and small sums (Groen & Parkman, 1972; Hamann & Ashcraft, 1985; Siegler & Shrager, 1984).

A developmental theory of the acquisition of expertise in addition includes specific assumptions about the state of both declarative and procedural knowledge at different points in

time. It includes the assumption that there is a gradual acquisition and strengthening of the network structure of addition facts. There is also a gradual acquisition of counting procedures that permit the calculation of answers when "facts" are not of sufficient strength to be retrieved. Preschoolers primarily depend on overt counting procedures to solve addition problems (Siegler & Shrager, 1984). Given instruction and practice in the early grades, there is a transition to more sophisticated and efficient counting procedures together with a transition from calculation via counting to direct retrieval. Thus, at any point in time from preschool age through at least fourth grade, a child will have some facts that can be retrieved and some that need to be calculated. From the fourth grade on through adulthood, simple addition problems are solved via retrieval with a continued strengthening of facts in the network resulting in further increases in the speed of retrieving all addition facts (Ashcraft, 1985).

Subtraction. This discussion has concentrated on addition but the issues raised about the nature of expertise and its acquisition are equally applicable to simple subtraction problems. One can posit exactly the same type of theory of expertise for subtraction, with a network of stored facts of varying strength and a set of procedures for calculation in the absence of directly retrievable information. It is also reasonable to assume that subtraction facts vary in strength (speed of retrieval) although far less is known about the details of such differences and whether they parallel the results for addition.

With regard to procedural knowledge and the acquisition of expertise, there is ample evidence that preschoolers and children in the early primary grades solve subtraction problems by counting procedures, both overt and covert (Fuson, 1984; Svenson & Hedenborg, 1979; Woods, Resnick & Groen, 1975). Considerable research has been done on the use of different counting procedures to solve subtraction problems and the difficulties children sometimes experience in understanding and using such procedures (Fuson, 1984). One is a decrementing procedure in which the child *counts down* from the larger number (e.g. 9) an amount equal to the smaller number (e.g., 2). Another is an incrementing procedure in which the child *counts on*

from the smaller number (e.g., 7) until the larger number (e.g., 9) is reached. These procedures not only differ in ease of use but also in efficiency depending on problem characteristics. A decrementing procedure is more efficient when there is a large difference between the numbers (e.g., 9-2), while the converse is true for the incrementing procedure (e.g., 9-7). There is some evidence that older children select the optimal counting procedure given such differences in problem characteristics (Svenson & Hedenborg, 1979; Woods et. al., 1975).

A theory of expertise in subtraction and its acquisition is similar to the theory for addition. Both emphasize the gradual acquisition of declarative knowledge facts. These changes in knowledge and processing occur over a period of several years. The rate of change both within and between individuals will vary with the experiential history and learning rate of each person. Thus, one must consider the possibility that the difficulties in mathematics manifest by some children are partially attributable to problems with basic facts. The facts may be sufficiently weak such that they cannot be retrieved and must therefore be computed, and the counting procedures for doing such computations may be slow and error prone.

Data on basic addition and subtraction performance suggest that children with mathematics difficulties often must compute rather than directly retrieve answers to problems (e.g., Connor, 1983; Goldman, Mertz, & Pellegrino, 1988; Russell & Ginsburg, 1984). Connor (1983) has reported results obtained by Fleishner and her colleagues from testing basic facts. Learning disabled students relied more on reconstructive counting strategies than the nondisabled students who tended to rely on direct retrieval. This agrees with the results obtained by Russell & Ginsburg (1984) who compared a group of math-disabled fourth graders to nondisabled third and fourth graders. They observed particular difficulties in retrieving addition facts by math-disabled students, with the children performing at a level below the nondisabled third graders. Svenson and Broquist (1975) have also reported results indicating that fifth-grade children with low mathematics achievement are particularly slow at answering simple addition problems. Although available data are suggestive of difficulties in simple

addition and subtraction, considerably more must be done to pursue these issues. The theory of expertise and its acquisition that has been outlined above provides a framework for systematically pursuing issues regarding both the assessment and instruction of basic skills (see also Baroody, Bajwa, & Eiland, 2009).

Mathematical Procedures: Subtraction. Knowledge and performance in basic skills are particularly important when we consider more complex mathematical procedures that require facility in such skills. For example, the typical course of instruction is to progress from single column addition and subtraction problems to multicolumn problems of increasing difficulty. The ultimate objective is knowledge of complex procedures such that the individual can solve any addition or subtraction problem of any length. What do individuals know and do when they are "experts" in multicolumn addition or subtraction? There are now explicit theories of the knowledge underlying such complex skills, with primary attention given to subtraction (e.g., Brown & Burton, 1978; Young & O'Shea, 1981). Part of the emphasis on subtraction is attributable to the difficulties children often have in solving subtraction problems with borrowing, especially "borrowing from zero."

Knowledge of subtraction can be conceptualized as a complex procedure with multiple parts, each of which represents a successive complication. The essential parts are (1) processing single columns in a right to left order, (2) borrowing when the bottom digit in a column is greater than the top digit, and (3) borrowing from zero. These three parts correspond to the typical sequence in learning how to subtract. The child first learns how to subtract a single column of numbers where the top number is always greater than the bottom number. Then this is expanded to multiple columns but in problems where borrowing is never needed. The assumption is frequently made that the child subtracts two numbers in a column by retrieving a "fact" from memory such as $7-5=2$. However, a child might actually perform the subtraction for single digits by a counting procedure. The next major stage is to introduce the borrowing part of the procedure. This involves a test to see if the top number is greater than the bottom number in

a column. If it is, then borrowing is needed and the sequence of steps is taught. In beginning instruction this usually takes the form of crossing out the top digit in the column to the left, decrementing it by one then writing the new digit in the top of that column. The child then writes a 1 in front of the top digit in the original column and now goes on to do the column subtraction by retrieving a fact such as $17-9=8$. Practice in borrowing is provided with a progression to problems with multiple columns that require borrowing. The final stage of instruction is the procedure for borrowing from zero. The original borrowing procedure is now expanded to include a test for whether the column to the left contains a zero. If a zero is present then a new set of operations must be executed which include changing zero to 9 and moving one column to the left, testing for zero again etc.

The preceding is a superficial description of the overall procedure for doing multicolumn subtraction, its separate subprocedures and the general sequence for acquiring the subprocedures. Adults typically have procedural knowledge of subtraction as well as declarative knowledge of the meaning (semantics) of individual actions such as borrowing or borrowing from zero relative to the base ten system. It is not clear, however, whether children comprehend the meaning of the procedures taught to them. Analyses of children's errors in subtraction suggest that they often follow faulty procedures that preserve "syntactic" aspects of subtraction procedures such as crossing things out or writing down a 1 while simultaneously violating the semantics of the procedures (see e.g., Resnick, 1982, 1984).

Expertise can be defined as being able to solve any subtraction problem, which minimally implies knowledge of all the elements of the subtraction procedure. Lower levels of expertise are defined by the probability that errors will occur. Errors in subtraction can imply (a) lapses of attention or memory, what have been termed slips (Norman, 1981), (b) the absence of a procedure or a step in a procedure, or (c) incorrect representation of a procedure or a step in a procedure. If errors are due to lapses of attention or memory failure such as retrieving 2 for 9-6, then there should be no pattern to the errors made by the child. However, if a child is lacking

knowledge or has incorrect knowledge of a procedure then systematic error patterns should be observed within a child. To the extent that many children experience similar difficulty in acquiring and/or representing complex procedures, then one would expect to find consistent error patterns across children.

Considerable effort has been expended on analyzing children's errors on subtraction problems (Burton, 1981; Brown & Burton, 1978; Brown & VanLehn, 1980; Friend & Burton, 1981; VanLehn, 1983, 1990; Young & O'Shea, 1981). It is now apparent that errors are not just random, i.e., they cannot be attributed primarily to slips. Instead, errors tend to be systematic and the systematicity can be directly related to one or more of the elements of the major subprocedures of the complete subtraction procedure. As might be suspected, most of the systematic errors involve borrowing in general and borrowing from zero in particular. A common error is "smaller from larger" in which the child subtracts the smaller digit in a column from the larger regardless of which one is on top. This may be due to a child's lack of knowledge about how to borrow, a failure to incorporate a test for borrowing, or a carryover from simple subtraction where the smaller number is always "taken away" from the larger number and position doesn't matter.

Many common errors involve borrowing from zero. An example is changing zero to 9 but failing to decrement the column to the left of zero. A different type of error is borrowing across zero such that the column to the left of the zero is decremented by one but the zero is left unchanged. One final example involving borrowing from zero is to stop the borrowing process at zero. In this case the child correctly adds ten to the column where the top digit is less than the bottom digit but fails to make any change in either the column to the left containing zero or the column to the left of the zero. Another major set of errors involves the process of subtracting from zero within a column. In these cases the child fails to use any borrowing procedure and instead writes 0 or N for the column 0-N. For a more complete discussion of the most frequently

occurring errors in children's subtraction see Brown and Burton (1978), VanLehn (1983, 1990), and Young and O'Shea (1981).

One way to conceptualize the underlying source of these types of errors is in terms of slightly flawed procedural knowledge. The child has represented the procedures for performing subtraction but one or more of the elements is incorrectly represented, i.e., the child has a "bug" in his program for doing subtraction. The term bug is taken from computer programming and reflects an algorithm that contains an incorrect operation. A systematic error is produced each time the program is run on the particular class of problems that requires execution of that operation. An alternative possibility is that the child is missing a piece of procedural knowledge, which is similar to a critical operation being omitted from a program. In a computer program, a missing operation will typically cause the program to "crash" and produce no output whatsoever. However, in the case of a child who knows that some response must be made, the child reaches an impasse. In order to move on the child attempts to repair that impasse by doing something. The something he or she does is an operation that may mimic syntactic but not semantic constraints of subtraction.

Given that children's errors in subtraction reflect slips, bugs, and impasses (VanLehn, 1983, 1990), there are several issues with respect to the applicability of such a theory of knowledge and performance. One issue is a diagnosis of a child's problem. It is a nontrivial exercise to develop tests capable of isolating the many different types of procedural bugs and impasses that can occur, often in peculiar interaction, as well as a scoring procedure to do the diagnosis (Burton, 1981). Furthermore, multiple samples of performance are needed to determine if there is a stable pattern of bugs and/or impasses (see VanLehn, 1983). There are, however, some systematic efforts in this direction using instructional materials and computer-based tests (VanLehn, 1983).

Other issues involve explaining the acquisition of flawed procedural knowledge and developing instructional methods that minimize such outcomes. A missing procedure that gives

rise to an impasse in solution may result from a failure on the part of a student to represent a specific operation. Thus, the child attempts to repair the overall subtraction procedure when an impasse is actually reached in solving a problem (Brown & VanLehn, 1980). If these repairs are practiced and fail to receive any corrective feedback they may become permanent bugs.

Another possibility is that a child initially misrepresents a procedure and then subsequently practices that flawed procedure, again without corrective feedback. Thus, bugs can arise from repairs to impasses, i.e., solution attempts for novel problems for which no procedure is represented. They can also arise from incorrect initial representations of correct procedures. In either case, the errors that children produce seem to follow many of the syntactic aspects of subtraction (crossing things out, writing 1 in a column, etc.) while violating some of the semantics of the procedures. Given this state of affairs, attempts have been made to investigate instructional methods that link more closely the semantics and syntax of complex procedures (Resnick, 1982, 1984). The hope is that such methods can minimize the development of flawed procedural knowledge.

It is almost a given that elementary school children experiencing difficulty in mathematics will demonstrate less than expert performance on problems requiring complex procedures. Concern then is whether the errors they make can be understood in terms of the theory of knowledge and performance described above. One possibility is that such children have all the correct procedures and that errors are due to slips and miscalculations associated with their weak "knowledge" of basic facts. This may be partially true (Russell & Ginsburg, 1984). A second possibility is that parts of the procedural knowledge are either missing or flawed, in which case the errors they make would be systematic. If there are systematic errors, then do these children exhibit "bugs" similar to those found in previous research or are their errors more bizarre? There is little in the way of systematic data to address these questions. Russell and Ginsburg (1984) have reported limited data indicating that math-disabled fourth graders have bugs similar to those exhibited by nondisabled, younger children. They offer a hypothesis of

"essential cognitive normality" in which math-disabled children are at the lower levels of expertise representing the knowledge and performance of younger children. Considerably more needs to be done to explore such a hypothesis as it applies to complex procedural skills, as well as other important aspects of mathematical proficiency as identified by the National Research Council (2001b).

Is All This Detail Necessary?

It is not uncommon for individuals to ask what useful purpose, beyond esoteric academic pursuits, is served by the foregoing consideration of what we know about the knowledge and cognitive processing underlying something as "simple" as basic reading or mathematics knowledge and skill? As mentioned earlier, the preceding was designed help those outside the research arena understand that even the "simplest" cognitive acts and instructional domains imply complicated forms of knowledge that are slowly acquired through experience and instruction. Furthermore, just as knowledge is not random, neither is performance, especially erroneous performance. In fact, some would argue that we can learn far more from mistakes than we do from correct answers. Unfortunately, test content and test scores focus on just the opposite. For one thing, test items are often far removed from a theory of the knowledge underlying the performance of interest, and test scores provide little in the way of information that is directly useful to teachers to guide instructional decision making. In a typical test, the items are sampled from some universe of possibilities and the emphasis is not on the individual problem but the score derived by aggregating over problems. This leads to a situation where the same score can have very different meanings but there is no way of knowing that since the focus is on the total score rather than the way in which the score was produced. If the research within cognitive science has told us anything it is that the process by which a response is produced is far more important than the product. The same products can often result from very

different thinking processes, and testing procedures are frequently insensitive to such differences.

Consider, for example, a case where two children have systematic but different misconceptions involving borrowing in multicolumn subtraction. They might well achieve the same score by missing different problems. Even if they miss the same problem the nature of their errors might be different. Typical tests and test scoring procedures do not discriminate among these possibilities because they were not designed to do so nor do they provide any information about the incorrect choices that were made. A similar situation could arise with respect to tests of basic math facts. Tests of basic addition and subtraction facts are usually timed. What matters is the number of correct answers within the time period allotted. What is often ignored is how the number correct relates to the number attempted and the nature of the errors made on those attempted. In this regard the author is reminded of an actual situation that arose when one of his children brought home a test of addition and subtraction basic facts. All of the addition facts were correct but almost all of the subtraction facts were wrong. The note on the paper said that he should memorize his basic math facts. He was clearly distressed because he didn't know what the teacher meant. I examined his test and noticed that for all the subtraction fact answers that were incorrect they were off by 1. This suggested that he was not recalling his facts from memory but was using a counting scheme that had a systematic flaw or bug. I sat him down and got him to explain how he arrived at his answers and discovered that he was using a "counting down" procedure but with an extra count. I showed him how to correct his "buggy" procedure, he practiced the new one for a while until it was reliable and off he went content that he wasn't stupid and that he could now get the right answers. It was true that he still didn't "know" his subtraction facts but eventually he would because the counting procedure would yield the right answers and this in turn would give way to retrieval from memory once each of the facts was sufficiently strong to be associatively retrieved. The point of this little true example is that tests and testing procedures need to be brought in correspondence with current

theories of the nature of expertise in the domain of interest and the nature of the acquisition process. It is far more helpful to know that a child understands how to do subtraction and what it means, albeit he is less than fluent in fact retrieval, than to know that he misses 70% of all his subtraction facts.

There is an obvious challenge in translating theories about content knowledge and the acquisition of expertise into acceptable and workable instructional and testing procedures. To think that this is an easy task is to seriously underestimate the practical problems of the translation and implementation process. On the one hand, researchers must be willing to expend the time and effort needed to articulate their theories and assessment procedures in ways that are operationally feasible. Assessment developers must be willing to adopt new measurement models, scoring and reporting procedures. Educational practitioners must be willing to articulate their needs regarding the instructional monitoring functions they would like to perform and then find ways to incorporate new teaching and assessment technologies into daily classroom practices.

Implications for Assessment Design

Existing guidelines for assessment design emphasize that the process of assessment design should begin with a statement of the purpose for the assessment and a definition of the content domain to be measured (AERA, APA, NCME, 1999). A central thesis of this chapter is that the targets of inference should also be largely determined by a model of cognition and learning that describes how people represent knowledge and develop competence in the domain (the cognition element of the assessment triangle). Starting with a model of learning is one of the main features that distinguishes the proposed approach to assessment design from typical current approaches. The model suggests the most important aspects of student achievement about which one would want to draw inferences, and provides clues about the

types of assessment tasks that will elicit evidence to support those inferences (see also NRC, 2001a; Pellegrino, 1988; Pellegrino, Baxter, & Glaser, 1999).

The model of learning that informs assessment design should have as many as possible of the following key features. Ideally, it should:

1. Be based on empirical studies of learners in the domain.
2. Identify performances that differentiate beginning and expert performance in the domain.
3. Provide a developmental perspective, laying out typical progressions from novice levels toward competence and then expertise, and noting landmark performances along the way.
4. Allow for a variety of typical ways in which children come to understand the subject matter.
5. Capture some, but not all, aspects of what is known about how students think and learn in the domain. Starting with a theory of how people learn the subject matter, the designers of an assessment will need to select a slice or subset of the larger theory as the targets of inference.
6. Lend itself to being aggregated at different grain sizes so that it can be used for different assessment purposes (e.g., to provide fine-grained diagnostic information as well as coarser-grained summary information).

As described earlier, research on cognition and learning has produced a rich set of descriptions of domain-specific performance that can serve as the basis for assessment design, particularly for certain areas of reading, mathematics and science (e.g., NRC, 1998; 2000; 2001b; 2005, 2007; AAAS, 2001). Yet much more research is needed, especially with regard to students who are low achievers and who may have various identifiable learning or cognitive disabilities. This is despite the fact that a significant body of work already exists regarding students with disabilities and their performance in aspects of mathematics (e.g., Baroody et al., 2009; Fuchs, et al., 2005; Goldman, et al., 1988; Miller, 1997; Russell & Ginsburg, 1984;

Swanson & Jerman, 2006) and their performance in aspects of reading (e.g., Connor, 1983; Fletcher et al., 2002; Foorman & Torgesen, 2001; O'Connor & Jenkins, 1999; Torgesen, 2002; Wagner et al., 1993; Wagner et al., 1997).

What follows are some of the implications of the knowledge we do have for multiple aspects of assessment design and use. We begin with a consideration of issues related to assessment purpose and move to implications of cognitive research and theory for assessment that occurs in the context of the classroom and for state-level large-scale accountability assessment. Many of the specific topics that are touched on subsequently related to assessment design and use, including issues of validity and fairness, are developed in much greater depth in the chapters that follow in Sections III and IV of this overall report.

Assessment Purposes, Levels & Timescales

Although assessments are currently used for many purposes in the educational system, a premise of the *Knowing What Students Know* report (NRC 2001a) is that their effectiveness and utility must ultimately be judged by the extent to which they promote student learning. The aim of assessment should be “to educate and improve student performance, not merely to audit it” (Wiggins, 1998, p.7). Because assessments are developed for specific purposes, the nature of their design is very much constrained by their intended use. The reciprocal relationship between function and design leads to concerns about the inappropriate and ineffective use of assessments for purposes beyond their original intent. To clarify some of these issues of assessment purpose, design, and use it is worth considering two pervasive dichotomies in the literature that are often misunderstood and conflated.

The first dichotomy is between “internal” classroom assessments administered by teachers, and “external” tests administered by districts, states, or nations. Ruiz-Primo, Shavelson, Hamilton, and Klein (2002) showed that these two very different types of assessments are better understood as two points on a continuum that is defined by the

“distance” from the enactment of specific instructional activities. They defined five discrete points on the continuum of assessment distance: *immediate* (e.g., observations or artifacts from the enactment of a specific activity), *close* (e.g., embedded assessments and semi-formal quizzes of learning from one or more activities), *proximal* (e.g., formal classroom exams of learning from a specific curriculum), *distal* (e.g., criterion-referenced achievement tests such as required by the U.S. No Child Left Behind legislation), and *remote* (broader outcomes measured over time, including norm-referenced achievement tests and some national and international achievement measures. Different assessments should be understood as different points on this continuum if they are to be effectively aligned with each other and with curriculum and instruction.

A second pervasive dichotomy is the one between “formative” assessments used to advance learning and “summative” assessments used to provide evidence of prior learning. Often it is assumed that classroom assessment is synonymous with formative assessment, and that large-scale assessment is synonymous with summative assessment. What are now widely understood as different types of assessment practices are more productively understood as different functions of assessment practice, and that summative *and* formative functions can be identified for most assessment activities regardless of the level on which they function.

Drawing from the work of Lemke (2001), it is apparent that different assessment practices can be understood as operating at different *timescales*. The timescales for the five levels defined above can be characterized as *minutes, days, weeks, months, and years*. Timescale is important because the different competencies that various assessments aim to measure (and therefore the appropriate timing for being impacted by feedback) are “timescale-specific.” The cycles, or periodicity, of educational processes build from individual utterances into an individual’s lifespan of educational development. What teachers and students say in class constitute verbal exchanges; these exchanges make up the lesson; a sequence of lessons make up the unit; units form a curriculum, and the curricula form an education. Each of these

elements operates on different cycles or timescales: second to second, day to day, month to month, and year to year.

The level at which an assessment is intended to function, which involves varying distance in “space and time” from the enactment of instruction and learning, has implications for how and how well it can fulfill various functions of assessment, be they formative, summative, or program evaluation (see NRC, 2003). As argued elsewhere (Hickey & Pellegrino, 2005; Pellegrino & Hickey, 2006), it is also the case that the different levels and functions of assessment can have varying degrees of match with theoretical stances about the nature of knowing and learning. With this in mind we now turn to the implications of cognitive theory and research for both classroom assessment practices and for large-scale assessment. These two contexts reflect some of the rich variation in assessment captured by the foregoing discussion of levels, functions, and timescales.

Implications of Cognitive Theory & Research for Classroom Assessment

Shepard (2000) discusses ways in which classroom assessment practices need to change to better support learning: the content and character of assessments need to be significantly improved to reflect contemporary understanding of learning; the gathering and use of assessment information and insights must become a part of the ongoing learning process; and assessment must become a central concern in methods courses in teacher preparation programs. Her messages are reflective of a growing belief among many educational assessment experts that if assessment, curriculum, and instruction were more integrally connected, student learning would improve (e.g., Pellegrino, Baxter, & Glaser, 1999; Stiggins, 1997).

Sadler (1989) provides a conceptual framework that places classroom assessment in the context of curriculum and instruction. According to this framework, three elements are required for assessment to promote learning:

- A clear view of the learning goals (derived from the curriculum)

- Information about the present state of the learner (derived from assessment)
- Action to close the gap (taken through instruction)

Furthermore, there are ongoing, dynamic relationships among formative assessment, curriculum, and instruction. That is, there are important bidirectional interactions among the three elements, such that each informs the other. For instance, formulating assessment procedures for classroom use can spur a teacher to think more specifically about learning goals, thus leading to modification of curriculum and instruction. These modifications can, in turn, lead to refined assessment procedures, and so on. The mere existence of classroom assessment along the lines discussed here will not ensure effective learning. The clarity and appropriateness of the curriculum goals, the validity of the assessments in relationship to these goals, the interpretation of the assessment evidence, and the relevance and quality of the instruction that ensues are all critical determinants of the outcome. Starting with a model of cognition and learning in the domain can enhance each of these determinants.

For most teachers, the ultimate goals for learning are established by the curriculum, which is usually mandated externally (e.g., by state content standards). However, teachers and others responsible for designing curriculum, instruction, and assessment must fashion intermediate goals that can serve as an effective route to achieving the ultimate goals, and to do so effectively they must have an understanding of how students represent knowledge and develop competence in the domain. National and state content standards set forth learning goals, but often not at a level of detail that is useful for operationalizing those goals in instruction and assessment. By dividing goal descriptions into sets appropriate for different age and grade ranges, current content standards provide broad guidance about the nature of the progression to be expected in various subject domains. Whereas this kind of epistemological and conceptual analysis of the subject domain is an essential basis for guiding assessment, deeper cognitive analysis of how people learn the subject matter is also needed. Formative assessment should be based in cognitive theories about how

people learn particular subject matter to ensure that instruction centers on what is most important for the next stage of learning, given a learner's current state of understanding.

It follows that teachers need training to develop their understanding of cognition and learning in the domains they teach. Preservice and professional development are needed to uncover teachers' existing understandings of how students learn and to help them formulate models of learning so they can identify students' naive or initial sense-making strategies and build on those to move students toward more sophisticated understandings. The aim is to increase teachers' diagnostic expertise so they can make informed decisions about next steps for student learning. This has been a primary goal of cognitively based approaches to instruction and assessment that have been shown to have a positive impact on student learning, including the Cognitively Guided Instruction program (Carpenter, Fennema, and Franke, 1996) and others (Cobb et al., 1991; Griffin & Case, 1997). Such approaches rest on a bedrock of informed professional practice.

Implications of Cognitive Research and Theory for Large-scale Assessment

Large-scale assessments are further removed from instruction but can still benefit learning if well designed and properly used. Substantially more valid, useful, and fair information could be gained from large-scale assessments if the principles of design set forth above and described subsequently were applied. However, fully capitalizing on contemporary theory and research will require more substantial changes in the way large-scale assessment is approached, and relaxation of some of the constraints that currently drive large-scale assessment practices.

Large-scale summative assessments should focus on the most critical and central aspects of learning in a domain as identified by content standards and informed by cognitive research and theory. Large-scale assessments typically will reflect aspects of the model of learning at a less detailed level than classroom assessments, which can go into more depth because they focus on a smaller slice of curriculum and instruction. For instance, one might need to know for summative purposes whether a student has mastered the more complex aspects of multicolumn subtraction,

including borrowing from and across zero, rather than exactly which subtraction bugs lead to mistakes. At the same time, while policymakers and parents may not need all the diagnostic detail that would be useful to a teacher and student during the course of instruction, large-scale summative assessments should be based on a model of learning that is compatible with and derived from the same set of knowledge and beliefs about learning as classroom assessment.

As described previously, research on cognition and learning suggests a broad range of competencies that should be assessed when measuring student achievement, many of which are essentially untapped by current assessments. Examples are knowledge organization, problem representation, strategy use, metacognition, and participatory activities (e.g., formulating questions, constructing and evaluating arguments, contributing to group problem-solving). Furthermore, large-scale assessments should provide information about the nature of student understanding, rather than simply ranking students according to general proficiency estimates.

Large-scale assessments not only serve as a means for reporting on student achievement, but also reflect aspects of academic competence societies consider worthy of recognition and reward. Thus, large-scale assessments can signal worthwhile targets for educators and students to pursue. Whereas teaching directly to the items on a test is not desirable, teaching to the theory of cognition and learning that underlies an assessment can provide positive direction for instruction.

A major problem is that only limited improvements in large-scale assessments are possible under current constraints and typical standardized testing scenarios. Large-scale assessments are designed to meet certain purposes under constraints that often include providing reliable and comparable scores for individuals as well as groups; sampling a broad set of content standards within a limited testing time per student; and offering cost-efficiency in terms of development, scoring, and administration. To meet these kinds of demands, designers typically create assessments that are given at a specified time, with all students being given the same (or parallel) tests under strictly standardized conditions (often referred to as “on-demand” assessment). Tasks are generally of the kind that can be presented in paper-and-pencil format, that students can respond

to quickly, and that can be scored reliably and efficiently. In general, competencies that lend themselves to being assessed in these ways are tapped, while aspects of learning that cannot be observed under such constrained conditions are not addressed. To design new kinds of situations for capturing the complexity of cognition and learning will require examining the assumptions and values that currently drive assessment design choices and breaking out of the current paradigm to explore alternative approaches to large-scale assessment.

The Design of Observational Situations

Once the purpose for an assessment, the underlying model of learning in the domain, and the desired types of inferences to be drawn from the results have been specified, situations must be designed for collecting evidence to support the desired inferences about what students know and can do.

Task design. The focus should be on the cognitive demands of tasks (the mental processes and knowledge required for successful performance), rather than primarily on surface features, such as how tasks are presented to students or the format in which students are asked to respond. For instance, it is commonly believed that multiple-choice items are limited to assessing low-level processes such as recall of facts, whereas performance tasks elicit more complex cognitive processes. However, research shows that the relationship between item format and cognitive demands is not so straightforward (Baxter and Glaser, 1998; Hamilton, Nussbaum, and Snow, 1997).

Linking tasks to the model of cognition and learning forces attention to a central principle of task design—that tasks should emphasize the features that are relevant to the construct being measured and minimize extraneous features (AERA, APA, NCME, 1999; Messick, 1993). Ideally, a task will not measure aspects of cognition that are irrelevant to the targeted performance. For instance, when assessing students' mathematical reasoning, one should avoid presenting problems in contexts that might be unfamiliar to a particular population of

students. Similarly, mathematics tasks should not make heavy demands for reading or writing unless one is explicitly aiming to assess students' abilities to read or communicate about mathematics. Surface features of tasks do need to be considered to the extent that they affect or change the cognitive demands of the tasks in unintended ways.

Task difficulty. The difficulty of tasks should be explained in terms of the underlying knowledge and cognitive processes required, rather than simply in terms of statistical item difficulty indices, such as the proportion of respondents answering the item correctly (which ignores that two tasks with similar surface features can be equally difficult, but for very different reasons). Beyond knowing that 80 percent of students answered a particular item incorrectly, it would be educationally useful to know *why* so many did so, that is, to identify the sources of the difficulty so they could be remedied.

Cognitive theory and analysis can be helpful here. For instance, cognitive research shows that a mathematics word problem that describes the combining of quantities and seeks the resultant total (e.g., John has 3 marbles and Mary has 5, how many do they have altogether?) is easier to comprehend than one that describes the same actors but expresses a comparison of their respective quantities (e.g., John has 3 marbles. He has 2 less than Mary. How many does she have?) (see e.g., Morales, Shute & Pellegrino, 1985; Riley, Greeno, & Heller, 1983). Part of the difficulty for children is the conflict between the relational expression *less than*, which implies subtraction, and the operation required, which involves addition.

The point is not that such sources of difficulty should necessarily be avoided. Rather, these kinds of cognitive complexities should be introduced into the assessment tasks in principled ways in those cases in which one wants to draw inferences about whether students can handle them. There are many reasons why educators might want to assess students' abilities to apply integrated sets of skills (e.g., literacy and mathematics capabilities) to complex problems. That is entirely consistent with the approach being set forth here, as long as assessment design begins with a model of learning that describes the complex of skills,

understandings, and communicative practices that one is interested in making inferences about, and tasks are specifically designed to provide evidence to support those inferences.

Scoring. Tasks and the procedures to be used for drawing the relevant evidence from students' responses to those tasks must be considered together. That is, the ways in which student responses will be scored should be conceptualized during the design of a task. A task may stimulate creative thinking or problem solving, but such rich information will be lost unless the means used to interpret the responses capture the evidence needed to draw inferences about those processes. Like tasks, scoring methods must be carefully constructed to be sensitive to critical differences in levels and types of student understanding identified by the model of learning. At times one may be interested in the quantity of facts a student has learned, for instance, when one is measuring mastery of the alphabet or multiplication table. However, a cognitive approach generally implies that when evaluating students' responses, the focus should be on the quality or nature of their understanding, rather than simply the quantity of information produced. In many cases, quality can be modeled quantitatively; that is, even in very qualitative contexts, ideas of order and orderliness will be present.

Task sets and assembly of an assessment instrument. An assessment should be more than a collection of items that work well individually. The utility of assessment information can be enhanced by carefully selecting tasks and combining the information from those tasks to provide evidence about the nature of student understanding. Sets of tasks should be carefully constructed and selected to discriminate among different levels and kinds of understanding that are identified in the model of learning. To illustrate this point simply, it takes more than one item or a collection of unrelated items to diagnose a procedural error in subtraction. If a student answers three of five separate subtraction questions incorrectly, one can infer only that the student is using some faulty process(es), but a carefully crafted collection of items can be designed to pinpoint the limited concepts or flawed rules the student is using.

Validation

Traditionally, validity concerns associated with achievement tests have tended to center around test content, that is, the degree to which the test samples the subject matter domain about which inferences are to be drawn. There is increasing recognition within the assessment community that traditional forms of validation — which emphasize expert appraisal of the alignment of tasks with content frameworks, and their statistical consistency with other measures — should be supplemented with evidence of the cognitive or substantive aspect of validity (e.g., AERA, APA, NCME, 1999; Messick, 1993). That is, the trustworthiness of the interpretation of test scores should rest in part on empirical evidence that the assessment tasks actually tap the intended cognitive processes and knowledge.

As described by Messick (1993) and summarized by Magone, Cai, Silver, and Wang (1994), a variety of techniques can be used to examine the processes examinees use during task performance to evaluate whether prospective items are functioning as intended. These techniques include *protocol analysis*, in which students are asked to think aloud as they solve problems or to describe retrospectively how they solved the problems. Another method is *analysis of reasons*, in which students are asked to provide rationales for their responses to the tasks. A third method is *analysis of errors*, in which one draws inferences about processes from incorrect procedures, concepts, or representations of the problems.

Situative and sociocultural research on learning suggests that validation should be taken a step further. This body of research emphasizes that cognitive processes are embedded in social practices. From this perspective, performance of students on tests is understood as an activity in the situation that the test presents and success depends on abilities to participate in the practices of test taking (Greeno, Pearson, & Schoenfeld, 1996). It follows that validation should include the collection of evidence that test-takers possess the communicative practices required for their responses to be actual indicators of their abilities, for instance, to understand

and reason. This has been demonstrated to be false in many cases (e.g., Cole, Gay, and Glick, 1968).

Reporting

Although reporting of results occurs at the end of an assessment cycle, assessments must be designed from the outset to ensure that reporting of the desired types of information will be possible. The familiar distinction between norm-referenced and criterion-referenced testing (Glaser, 1963) is salient in understanding the central role of a model of learning in the reporting of assessment results. The notion of criterion-referenced testing has gained popularity in the last several decades, particularly with the advent of standards-based reforms in the 1990s. As a result of these reforms, many states are implementing tests designed to measure student performance against standards in core content areas.

Because criterion-referenced interpretations depend so directly on a clear explication of what students can or cannot do, well-delineated descriptions of learning in the domain are key to their effectiveness in communicating about student performance. Test results should be reported in relation to a model of learning. The ways people learn the subject matter and different states of competence should be displayed and made as recognizable as possible to educators, students, and the public to foster discussion and shared understanding of what constitutes academic achievement.

Fairness

Fairness in testing is defined in many ways (see AERA, APA, NCME, 1999), but at its core is the idea of comparable validity: a fair test is one that yields comparably valid inferences from person to person and group to group (NRC, 1999d). An assessment task is considered biased if construct-irrelevant characteristics of the task result in different meanings for different subgroups. Currently, bias tends to be identified through expert review of items. Such a finding is merely judgmental, however, and in and of itself may not warrant removal of items from an

assessment. Also used are statistical DIF (differential item functioning) analyses, which identify items that produce differing results for members of particular groups after the groups have been matched in ability with regard to the attribute being measured. However, DIF is a statistical finding and again may not warrant removal of items from an assessment. Some researchers have therefore begun to supplement existing bias-detection methods with cognitive analyses designed to uncover the reasons why items are functioning differently across groups, in terms of how students think about and approach the problems (e.g., Lane, Wang, and Magone, 1996; Zwick and Ercikan, 1989).

A particular set of fairness issues involves the testing of students with disabilities. A substantial number of children who participate in assessments do so with accommodations intended to permit them to participate meaningfully. For instance, a student with a severe reading and writing disability might be able to take a chemistry test with the assistance of a computer-based reader and dictation system. Unfortunately, little evidence currently exists about the effects of various accommodations on the inferences one might wish to draw about the performance of individuals with disabilities (NRC, 1997), though some researchers have taken initial steps in studying these issues (Abedi, Hofstetter, & Baker, 2001). Therefore, cognitive analyses are also needed to gain insight into how accommodations affect task demands, as well as the validity of inferences drawn from test scores obtained under such circumstances.

Conclusions & Caveats

A major thesis of this chapter is that the task of developing assessments tied to modified achievement standards needs to take seriously what we know about aspects of human cognition and its development, especially in domains of academic achievement and performance. This is much easier said than done, especially when we try to consider students who are hard to define and classify (see Quenemoen, Chapter 2, this volume) and whose

performance in the regular education context and on general academic achievement tests consistently leaves much to be desired. Many of the issues raised in this chapter pertain to the assessment of any and all students. It would behoove us to pay as much attention to a careful definition of academic achievement standards and assessments for the majority of students as we might propose to give to a subgroup of the population for whom we wish to define “modified achievement standards” and develop appropriate assessments. While there is much we know about aspects of the development of competence in the regular education population there is much that we don’t know about “cognition” in selective parts of the school age population. This knowledge gap has major implications for defining appropriate assessment targets and appropriate modes of assessment.

Given that we can never really know all that we need to know, and that there are pragmatic problems to be solved in the design of assessments for students with typical low levels of achievement, it is perhaps useful to consider some of the modifications that have been proposed for the assessment of this group of students and whether such design features can be justified. One such example is reducing cognitive load in questions through various means such as presenting a smaller number of distracters. While such a modification might well reduce construct irrelevant variance associated with working memory or metacognitive monitoring issues, it is a far cry from the type of deep engagement with assessment design issues related to the measurement construct that one would like to see. Such modifications may in fact change the nature of the construct assessed, but they do little to engage the issue of what the targets for learning should be and how such targets are manifest in the types of knowledge tapped by specific test items. Furthermore, work needs to be done to establish the cognitive processing validity of such modifications, which includes ruling out the possibility that improved performance has little to do with the nature of the knowledge and skill that is the intended achievement target. This is not to say, however, that attention to issues of processing load and

“construct irrelevant variance” should not be considered in the design of assessment situations for students who may experience attentional processing issues.

A second example is choosing items that have a lower level of cognitive difficulty or “depth of knowledge” but that still represent the appropriate grade-level content standards. Aside from the fact that item difficulty can be driven by multiple factors that vary in their relationship to the construct one proposes to measure, it is important that any such efforts grapple in some detail with precisely what is meant by depth of knowledge. This can only be done by taking seriously an analysis of the nature of domain knowledge and competence and then selecting items that are purposely designed to assess some restricted aspects of the domain. Such a choice can be done in a principled way and it necessarily brings with it the implication that the new assessment has a different construct representation than the general education assessment. Regardless of how one chooses to approach the process of changing difficulty, decisions can’t be made on the basis of a superficial level of analysis of the nature of knowledge and skill desired at a particular grade level nor a similarly superficial analysis of how it articulates with some progression of knowledge and skill at both higher and lower grade levels.

A third possible approach is using a “dynamic assessment” procedure with varying levels of prompting and scaffolding. Such a procedure might well be justifiable when assessment is viewed as a test of transfer and the goal is to see how far a student is capable of transferring his or her knowledge. As was noted above, one must still engage deeply with an analysis of what defines competence and acceptable achievement at the particular grade level. On that basis one can determine how a dynamic assessment or scaffolding process can be used to provide an estimate of the intended construct at the desired level of attainment.

One way to view the preceding is as a plea for states, and the field in general, to move cautiously in the development of modified achievement standards and in the implementation of assessments designed to provide valid measures of those standards. Clearly, there is much

that we still need to know. A “design-based” research strategy (e.g., Kelly, Lesh, & Baek, 2008), in which there are serious, small-scale attempts to design and validate such assessments through iterative cycles of empirical testing, redesign, and refinement, may be the most appropriate and cost-effective model for the field at this point in time. Such an approach has multiple advantages including avoiding considerable investment in an approach to assessment that may have limited validity and utility. One critical goal of the pursuit of modified achievement standards should be to provide resources and information that allow educators to engage in a better integration of curriculum, instruction, and assessment for precisely those students whose achievement is characteristically well below desired levels.

References

- Abedi, J., Hofstetter, & Baker, E. (2001). *NAEP math performance and test accommodations: Interactions with student language background* (CSE Technical Report 536). Los Angeles: National Center for Research on Evaluation, Standards, and Student Testing (CRESST), University of California, Los Angeles.
- Adams, J., Treiman, R., & Pressley, M. (1998). Reading, writing, and literacy. In I.E. Sigel and K.A. Renninger, (Eds.), *Handbook of Child Psychology, Fifth Edition, Vol. 4: Child Psychology in Practice* (pp. 275-355). New York: Wiley.
- American Association for the Advancement of Science. (2001). *Atlas of science literacy*. Washington, DC: Author.
- American Educational Research Association, American Psychological Association, & National Council of Measurement in Education (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
- Anderson, J.R. (Ed.) (1981). *Cognitive skills and their acquisition*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Ashcraft, M.H. (1982). The development of mental arithmetic: A chronometric approach. *Developmental Review, 2*, 213-236.
- Ashcraft, M.H. (1983). *Simulating network retrieval of arithmetic facts* (Report No. 1983/10). Pittsburgh: University of Pittsburgh, Learning Research and Development Center.
- Ashcraft, M.H. (1985). *Children's knowledge of simple arithmetic: A developmental model and simulation*. Unpublished manuscript, Cleveland State University.
- Ashcraft, M.H. (1987). Children's knowledge of simple arithmetic: A developmental simulation. In J. Bisanz, C. Brainerd, & R. Kail (Eds.), *Formal methods in developmental psychology* (pp. 302-338). New York: Springer-Verlag.
- Ashcraft, M.H., & Battaglia, J. (1978). Cognitive arithmetic: Evidence for retrieval and decision processes in mental addition. *Journal of Experimental Psychology: Human Learning and Memory, 4*, 527-538.
- Ashcraft, M.H., & Fierman, B.A., & Bartolotta, R. (1984). The production and verification tasks in mental addition: An empirical comparison. *Developmental Review, 4*, 157-170.
- Ashcraft, M.H., & Stazyk, E.H. (1981). Mental addition: A test of three verification models. *Memory & Cognition, 9*, 185-196.
- Baddeley, A. (1986). *Working memory*. Oxford: Oxford University Press.
- Baroody, A.J., Bajwa, N.P., & Eiland, M. (2009). Why can't Johnny remember the basic facts? *Developmental Disabilities Research Reviews, 15*(1), 69 – 79.
- Baxter, G. P., & Glaser, R. (1998). Investigating the cognitive complexity of science assessments. *Educational Measurement: Research and Practice, 17*(3), 37-45.

- Beck, I.L., & McKeown, M.G. (2001). Text talk: Capturing the benefits of read-aloud experiences for young children. *The Reading Teacher*, 55(1), 10-20.
- Beck, I.L., McKeown, M.G., Hamilton, R.L., & Kucan, L. (1997). *Questioning the author: An approach for enhancing student engagement with text*. Newark, DE: International Reading Association.
- Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education*, 5(1), 7-73.
- Brown, J.S., & Burton, R.R. (1978). Diagnostic models for procedural bugs in mathematics. *Cognitive Science*, 4, 379-426.
- Brown, J.S., & Van Lehn, K. (1980). Repair theory: A generative theory of bugs in procedural skills. *Cognitive Science*, 4, 379-426.
- Burton, R.B. (1981). DEBUGGY: Diagnosis of errors in basic mathematical skills. In D.H. Sleeman & J.S. Brown (Eds.), *Intelligent tutoring systems*. London: Academic Press.
- Carpenter, T.P. (1985). Learning to add and subtract: An exercise in problem solving. In E.A. Silver (Ed.), *Teaching and learning mathematical problem solving: Multiple research perspectives*. (pp. 17-40). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Carpenter, T.P., Moser, J.M., & Romberg, T.A. (Eds.). (1982). *Addition and subtraction: A cognitive perspective*. Hillsdale, N.J.: Lawrence Erlbaum Associates.
- Carpenter, T., Fennema, E., & Franke, M. (1996). Cognitively guided instruction: A knowledge base for reform in primary mathematics instruction. *Elementary School Journal*, 97(1), 3-20.
- Chi, M.T.H., Feltovich, P. J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science*, 5, 121-152.
- Cobb, P., Wood, T., Yackel, E., Nicholls, J., Wheatley, G., Trigatti, B., & Perlwitz, M. (1991). Assessment of a problem-centered second-grade mathematics project. *Journal for Research in Mathematics Education*, 22(1), 3-29.
- Cole, M., Gay, J., & Glick, J. (1968). Some experimental studies of Kpelle quantitative behavior. *Psychonomic Monograph Supplements*, 2(10), 173-190.
- Connor, F. P. (1983). Improving school instruction for learning disabled children: The Teachers College Institute. *Exceptional Education Quarterly*, 4, 23-44.
- Corcoran, T.B., Mosher, F.A. & Rogat, A. (2009). *Learning progressions in science: An evidence-based approach to reform*. Consortium for Policy Research in Education, Center on Continuous Instructional Improvement, Teachers College, Columbia University, New York.
- Dickinson, D.K., & Sprague, K. E. (2001) The nature and impact of early childhood care environments on the language and early literacy development of children from low-

- income families. In S.B. Neuman and D.K. Dickinson (Eds.), *Handbook of early literacy research* (pp. 263-280). New York: Guilford Press.
- Duncan, R.G., & Hmelo-Silver, C. (2009). Learning progressions: Aligning curriculum, instruction, and assessment. *Journal for Research in Science Teaching*, in press.
- Dweck, C., & Legget, E. (1988). A social-cognitive approach to motivation and personality. *Psychological Review*, 95, 256-273.
- Ehri, L.C. (1998) Grapheme-phoneme knowledge is essential for learning to read words in English. In J.L. Metsala and L.C. Ehri, (Eds.), *Word Recognition in Beginning Literacy* (pp. 3-40). Mahwah, NJ: Erlbaum.
- Fay, A. & Klahr, D. (1996). Knowing about guessing and guessing about knowing: Preschoolers' understanding of indeterminacy. *Child Development*, 67, 689-716.
- Fletcher, J.M., Foorman, B. R., Boudousquie, A., Barnes, M., Schatschneider, C., & Francis, D.J. (2002). Assessment of reading and learning disabilities: A research-based, treatment-oriented approach. *Journal of School Psychology*, 40, 27-63.
- Foorman, B.R., & Torgesen, J.K. (2001). Critical elements of classroom and small-group instruction promote reading success in all children. *Disabilities Research and Practice*, 16, 202-211.
- Fredriksen, N., Lesgold, A., Glaser., & Shafton, M. (Eds.) (1988). *Diagnostic monitoring of skill and knowledge*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Friend, J., & Burton, R. (1981). *A teacher's manual of subtraction bugs* (working paper). Palo Alto, CA: Xerox Palo Alto Research Center.
- Fuchs, L.S., Compton, D.L., Fuchs, D., Paulsen, K., Bryant, J.D., Hamlett, C.L. (2005). The prevention, identification, and cognitive determinants of math difficulty. *Journal of Educational Psychology*. 97(3), 493-513.
- Fuson, K.C. (1982). An analysis of the counting-on procedure in addition. In T.P. Carpenter, J.M. Moser, & T.A. Romberg (Eds.) *Addition and subtraction: A cognitive perspective* (pp. 67-81). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Fuson, K.C. (1984). More complexities in subtraction. *Journal for Research in Mathematics Education*, 15, 214-225.
- Fuson, K.C., & Hall, J.W. (1983). The acquisition of early number word meanings: A conceptual analysis and review. In H.P. Ginsburg, (Ed.), *The development of mathematical thinking* (pp. 49-107). New York: Academic Press.
- Geary, D. (1995). Reflections of evolution and culture in children's cognition: Implications for mathematical development and instruction. *American Psychologist*, 50(1), 24-37.
- Gelman, R. (1990). First principles organize attention to and learning about relevant data: Number and the animate-inanimate distinction as examples. *Cognitive Science*, 14, 79-106.

- Gelman, R., & Gallistel, C.R. (1978). *The child's understanding of number*. Cambridge, MA: Harvard University Press.
- Ginsburg, H.P., Greenes, C., & Balfanz, R. (2003). *Big math for little kids*. Parsippany, NJ: Dale Seymour Publications.
- Glaser, R. (1963). Instructional technology and the measurement of learning outcomes: Some questions. *American Psychologist*, 18, 519-521.
- Glaser, R. (1992). Expert knowledge and processes of thinking. In D. F. Halpern (Ed.), *Enhancing thinking skills in the sciences and mathematics* (pp. 63-75). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Goldman, S.R., Mertz, D., & Pellegrino, J.W. (1988). Extended practice of basic addition facts: Strategy changes in learning disabled students. *Cognition & Instruction*, 5(3), 223-265.
- Graesser, A.C., Millis, K.K. & Zwaan, R.A. (1997). Discourse comprehension. *Annual Review of Psychology*, 48, 163-189.
- Graves, M.F. & Slater, W.H. (1987). *Development of reading vocabularies in rural disadvantaged students, intercity disadvantaged students and middle class suburban students*. Paper presented at the conference of the American Educational Research Association, Washington, DC, April.
- Greeno, J.G. (1978). A study of problem solving. In R. Glaser (Ed.) *Advances in instructional psychology* (Vol. 1) (pp.13-75). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Greeno, J. G., Pearson, P. D., & Schoenfeld, A. H. (1996). *Implications for NAEP of research on learning and cognition. Report of a study commissioned by the National Academy of Education*. Panel on the NAEP Trial State Assessment, Conducted by the Institute for Research on Learning. Stanford, CA: National Academy of Education.
- Griffin, S., & Case, R., (1997). Re-thinking the primary school math curriculum: An approach based on cognitive science. *Issues in Education*, 3(1), 1-49.
- Griffin, S. A., Case, R., & Siegler, R. S. (1994). Rightstart: Providing the central conceptual prerequisites for first formal learning of arithmetic to students at risk for school failure. In K. McGilly (Ed.), *Classroom lessons: Integrating cognitive theory and classroom practice* (pp. 1-50). Cambridge, MA: MIT Press/Bradford Books.
- Groen, G.J., & Parkman, J.M. (1972). A chronometric analysis of simple addition. *Psychological Review*, 79, 329-343.
- Hamann, M.S., & Ashcraft, M.H. (1985). Simple and complex mental addition across development. *Journal of Experimental Child Psychology*, 40, 49-72.
- Hamilton, L. S., Nussbaum, E. M., & Snow, R. E. (1997). Interview procedures for validating science assessments. *Applied Measurement in Education*, 10(2), 181-200.
- Hart, B., & Risley, T.R. (1995). *Meaningful differences in the everyday experience of young American children*. Baltimore: Paul H. Brookes Publishing Co.

- Hatano, G. (1990). The nature of everyday science: A brief introduction. *British Journal of Developmental Psychology*, 8, 245-250.
- Hickey, D., & Pellegrino, J.W. (2005). Theory, level, and function: Three dimensions for understanding transfer and student assessment. In J.P. Mestre (Ed.). *Transfer of learning from a modern multidisciplinary perspective* (pp. 251-293). Greenwich, CO: Information Age Publishing.
- Houlihan, D.M., & Ginsburg, H.G. (1981). The addition methods of first- and second-grade children. *Journal for Research in Mathematics Education*, 12, 95-106.
- Huttenlocher, J. (1998). Language input and language growth. *Preventive Medicine*, 27, 195-199.
- Kaiser, M.K., Proffitt, D.R., and McCloskey, M. (1985). The development of beliefs about falling objects. *Perception & Psychophysics*, 38(6), 533-539
- Kalchman, M., Moss, J., & Case, R. (2001). Psychological models for development of mathematical understanding: Rational numbers and functions. In S. Carver and D. Klahr, (Eds.), *Cognition and Instruction: Twenty-Five Years of Progress* (pp. 1-38). Mahwah, NJ: Erlbaum.
- Kelly, A.E., Lesh, R.A., & Baek, J.Y. (Eds.) (2008). *Handbook of design research methods in education: Innovations in science, technology, engineering, and mathematics learning and teaching*. New York: Routledge.
- Kintsch, W. (1998). *Paradigms of comprehension*. Oxford, England: Oxford University Press.
- Lampert, M. (1986). Knowing, doing, and teaching multiplication. *Cognition and Instruction*, 3, 305-342.
- Lane, L, Wang, N, & Magone, M. (1996). Gender-related differential item functioning on a middle-school mathematics performance assessment. *Educational Measurement: Issues and Practice*, 15(4), 21-28.
- Lave, J. (1988). *Cognition in practice*. Cambridge, England: Cambridge University Press.
- Lemke, J. J. (2000). Across the scale of time: Artifacts, activities, and meaning in ecosocial systems. *Mind, Culture, and Activity*, 7 (4), 273-290.
- Lesgold, A., & Perfetti, C.A. (Eds.). (1981). *Interactive process in reading*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Lesh, R., & Landau, M. (Eds.). (1983). *Acquisition of mathematics concepts and processes*. New York: Academic Press.
- Magone, M. E., Cai, J., Silver, E. A., & Wang, N. (1994). Validating the cognitive complexity and content quality of a mathematics performance assessment. *International Journal of Educational Research*, 21(3), 317-340.
- Mannes, S.M., & Kintsch, W. (1987). Knowledge organization and text organization. *Cognition*

and Instruction, 4, 91-115.

Massey, C.M., & Gelman, R. (1988). Preschoolers decide whether pictured unfamiliar objects can move themselves. *Developmental Psychology*, 24, 307-317.

Messick, S. (1993). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed.) (pp. 13-103). Phoenix, AZ: Oryx Press.

Miller, S. P. (1997). Educational aspects of mathematics disabilities. *Journal of Learning Disabilities*, 30(1), 47-56.

Miyake, A., Just, M.A., & Carpenter, P.A. (1994). Working memory constraints on the resolution of lexical ambiguity: Maintaining multiple interpretations in neutral contexts. *Journal of Memory and Language*, 33(2), 175-202.

Morales, R., Shute, V., & Pellegrino, J.W. (1985). Developmental differences in understanding and solving simple mathematics word problems. *Cognition and Instruction*, 2, 41-57.

National Council of Teachers of Mathematics (1995). *Assessment standards for school mathematics*. Reston, VA: Author.

National Council of Teachers of Mathematics (2000). *Principles and standards for school mathematics*. Reston, VA: Author.

National Institute of Child Health and Human Development (2000). *Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction*. Report of the National Reading Panel. NIH Publication No. 00-4769. Washington, DC: U.S. Department of Education.

National Research Council (1997). *Educating one and all: Students with disabilities and standards-based reform*. Committee on Goals 2000 and the Inclusion of Students with Disabilities. L. M. McDonnell, M. J. McLaughlin, and P. Morison, Eds. Washington, DC: National Academy Press.

National Research Council (1998). *Preventing reading difficulties in young children*. Committee on the Prevention of Reading Difficulties in Young Children. C.E. Snow, M. Burns, and M. Griffin, Eds. Washington, DC: National Academy Press.

National Research Council (1999a). *Grading the nation's report card: Evaluating NAEP and transforming the assessment of educational progress*. Committee on the Evaluation of National and State Assessments of Educational Progress. J. W. Pellegrino, L. R. Jones, and K. J. Mitchell, Eds. Washington, DC: National Academy Press.

National Research Council (1999b). *How people learn: Brain, mind, experience, and school*. Committee on Developments in the Science of Learning. J. D. Bransford, A. L. Brown, and R. R. Cocking, Eds. Washington, DC: National Academy Press.

National Research Council (1999c). *How people learn: Bridging research and practice*. Committee on Learning Research and Educational Practice. M. S. Donovan, J. D. Bransford, and J. W. Pellegrino, Eds. Washington, DC: National Academy Press.

- National Research Council (1999d). *High stakes: Testing for tracking, promotion, and graduation*. Committee on Appropriate Test Use. J. P. Heubert and R. M. Hauser, Eds. Washington, DC: National Academy Press.
- National Research Council (2000). *How people learn: Brain, mind, experience, and school* (Expanded Edition). Committee on Developments in the Science of Learning and Committee on Learning Research and Educational Practice. J. D. Bransford, A. L. Brown, R. R. Cocking, M. S. Donovan, & J. W. Pellegrino, Eds. Washington, DC: National Academy Press.
- National Research Council (2001a). *Knowing what students know: The science and design of educational assessment*. Committee on the Foundations of Assessment. J. Pellegrino, N. Chudowsky, and R. Glaser, Eds. Washington, DC: National Academy Press.
- National Research Council (2001b). *Adding It Up: Helping children learn mathematics*. Mathematics Learning Study Committee. J. Kilpatrick, J. Swafford, and B. Findell, Eds. Washington, DC: National Academy Press.
- National Research Council (2003). *Assessment in support of learning and instruction: Bridging the gap between large-scale and classroom assessment*. Committee on Assessment in Support of Learning and Instruction. Washington, DC: National Academies Press.
- National Research Council (2005). *Systems for state science assessments*. Committee on Testing Design for K-12 Science Achievement. M.R. Wilson, & M.W. Bertenthal, Eds. Washington DC: The National Academy Press.
- National Research Council (2007). *Taking science to school: Learning and teaching science in grade K-8*. Committee on Science Learning, Kindergarten through eighth grade. R.A. Duschl, H.A., Schweingruber, & A.W. Shouse, Eds. Washington DC: The National Academy Press.
- Newcombe, N.S., & Huttenlocher, J. E. (2000). *Making space*. Cambridge, MA: MIT Press.
- Norman, D.A. (1981). Categorization of action slips. *Psychological Review*, 88, 1-15.
- O'Connor, R.E., & Jenkins, J.R. (1999). The prediction of reading disabilities in kindergarten and first grade. *Scientific Studies of Reading*, 3, 159-197.
- Olson, D.R. (1977). From utterance to text: The bias of language in speech and writing. *Harvard Educational Review*, 47, 257-281.
- Pearson, P.D., & Hamm, D.N. (2002). *The assessment of reading comprehension: A review of practice—past, present, and future*. Santa Monica, CA: RAND Corporation.
- Pellegrino, J.W. (1988). Mental models and mental tests. In H. Wainer & H.I. Braun (Eds.), *Test validity* (pp. 49-60). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Pellegrino, J. W., Baxter, G. P., and Glaser, R. (1999). Addressing the "two disciplines" problem: Linking theories of cognition and learning with assessment and instructional practice. In A. Iran-Nejad and P. D. Pearson (Eds.), *Review of research in education* (vol. 24) (pp. 307-353). Washington, DC: American Educational Research Association.

- Pellegrino, J.W. & Hickey, D. (2006). Educational assessment: Towards better alignment between theory and practice. In L. Verschaffel, F. Dochy, M. Boekaerts, & S. Vosniadou (Eds.), *Instructional psychology: Past, present and future trends. Sixteen essays in honour of Erik De Corte* (Advances in Learning and Instruction Series) (pp 169-189). Oxford: Elsevier.
- Perfetti, C.A. 1992 The representation problem in reading acquisition. In P.B. Gough, L.C. Ehri, and R. Treiman (Eds.), *Reading Acquisition* (pp. 145-174). Hillsdale, NJ: Lawrence Erlbaum.
- Pressley, M., Johnson, C.J., Symons, S., McGoldrick, J.A., & Kurita, J.A. (1989). Strategies that improve children's memory and comprehension of text. *Elementary School Journal*, 90(1), 3-32.
- RAND (2002a). *Toward an R and D program in reading comprehension*. RAND Reading Study Group, Catherine Snow, Chair. Santa Monica, CA: RAND.
- RAND (2002b). *Mathematical proficiency for all students: Toward a strategic research and development program in mathematics education*. RAND Mathematics Study Panel, Deborah Loewenberg Ball, Chair. DRU-2773-OERI. Santa Monica, CA: RAND.
- Resnick, L.B. (1982). *Syntax and semantics in learning to subtract*. Report # LRDC-1982/8, Learning Research and Development Center, University of Pittsburgh, Pittsburgh, PA.
- Resnick, L.B. (1984). Beyond error analysis: The role of understanding in elementary school arithmetic. In H.N. Creek (Ed.), *Diagnostic and prescriptive mathematics: Issues, ideas, and insight* (pp. 181-205). Kent, OH: Research Council for Diagnostic and Prescriptive Mathematics.
- Riley, M., Greeno, J., & Heller, J. (1983). Development of children's problem-solving ability in arithmetic. In H. Ginsburg (Ed.), *The development of mathematical thinking* (pp. 153–196). New York: Academic Press.
- Rogoff, B. (1990). *Apprenticeship in thinking: Cognitive development in social context*. New York: Oxford University Press.
- Rosenbloom, P., & Newell, A. (1987). Learning by chunking: A production system model of practice. In D. Klahr & P. Langley (Eds.), *Production system models of learning and development* (pp. 221-286). Cambridge, MA: MIT Press.
- Ruiz-Primo, M.A., Shavelson, R.J., Hamilton, L., & Klein, S. (2002). On the evaluation of systemic science education reform: Searching for instructional sensitivity. *Journal of Research in Science Teaching*, 39, 369-393.
- Russell, R.L., & Ginsburg, H.P. (1984). Cognitive analysis of children's mathematics difficulties. *Cognition and Instruction*, 1, 217-244.
- Sadler, R. (1989). Formative assessment and the design of instructional systems. *Instructional Science*, 18, 119-144.
- Schoenfeld, A.H. (1985). *Mathematical problem solving*. Orlando, FL: Academic Press.

- Shepard, L. A. (2000). The role of assessment in a learning culture. *Educational Researcher*, 29 (7), 4-14.
- Siegler, R. S. (1998). *Children's thinking* (3rd ed.). Upper Saddle River, NJ: Prentice Hall.
- Siegler, R.S., & Shrager, J. (1984). Strategy choices in addition and subtraction: How do children know what to do? In C. Sophian (Ed.), *Origins of cognitive skills* (pp. 229-293). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Simon, H.A. (1980). Problem solving and education. In D.T. Tuma and F. Reif (Eds.), *Problem solving and education: Issues in teaching and research* (pp. 81-96). Hillsdale, NJ: Erlbaum.
- Smith, C, Wiser, M., Anderson, C.W, & Krajcik, J. (2006). Implications of children's learning for assessment: A proposed learning progression for matter and the atomic molecular theory. *Measurement*, 14(1&2), 1-98.
- Starkey, P., & Cooper, R.G. (1980). Perception of numbers by human infants. *Science*, 210, 1033-1035.
- Starkey, P., & Gelman, R. (1982). The development of addition and subtraction abilities prior to formal schooling in arithmetic. In T.P. Carpenter, J.M. Moser, & T.A. Romberg (Eds.), *Addition and subtraction: A cognitive perspective* (pp. 99-116). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Steffe, L., Thompson, P., & Richards, J. (1982). Children's counting in arithmetical problem solving. In T.P. Carpenter, J.M. Moser, & T. Romberg (Eds.). *Addition and subtraction: A cognitive perspective*. Hillsdale, N.J.: Lawrence Erlbaum Associates.
- Steffe, L.P., von Glaserfeld, E., Richards, J., & Cobb, P. (1983). *Children's counting types: Philosophy, theory, and application*. New York: Praeger Scientific.
- Stiggins, R. J. (1997). *Student-centered classroom assessment*. Upper Saddle River, NJ: Prentice-Hall.
- Svenson, O. (1975). Analysis of time required by children for simple additions. *Acta Psychologica*, 39, 289-302.
- Svenson, O., & Broquist, S. (1975). Strategies for solving simple addition problems: A comparison of normal and subnormal children. *Scandinavian Journal of Psychology*, 16, 143-151.
- Svenson, O., & Hedenborg, M.L. (1979). Strategies used by children when solving simple subtractions. *Acta Psychologica*, 43, 477-489.
- Swanson, H.L., & Jerman, O. (2006). Math disabilities: A selective meta-analysis of the literature. *Review of Educational Research*, 76(2), 249-274.
- Thorndike, E.L. (1931). *Human learning*. New York: Century.

- Torgesen, J.K. (2002). The prevention of reading difficulties. *Journal of School Psychology, 40*(1), 7-26.
- Unsworth, N. & Engle, R.W. (2007). The nature of individual differences in working memory capacity: Active maintenance in primary memory and controlled search from secondary memory. *Psychological Review, 114*(1), 104-132.
- Van Lehn, K. (1983). Bugs are not enough: Empirical studies of bugs, impasses and repairs in procedural skills. *Journal of Mathematical Behavior, 3*, 3-71.
- Van Lehn, K. (1990). *Mind bugs: the origins of procedural misconceptions*. Cambridge, MA: MIT Press.
- Wagner, R.K., Torgesen, J.K., Laughon, N.P., Simmons, K., & Rashotte, C.A. (1993). Development of young readers' phonological processing abilities. *Journal of Educational Psychology, 85*, 83-103.
- Wagner, R.K., Torgesen, J.K., Rashotte, C.A., Hecht, S.A., Barker, T.A., Burgess, S.R., Donahue, J., & Garon, T. (1997). Changing relations between phonological processing abilities and word-level reading as children develop from beginning to skilled readers: A 5-year longitudinal study. *Developmental Psychology, 33*, 468-479.
- Webb, N. L. (1997). *Criteria for alignment of expectations and assessments in mathematics and science education*. National Institute for Science Education and Council of Chief State School Officers Research Monograph No. 6. Washington, DC: Council of Chief State School Officers.
- Whitehurst, G.J., & Lonigan, C.J. (2001). Emergent literacy: Development from prereaders to readers. In S.B. Neuman and D.K. Dickinson, eds. *Handbook of Early Literacy Research* (pp. 11-29). New York: Guilford Press.
- Wiggins, G. (1998) *Educative assessment: Designing assessments to inform and improve student performance*. San Francisco: Jossey-Bass.
- William, D. (2007). Keeping learning on track: formative assessment and the regulation of learning. In F. K. Lester Jr. (Ed.), *Second handbook of mathematics teaching and learning* (pp. 1053-1098). Greenwich, CT: Information Age Publishing.
- Winkelman, H.J., & Schmidt, J. (1974). Associative confusions in mental arithmetic. *Journal of Experimental Psychology, 102*, 734-736.
- Woods, S.S., Resnick, L.B., & Groen, G.J. (1975). An experimental test of five models for subtraction. *Journal of Educational Psychology, 67*, 17-21.
- Young, R.M., & O'Shea, T. (1981). Errors in children's subtraction. *Cognitive Science, 5*, 153-177.
- Zwick, R., & Ercikan, K. (1989). Analysis of differential item functioning in the NAEP history assessment. *Journal of Educational Measurement, 26*(1), 55-66.