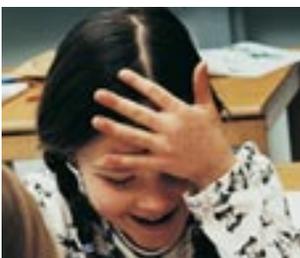




Alternate assessments for special education students in the Southwest Region states



Institute of Education Sciences
U.S. Department of Education



Alternate assessments for special education students in the Southwest Region states

February 2008

Prepared by

Stanley Rabinowitz, Ph.D.
WestEd

Edynn Sato, Ph.D.
WestEd

Betsy J. Case, Ph.D.
WestEd

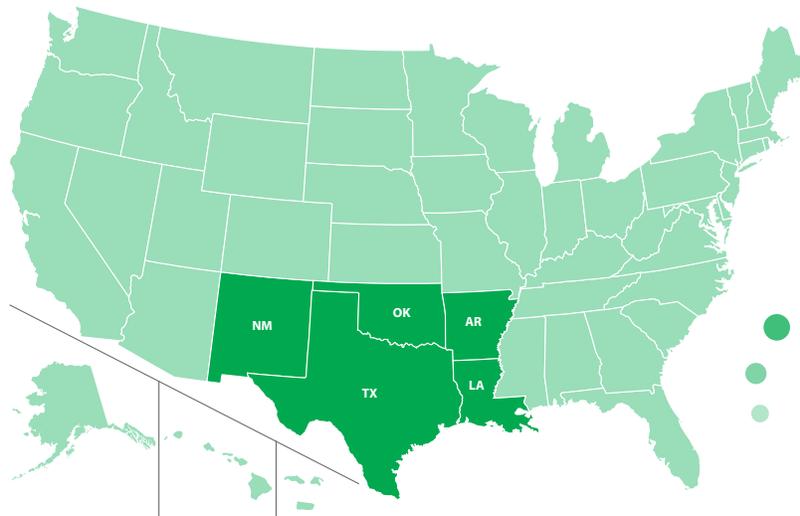
Debra Benitez, Ph.D.
WestEd

Kevin Jordan
WestEd



Institute of Education Sciences

U.S. Department of Education



Issues & Answers is an ongoing series of reports from short-term Fast Response Projects conducted by the regional educational laboratories on current education issues of importance at local, state, and regional levels. Fast Response Project topics change to reflect new issues, as identified through lab outreach and requests for assistance from policymakers and educators at state and local levels and from communities, businesses, parents, families, and youth. All Issues & Answers reports meet Institute of Education Sciences standards for scientifically valid research.

February 2008

This report was prepared for the Institute of Education Sciences (IES) under Contract ED-06-CO-0017 by Regional Educational Laboratory Southwest administered by Edvance Research. The content of the publication does not necessarily reflect the views or policies of IES or the U.S. Department of Education nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government.

This report is in the public domain. While permission to reprint this publication is not necessary, it should be cited as:

Rabinowitz, S., Sato, E., Case, B. J., Benitez, D., & Jordan K. (2008). *Alternate assessments for special education students in the Southwest Region states* (Issues & Answers Report, REL 2008–No. 044). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Southwest. Retrieved from <http://ies.ed.gov/ncee/edlabs>

This report is available on the regional educational laboratory web site at <http://ies.ed.gov/ncee/edlabs>.

Alternate assessments for special education students in the Southwest Region states

In 2003 the U.S. Department of Education issued regulations allowing states to develop alternate standards and assessments for students with the most significant cognitive disabilities. This study reviews and summarizes alternate assessment policies and practices—and their implementation and impact—for the most significantly cognitively disabled students, across the five states in the Southwest Region.

The No Child Left Behind Act of 2001 was the first federal act to require including all students in state and district accountability systems. In 2003 the U.S. Department of Education issued regulations allowing states to develop alternate assessment standards for students with the most significant cognitive disabilities—and to include some results from these assessments in annual school, district, and state accountability formulas as long as the number of such inclusions did not exceed 1 percent of the combined population of students taking general and alternate assessments statewide (U.S. Department of Education, 2003b). The Individuals with Disabilities Education Act of 2004 and U.S. Department of Education (2006b) regulations issued in August 2006 further clarified the requirements for assessing students with the most significant cognitive disabilities. One important

change was that states now needed to link alternate assessment standards to general education standards.

Many states are struggling to identify alternate content standards, to find curricula that address these standards while meeting student needs, to locate teachers who can implement the curricula, and to ensure that alternate standards are demonstrably linked to general education standards in accordance with expectations set by the No Child Left Behind Act. Most also face great challenges developing and implementing reliable and valid alternate assessments that can be implemented efficiently and comparably across the state.

The survey and interviews conducted for this study suggest that the Southwest Region states have been tracking changes in their curricular and assessment focus from functional to academic content. State representatives believe that changes in policies and practices have improved each state's approach and emphasis, though they admit a need for more rigorous analysis of these relationships.

For 2007/08 four of the five Southwest Region states will have instituted alternate portfolio assessments (based on work samples) or performance assessments (based on exemplars of proficient performance) for students with

significant cognitive disabilities. Louisiana, New Mexico, and Texas have transitioned dramatically. Louisiana has changed from a checklist to a multiple-choice measure. New Mexico has switched from a checklist based wholly on functional achievement to performance-based tasks that are linked to alternate achievement standards. Texas has created alternate achievement standards based on state general content standards and is transitioning from local choice in testing to a uniform state-developed portfolio system with a checklist.

Given the range of the student cognitive and physical disabilities that definitions cover, a one-size alternate assessment will not fit all.

The wide range of skills and tasks targeted by alternate assessments creates challenges for comparability and for determinations of across-the-board technical adequacy. Much work is needed to establish alternate assessments that reflect adequate psychometric properties, instructional relevance, validity, reliability, and usability. The Southwest Region states share many of the same needs. Each, however, has its own unique histories, values, populations, approaches, resources, and constraints—which must be taken into account in any attempt to address a particular state’s requirements or to study them further.

February 2008

TABLE OF CONTENTS

Overview	1
Why this study?	4
Challenges to designing and implementing alternate assessments	7
Deciding who should participate	7
Deciding what content alternate assessments should measure	9
Defining technical adequacy for alternate assessments	12
Creating reliable and valid alternate assessments	15
Defining proficient performance (setting standards)	18
A review of alternate assessments across the Southwest Region states	19
What do alternate assessments across the Southwest Region states look like?	19
What training or professional development is provided for teachers on alternate assessments?	23
How are results collected and used at the state, district, school, and student levels?	26
To what extent do state alternate assessments capture the same or similar skills as state tests designed for the general student population?	28
What technical issues are states facing in developing and implementing reliable and valid alternate assessments?	30
Far-reaching questions, existing approaches, further research	30
Appendix A Glossary	33
Appendix B Southwest Region state demographic information	37
Appendix C Side-by-side comparison of Southwest Region states' alternate assessment policies and practices	38
Appendix D Study methods and limitations	54
Appendix E Alternate assessment interview for the Southwest Region states	56
References	58
Boxes	
1	Definitions of key terms 2
2	Study methods 3
Tables	
1	Alternate assessment approaches commonly used by states (number and percent of states), 1999–2005 16
2	Summary of state demographic statistics for the Southwest Region states 20
3	Students taking alternate assessments in Southwest Region states, by grade, 2005/06 22
4	Southwest Region state alternate assessment approaches for 2007/08 23

5	Students scoring at proficient or above on alternate assessments in math and reading by grade and state, 2005/06 (percent)	27
6	Students scoring proficient or above for general and alternate assessments in math and reading in grades 4 and 5 in the Southwest Region states, 2005/06	28
B1	State demographic statistics	37
C1	State assessments by grade and subject	38
C2	Southwest Region state laws, regulations, rules, and administrative code for alternate assessment	40
C3	Southwest Region state definition of significant cognitive disability	42
C4	Southwest Region state education agency guidelines for participation in alternate assessment	44
C5	Assessment approaches and tasks	46
C6	Southwest Region state-provided training on administration and use of results	48
C7	Alignment and standard setting in the Southwest Region states	50
C8	Review procedures and recent changes	52

In 2003 the U.S. Department of Education issued regulations allowing states to develop alternate standards and assessments for students with the most significant cognitive disabilities. This study reviews and summarizes alternate assessment policies and practices—and their implementation and impact—for the most significantly cognitively disabled students, across the five states in the Southwest Region.

OVERVIEW

This study reviews policies and practices related to alternate assessment for the most significantly cognitively disabled students (see box 1 and appendix A for definitions of key terms). It examines challenges that states across the nation encounter when implementing these policies and practices, and it presents information about alternate assessments in the five states served by the Southwest Regional Educational Laboratory: Arkansas, Louisiana, New Mexico, Oklahoma, and Texas (see appendix B for student demographic information for each state). The study's findings emerge from documents, a survey, and interviews, summarized across the five states and compared with national findings and trends.

Most students with disabilities participate in state and district assessments by taking existing assessments with testing accommodations. But a small percentage of students have disabilities that make their participation in general state and district tests impractical, if not impossible. Such participation is likely to yield inaccurate measures of academic achievement. Alternate assessments are intended for students unable to participate in state and district assessment systems, even with accommodations (Thompson, Johnstone, Thurlow, & Altman, 2005; Thurlow & Case, 2004). On December 9, 2003, the U.S. Department of Education issued regulations that allowed states to use—for accountability purposes—alternate assessments based on alternate achievement standards for students with the most significant cognitive disabilities.

A primary goal of both the No Child Left Behind Act of 2001 and the Individuals with Disabilities Education Act of 2004 was to include more students with significant cognitive disabilities in state and district accountability systems. For 2007/08 all five Southwest Region states will have approached this goal—for example, by moving from checklists or assessments based on functional achievement to portfolio or performance tasks that are linked to alternate standards based on general

BOX 1

Definitions of key terms

1 percent rule. When measuring adequate yearly progress, states and school districts have the flexibility to count the “proficient” and “advanced” scores of students with the most significant cognitive disabilities who take alternate assessments based on alternate achievement standards—as long as the number of scores so counted does not exceed 1 percent of all students in the grades assessed (or about 9 percent of students with disabilities). (U.S. Department of Education, 2003a)

2 percent rule. When measuring adequate yearly progress, states and local education agencies may count the “proficient” and “advanced” scores of certain students who take alternate assessments even though they are not identified as having the most significant cognitive disabilities—as long as the number of scores so counted does not exceed 2 percent of all students in the grades assessed (or about 20 percent of students with disabilities). (U.S. Department of Education, 2007)

Accommodations. A change in the administration of an assessment (setting, scheduling, timing, presentation, response mode) to achieve equity, not advantage, that does not

change the construct to be measured or the meaning of the resulting scores. Accommodations should be identified in the student’s individualized education program or an accommodation plan under Section 504 of the Rehabilitation Act of 1973 and used regularly during instruction and classroom assessment. (Policy to Practice Study Group, 2003)

Adequate yearly progress. A provision of the No Child Left Behind legislation requiring schools, districts, and states to demonstrate on the basis of test scores that students are making academic progress. Each state was required to submit by January 31, 2003, a specific plan for monitoring adequate yearly progress. (Policy to Practice Study Group, 2003)

Alternate assessment. An instrument used to gather information on the standards-based performance and progress of a relatively small population of students who are unable to participate in the general assessment system, such as those whose disabilities preclude their valid and reliable participation in general assessments. (Policy to Practice Study Group, 2003)

Individualized education program. A document that reflects the decisions made by an interdisciplinary team, including the parent and student

when appropriate, and identifies the abilities and disabilities of a disabled student. (Policy to Practice Study Group, 2003)

Peer review. The review of a state’s standards and assessment system by state practitioners and experts to determine whether it meets the requirements of the No Child Left Behind Act.

Reliability. The consistency of the test instrument; the extent to which it is possible to generalize a specific behavior observed at a specific time by a specific person to observations of similar behavior at different times or by different behaviors. (Policy to Practice Study Group, 2003)

Technical adequacy. The extent to which an assessment meets the requirements for validity, reliability, accessibility, objectivity, and consistency with nationally recognized professional and technical standards. Evidence for technical adequacy can include information on administration, scoring, interpretation, and technical data. (U.S. Department of Education, 2007)

Validity. The extent to which a test measures what it was designed to measure. (Policy to Practice Study Group, 2003)

content standards. But peer review letters from the U.S. Department of Education to each Southwest Region state (U.S. Department of Education, 2006a) reveal that although most states have done a good deal of work to implement the federal requirements, few have met the technical challenges related to implementing alternate assessments that are linked to their general education counterparts.

To provide an overview of the technical and support challenges that states face as they build alternate assessments for students with the most significant cognitive disabilities, and to supply a context for the study’s findings, researchers reviewed studies of the development, implementation, and validation of the assessments and surveyed or interviewed state department of education staff (see box 2).

BOX 2

Study methods

To investigate the challenges to designing and implementing alternate assessments across the Southwest Region states, the researchers developed six research questions:

1. What challenges are states encountering when implementing new alternate assessment policies and practices?
2. What do alternate assessments across the Southwest Region states look like?
3. What training or professional development is provided for teachers on alternate assessments?

4. How are results collected and used at the state, district, school, and student levels?
5. To what extent do state alternate assessments capture the same or similar skills as state tests designed for the general student population?
6. What technical issues are states facing in developing and implementing reliable and valid alternate assessments?

The study replicated the procedures used by Browder et al. (2005), a mixed-methods approach to research that includes qualitative and quantitative methods. When systematically combined, the methods provide rigorous, methodologically

sound investigations in a range of fields (Creswell, Fetters, & Ivankova, 2004). The quantitative data collection involved review of state materials, surveys, and interviews using descriptive techniques. The qualitative data collection consisted of semistructured surveys and a review of documents.

The researchers organized the data into a matrix to compare state policies, practices, and procedures. They integrated the results of the quantitative method (survey) with that of the qualitative data (review of documents) and systematically analyzed the data and reported the results. (For a fuller account of the study methodology, see appendix D.)

They identified the challenges states are encountering when implementing new alternate assessment policies and practices. They described what alternate assessments across the Southwest Region states look like and what training or professional development is provided for educators on alternate assessments. They looked at how results were being collected and used at the state, district, school, and student levels and the extent to which states' alternate assessments capture the same or similar skills as state tests designed for the general student population. They also considered what technical issues states face in developing and implementing reliable and valid alternate assessments. They summarized current policies and practices in the five Southwest Region states and connected these practices to what is known nationally.

All the Southwest Region states report including increasing numbers of students with significant cognitive disabilities in state assessments and implementing changes in special education curricula and instruction (see appendix C). Arkansas and Oklahoma, which have achieved full approval

status according to the federal peer review (U.S. Department of Education, 2006a), report that they are continuing to improve programming and instruction by refining content achievement standards to reflect a blend of functional and academic skills linked to content standards.

Louisiana has changed from a checklist to a performance-based assessment and will be changing to a more traditional testing approach. New Mexico has switched from a checklist based wholly on functional achievement to performance-based tasks that are linked with alternate achievement standards. Texas has created alternate achievement standards based on state general content standards and is transitioning from locally selected alternate assessments and an optional state-developed alternate assessment to a uniform state-developed portfolio system with a checklist.

A linchpin of the No Child Left Behind Act is the need to attend to what gets reported and used for accountability. This study revealed disparities in the five Southwest Region states' definitions of *significant cognitive disability*, their criteria

for alternate assessment participation, their task selection, their scoring, and their reporting. Representatives of all five states openly discussed the technical challenges involved (see appendix C).

As states conduct new alignment studies and standard setting activities, they are required to demonstrate the alignment of alternate assessment achievement standards with grade-level content standards and alternate assessments—a challenging task (especially when states allow teachers to select the content standards being measured or the tasks for assessment). Because of changes to content standards, achievement standards, and assessment approaches in the past several years, three states lack continual, year-to-year trend data that could be used to measure progress and growth. Those three states are exploring how to conduct consequential validity studies to examine policies' positive and negative effects.

Representatives from Louisiana, New Mexico, and Texas underscored their states' needs for additional support to fully implement federal requirements. States report that they might benefit from technical help as they develop robust strategies for collecting and using results at the state, district, school, and student levels. Despite the challenges, some districts are attempting to support the needs of students with disabilities and the current federal mandates. Researchers found similar needs for further research for all five states.

WHY THIS STUDY?

Alternate assessments for students with the most significant cognitive disabilities are fairly new in most states. Before the federal Individuals with

Disabilities Education Act of 1997 most students with disabilities either were not included in state testing or participated inconsistently (Thurlow, 2004). The 1997 act required states to include students with disabilities in state testing, and it boosted the use of

alternate assessments to assess these students—requiring states to create alternate assessments by the end of 2000. Besides requiring full and fair participation in assessment systems, the 1997 act underscored that accommodations must be provided for students with disabilities and that individualized education program teams (IEP teams; see box 1) must make determinations about participation in alternate assessments according to state guidelines.

The No Child Left Behind Act was the first federal act to require including all students in state and district accountability systems. A primary goal of both it and the Individuals with Disabilities Education Act of 2004 was to include more students with significant cognitive disabilities in state assessment and accountability systems. With the No Child Left Behind Act the movement toward alternate assessments picked up speed and urgency. It mandated not only that students with disabilities be included in state assessment programs, but also that they count equally with other designated subgroups for state and federal accountability. Follow-up regulations created the 1 percent group, or students with the most significant cognitive disabilities who could be assessed based on alternate content standards.

The Individuals with Disabilities Education Act of 1997 did not mandate how states should develop alternate assessment policies or procedures. So, states created different versions of alternate assessments based on their own special education student demographics, knowledge of the populations, and requirements for technical adequacy in assessments (Browder, Spooner, Algozzine, et al., 2003; Thurlow, 2004). Virtually all states have developed alternate achievement standards along with some form of alternate assessment and are attempting to link these to reform instructional practice. But recent peer review letters from the U.S. Department of Education (a formal review and approval process for state standards and assessment systems) reveal that although most of the five Southwest Region states have done a good deal of work to implement federal requirements, few have met the challenges to implementing alternate assessments on a par

With the No Child Left Behind Act the movement toward alternate assessments picked up speed and urgency

with their general education counterparts (U.S. Department of Education, 2006a).

Referring to the assessment of students with significant cognitive disabilities under the No Child Left Behind Act, the U.S. Department of Education (2003b) defines an alternate assessment as “an assessment designed for the small number of students with disabilities who are unable to participate in the regular grade-level state assessment, even with appropriate accommodations.” Nonregulatory guidance also included the following language:

To qualify as an assessment under Title I, an alternate assessment must be aligned with the state’s content standards, must yield results separately in both reading and language arts and mathematics, and must be designed and implemented in a manner that supports use of the results as an indicator of adequate yearly progress. Alternate assessments can measure progress based on alternate achievement standards...and can also measure proficiency based on grade-level achievement standards. Alternate assessments may be needed for students who have a broad variety of disabilities; consequently, a state may employ more than one alternate assessment. When used as part of the state assessment program, alternate assessments must have an explicit structure, guidelines for which students may participate, clearly defined scoring criteria and procedures, and a report format that communicates student performance in terms of the academic achievement standards defined by the state (U.S. Department of Education, 2005, p. 15).

The definition and guidance reinforced the idea that alternate assessments were appropriate and allowable for the full range of the intended special education subpopulation. But they gave states considerable flexibility in developing alternate assessment policies, structures, and formats.

Proponents of including students with significant cognitive disabilities as full participants in state

assessments determined that the change would give them a greater voice in the education system and shift accountability to the principle that “all means all” (Thurlow, 2004). The academic achievement of students with the most significant cognitive disabilities has consistently been underestimated by traditional paper-and-pencil tests because of students’ inability to function in an on-demand environment and respond in the limited manner allowed. States saw alternate assessments built to address the needs of this hard-to-assess population as a valid means to increase access and, consequently, make schools’ adequate yearly progress results more fair.

In addition, given the traditional underemphasis on academic instruction for students with disabilities as opposed to those with functional skills (Kleinert & Kearns, 1999), full inclusion with formal accountability requirements was seen as a way to reform the education of students with significant cognitive disabilities. Advocates felt that information gleaned from test results could help guide instructional programming and curricula (Crawford & Tindal, 2006). The outcomes identified as a consequence of the alternate assessment would become a part of the daily education routine (Ford, Davern, & Schnorr, 2001).

A review of 19 research-based articles related to alternate assessment policies (Browder, Spooner, Algozzine, et al., 2003) summarized the underlying principles of alternate assessments following the 1997 federal legislation. The researchers found four reasons why alternate assessments initially showed promise for state accountability and for students with significant cognitive disabilities. Advocates believed that:

1. Greater consideration would be given to the students because they were participating in alternate assessments.

The academic achievement of students with the most significant cognitive disabilities has been underestimated by traditional paper-and-pencil tests

2. Perceptions of low expectations for the students would shift, affording students more opportunities for success.
3. All students would have access to the general education curriculum and, thus, to the same academic content standards.
4. The quality of instructional programming would increase for students with significant cognitive disabilities—a change with implications for the preparation of instructors.

A research study published in 2005 indicated that since December 9, 2003—when the federal government announced the 1 percent rule (see box 1 and appendix A) as a new provision of No Child Left Behind legislation—22 states, or 44 percent, had changed their alternate assessment policies to use alternate achievement standards (Thompson, Johnstone, Thurlow, & Altman 2005). In particular, the researchers believe that two important changes have occurred under the No Child Left Behind Act:

- The focus of alternate assessments has shifted greatly toward measuring academic skills linked to state content standards.
- Given the attention that alternate assessments now receive because of their inclusion in federal and state formal accountability systems, states have attempted to increase the technical adequacy of their alternate assessments.

States across the country are struggling to address the needs of students with the most significant cognitive disabilities; to validly, reliably, and accurately measure student performance; and to develop alternate assessment systems that yield high-quality data for evaluating school, district, and state performance and improving instruction.

Although the five Southwest Region states have made efforts to include students with significant

cognitive disabilities, they all still face great challenges. The Southwest Region states differ in their definitions of *significant cognitive disability* and in their guidelines for student participation in alternate assessments. And the technical quality of their alternate assessments varies. Peer review letters from the U.S. Department of Education indicate that the five states in the Southwest Region received full approval, approval expected, or approval pending for their alternate assessments (U.S. Department of Education, 2006a). But they needed to provide extensive additional evidence from their 2005/06 alternate assessments in six areas:

1. *Academic achievement standards.* Three states needed to submit science-achievement standards. Two states needed to document cutscores and performance descriptors for other content areas measured.
2. *Full assessment system.* Two states needed to submit alignment-study results and plans for completing alignment studies.
3. *Technical quality.* Three states needed to submit evidence of validity, reliability, and usability. All three needed to prepare and submit technical manuals.
4. *Alignment results.* Three states had to complete studies of the alignment of alternate achievement standards to assessments and show their plans for filling any gaps in coverage. The alignment issue is especially complex for alternate assessments because greater local choice and range of artifacts (portfolio work samples, such as photographs or videotapes of the student performing a task, audiotapes, writing samples, drawings, and tests) in alternate assessments mean that the method for demonstrating alignment between content standards and performance assessments is less developed than for on-demand tests.
5. *Inclusion.* Three states needed to show that all students were included in testing (including groups such as migrant students).

Although the five Southwest Region states have made efforts to include students with significant cognitive disabilities, they all still face great challenges

6. *Reporting.* Three states needed to document and clarify their reporting practices.

This descriptive study is intended to contribute to the ongoing discussion of promising and best practices for the assessment of students with significant cognitive disabilities.

- Creating reliable and valid alternate assessments.
- Defining proficient performance (setting standards).

Each of these challenges is discussed below.

CHALLENGES TO DESIGNING AND IMPLEMENTING ALTERNATE ASSESSMENTS

The No Child Left Behind Act and the Individuals with Disabilities Education Act of 2004 represent a paradigm shift in instructing and assessing students with significant cognitive disabilities. According to one study, these acts have moved special education from “a culture of compliance to a culture of accountability for results” (Manasevit & Maginnis, 2005, p. 51). Yet they have also confronted states with technical and logistical challenges far greater than those created by assessments designed chiefly for general education students. Issues of bias, validity, and reliability (see appendix A)—as well as approaches to training and monitoring—are complicated by the heterogeneity and varying degrees of disability in the targeted population and by the nature of the performance assessments developed to increase access for this student group.

Of the six research questions that guided this study (see appendix D), the first—“What challenges are states encountering when implementing new alternate assessment policies and practices?”—prompted researchers to review the national literature. They found that states across the country face five technical and logistical challenges as they struggle to develop standards-based alternate assessments :

- Deciding who should participate.
- Deciding what content alternate assessments should measure.
- Defining technical adequacy for alternate assessments.

Deciding who should participate

Alternate assessments are reserved for students with the most significant cognitive disabilities, but the phrase *significant cognitive disabilities*—used extensively in the literature and in federal guidelines—has no single authoritative definition. The Individuals with Disabilities Education Act of 2004, a common source of special education terminology, does not define it. And only one study (Almond & Bechard, 2005) has formally addressed the characteristics of students who take alternate assessments.

Alternate assessments are reserved for students with the most significant cognitive disabilities, but this phrase has no single authoritative definition

Many states have defined *significant cognitive disabilities* using general language, such as “unable to participate in the general assessment even with accommodations.” But such definitions, which focus on deficits in functioning, do not say much about students as individuals (McDonnell, Hardman, & McDonnell, 2003). To design an assessment one needs a thorough understanding of the targeted student populations and their relevant characteristics. And the students now being targeted by alternate assessments form a widely heterogeneous group in disability characteristics, capabilities, and education needs (Snell & Brown, 2006).

Although the states have not precisely identified the characteristics of students who should participate in alternate assessments, many states include students with autism, moderate to severe mental retardation, multiple disabilities, and traumatic brain injury—some of the 13 distinct disability categories defined by the Individuals with Disabilities Education Act

of 2004 and its corresponding Code of Federal Regulations (34 CFR, 300.7 and 300.8). Students in several of these federally defined categories are considered to have significant cognitive disabilities. But using these federally defined labels to determine eligibility for participation in alternate assessments can be problematic. First, the labels are operationally defined in different ways across local and state education agencies. Second, states have typically used a checklist, or series of questions, completed by a student's IEP team or committee to guide them through the state's process for determining eligibility. But IEP teams vary in their ability to classify students properly, especially students with multiple disabilities. Finally, decisionmaking templates vary widely across states, ranging from a few general questions about a student's level of functioning to extensive multistep procedures that require considering the student's curriculum and documenting possible testing accommodations. Researchers know little about these frameworks' validity or about whether different state processes yield comparable decisions about participation (Almond & Bechard, 2005; Yovanoff & Tindal, 2007).

The procedures and guidelines used to determine student eligibility must be linked to the curriculum

The 1 percent rule makes it essential that IEP teams correctly identify students for alternate assessment participation. The procedures and guidelines used to determine student eligibility must be linked to the curriculum

(Almond and Bechard, 2005). To ensure such linkage—and to make informed decisions about whether to use an alternate assessment rather than accommodations on a large-scale standardized test—IEP teams must be thoroughly familiar with the format and content of state assessments and with state policies on testing accommodations. But Yovanoff and Tindal (2007) indicate that this is not always the case. Browder, Spooner, Ahlgrim-Dezell, et al. (2003) identify this as an area where training is sorely needed.

The addition of a new student population under the 2 percent rule (students eligible for alternate assessments even though they are not identified

as significantly cognitively disabled; see box 1) has further complicated the identification of students eligible for alternate assessments. States may choose to implement an additional assessment based on modified academic achievement standards. They have the flexibility to determine who may take the modified assessment but must follow the guidelines set forth by the U.S. Department of Education.

According to the nonregulatory guidance, eligible students have disabilities under section 602(3) of the Individuals with Disabilities Education Act and have access to a curriculum based on grade-level standards. In each case objective evidence—including performance on state or other assessments that validly document academic achievement—must show that a disability has precluded the student from achieving grade-level proficiency. The student's IEP team must be reasonably certain that, even if significant growth occurs, the student will not achieve grade-level proficiency within the year covered by the individualized education program. The team must base this determination on the student's progress in response to appropriate instruction (including special education and related services designed to address individual needs) and on multiple valid measures of student progress over time. Finally, the student's individualized education program must include goals based on academic content standards for the grade in which the student is enrolled (U.S. Department of Education, 2007, p. 16).

For adequate yearly progress calculations, states may claim that students meeting these criteria and included in the modified assessments are “proficient” or “advanced” so long as the number of such scores does not exceed 2 percent of all students in the grades assessed (about 20 percent of students with disabilities). Like the 1 percent rule, the 2 percent rule does not limit the number of students who may participate in a modified assessment.

Though little is known about whether states are correctly identifying students with disabilities (mild to moderate or more severe), states are required to develop valid and reliable assessments

for them. The Individuals with Disabilities Education Act of 2004 instructs states to develop “guidelines for participation of students in alternate assessment for those children who cannot participate in state- and district-wide assessment programs” (§300.138, part B). It requires students’ individualized education programs to document the justification for their exclusion from the general large-scale assessment and to describe how they will be assessed using an alternate method. The challenge of deciding who should participate in alternate assessments needs further study, regionally and nationally. Across the five Southwest Region states definitions and parameters for eligibility differ substantially (see tables C2 and C3 in appendix C). The diversity of definitions across the country is even greater.

Deciding what content alternate assessments should measure

The Individuals with Disabilities Education Act of 2004 states that students with disabilities should have access to general education curricula and academic standards. Students with significant disabilities must have instruction and accommodations that promote their progress, no matter how modest, toward meeting state and district academic standards for the larger student population. The emphasis on common standards and curricula is a paradigm change from traditional curricula and inclusion practices:

Although the law still maintains the right of each student with disabilities to an individually referenced curriculum, outcomes linked to the general education program have become the optimal target. It is no longer enough for students with disabilities to be present in a general education classroom (Pugach & Warger, 2001, p. 194).

Policymakers and researchers increasingly agree that alternate assessments are intended to function as one component in a larger accountability system and to measure progress toward general education expectations. A state’s general education academic

standards should form the foundation for alternate assessment, and evidence of this link should routinely be available (U.S. Department of Education, 2003b, 2006b; U.S. Department of Education, Office of Elementary & Secondary Education, 2004).

States are now struggling to identify outcomes on which to base alternate content standards and to find curriculum models that meet students’ needs in addressing these stan-

dards. In addition, they are struggling to link the alternate standards to their grade-level counterparts in accordance with expectations set by the No Child Left Behind Act.

Across the five Southwest Region states definitions and parameters for eligibility differ substantially

Outcomes and curriculum models. Test developers and policymakers struggle over the content and focus of state alternate assessments. Should alternate assessments focus on “the content standards (or core learning outcomes) identified for all students” or on “a separate, more ‘functional’ set of learner outcomes” (Kleinert & Kearns, 1999, p. 101)?

The functional-skills curriculum model for students with significant cognitive disabilities was intended to promote community inclusiveness. It was a paradigm shift from previous developmental models based largely on infant and early childhood curricula. Developmental models hinged on the belief that many students with significant cognitive disabilities would not continue to develop intellectually as their typically developing peers would (Browder et al., 2004) and were in essence based on students’ mental rather than chronological age. Early functional curriculum models, by contrast, focused primarily on skills for independent living, such as cooking, shopping, managing money, using public transportation, and living in the community (National Center on Learning Disabilities, 2007).

Functional curricula vary from district to district, and often from classroom to classroom, depending on student needs or the mandates of an individualized education program. One functional

curriculum includes personal-care skills (grooming, health, dressing, attending to medical needs), domestic skills (shopping, cleaning, cooking, budgeting, planning), recreation skills (making social connections, using the library, swimming, biking), and employment skills (prevocational, vocational, on-the-job training, community-based

job experiences; Provincial Outreach Program for Autism, 2007). A functional-skills curriculum is child-centered, not curriculum-centered. It is fluid, changing with the needs of the student, and is teacher-selected and teacher-directed to emphasize academic tasks that the student will use daily and can apply in real life.

The shift from a functional-skills assessment approach to one based more on academic skills has advanced the trend toward alternate assessments

The shift from a functional-skills assessment approach to one based more on academic skills has advanced the trend toward alternate assessments across the country. The states, however, had to decide how to relate academic content standards to the alternate assessments. Would they keep standards identical with general education standards, revise or amend the general education standards, or develop separate alternate assessment standards (Thurlow, 2004)?

A state's response to federal alternate assessment requirements would reflect decisions about its education values and priorities. Many educators struggled with the expectation that they would teach students to read and solve mathematical problems when many of the students still had not mastered basic life skills, such as using a consistent mode for communication (speaking, gesturing, using eye gaze), lifting their heads, or grooming themselves. As states began to implement the mandates of the Individuals with Disabilities Education Act of 1997, they faced the challenge of negotiating between the federal requirements and the pervasive, often competing needs of students (Ysseldyke & Olsen, 1999).

To meet federal requirements and provide students with disabilities access to the general curriculum,

states are attempting to combine the functional curriculum with the more traditional academic curriculum. Because functional academic curricula vary from state to state and from classroom to classroom, and research on functional academics is limited (Browder et al., 2004), it is difficult to identify the best blend. The Chicago Public Schools offers high school classes in functional academics developed specifically to address the goals and objectives of the Illinois Alternate Assessment. Most students in the classes have significant cognitive disabilities, but the classes are not limited to these students. Besides instruction in four main content areas, students take a variety of community-based classes, including (Chicago Public Schools, 2007):

- *Reading.* This class teaches reading comprehension and critical-thinking skills to students who can benefit from classroom reading instruction. Students are exposed to Chicago public libraries and receive weekly peer tutoring.
- *Math.* Available to students who function at a basic math level necessary for further development, this class is geared to the individual. It covers such areas as time, computation, money usage, and basic functional math.
- *Communication.* This class stresses functional vocabulary. Students who have difficulty receiving or sending spoken language explore the use of augmentative communication systems, such as pictures and signing, to make requests and share experiences. Students learn to identify, trace, or write sight words, depending on ability. Modified communication is geared to students who require more intensive instruction. Students learn basic functional academics such as counting, number recognition, following directions, learning community words and signs, and becoming more aware of their environment.
- *Computers.* Students are introduced to a computer's parts and functions and to various software programs. Individualized

enrichment activities are available depending on student needs. Adaptive equipment assists students with various degrees of physical involvement. And classes address various levels of computer literacy.

States have not made academic content standards the only foundation for alternate assessments (Browder, 2006; Towles-Reeves, Muhomba, & Kleinert, in press). Functional skills remain a priority in instruction and have not been eliminated from alternate assessment models. But many states, to meet policy regulations and more broadly educate their students, are using academic tasks and contexts for alternate assessments in reading and math. Thus, states are likely using a combination of academic and functional skills linked to state standards (Browder, Spooner, Ahlgrim-Delzell, et al., 2003). Most are leaning toward linking general education content standards to alternate assessments.

As states move from functional skills to more academic skills, teachers will need to understand the broader policy context and to be able to implement new curriculum expectations. Ascertaining the right blend of skills can be challenging. Teachers' abilities vary, as do student populations and state guidelines (Thurlow, 2004). Although some local education agencies and institutions of higher education are stepping up to support educators who must implement the assessments and align special education curricula, more data need to be collected in this area.

Alignment and linkage. Several researchers have developed models that differentiate alignment and linkage. *Alignment* designates the degree to which content (such as skills and concepts) concurs in two sets of standards or in an assessment and a set of standards. Alignment relationships tend to be direct relationships (skill-content matches) and are typically observed between standards and assessments for a single student population (such as general education, special education, or English language learners). *Linkage* refers to relationships that tend to be developmental, foundational, or

proximal and are typically observed between standards or assessments developed for different populations (such as general education standards and alternate standards; WestEd, 2004).

Researchers performing linkage studies need to include reviewers with content expertise and knowledge of the targeted student population. Relationships between sets of standards are typically less direct than those found with alignment studies, often linking precursor or support skills with general education grade-level expectations.

The Individuals with Disabilities Education Act of 2004 and the No Child Left Behind Act support the design of alternate assessments as an extension or modification of state assessment systems based on general education standards. States are developing and refining their alternate assessments to focus more on these standards: "In 1992, 32 percent of states were using only functional skills for their alternate assessments with no link to state standards[;] by 2001 only 8 percent were doing so" (Browder, Fallin, Davis, & Karvonen, 2003, p. 259). Similarly, a 2005 survey of state departments of education indicated that 90 percent of states were adhering to the requirements of the No Child Left Behind Act and the Individuals with Disabilities Education Act of 2004, using academic content standards as the basis of their alternate assessments in linking functional skills to content standards (Thompson et al., 2005).

Aligning alternate assessments to state academic content standards is critical. But such alignment remains a complex issue

Aligning alternate assessments to state academic content standards is critical for two reasons. First, it is intended to provide access to the general curriculum for students with significant cognitive disabilities, setting high expectations for students. Both the letter and spirit of the law reflect this intention (U.S. Department of Education, 2005, p. 6; Ysseldyke, Dennison & Nelson, 2004). Second, such assessment data are believed to lead to practices that improve instruction quality and

assessment policies (McDonnell, McLaughlin & Morison, 1997; Ysseldyke, Dennison & Nelson, 2004). But alignment remains a complex issue for alternate assessments. Because of the greater local choice and range of artifacts in these assessments, the method for demonstrating alignment between content standards and performance assessments is less developed than for on-demand tests.

Browder, Spooner, Ahlgrim-Delzell, et al. (2003) question how well versed special education teachers are in linking functional skills to content standards, concluding that teachers need to be trained on how to use functional academics. For example, while teachers are adept at using sight words in instruction, they need more training on how to teach story patterns and tie in their personal experiences (Browder, 2006; D. Farley, Education Consultant, Special Education Bureau at the New Mexico Public Education Department, personal communication, June 27, 2007; Louisiana Department of Education, 2006; Towles-Reeves et al., in press). Towles-Reeves et al. identified 18 studies conducted between December 2002 and 2006 examining the readiness of special education teachers to implement a

standards-based curriculum. Their summary provides further evidence of a disconnect between the intent to provide instruction based on academic content to students with significant cognitive disabilities and the practical abilities of teachers.

Browder et al. (2005) have explored the problem of validating performance indicators with content experts and stakeholders. They found that 42 states developed rubrics to measure student performance, but all states needed help providing evidence of reliability and validity from scores (see appendix A).

Other researchers have observed a change in states' curriculum philosophies, as they shift from functional to more academic alternate assessment achievement standards and performance descriptors. Because the academic focus is fairly narrow,

researchers have called for more studies to find the right blend of functional and academic standards (including performance indicators that reflect that blend), to help states give students with severe disabilities access to the general curriculum, and to improve alignment criteria (Browder et al., 2005; Browder et al., 2002; Browder, Fallin, et al., 2003; Browder, Spooner, Ahlgrim-Delzell, et al., 2003).

Defining technical adequacy for alternate assessments

The development or redesign of an alternate assessment will be driven by the design imperatives for technically sound assessments. All good assessments must be valid, reliable, and usable (Forte Fast, 2004; E. Forte, president, edCount, LLC, personal communication, February, 2006; Rabinowitz & Sato, 2005). But models for investigating the technical adequacy of alternate assessments are scarce or nonexistent. Only a few research studies are available on this topic. Although criteria for evaluating the validity of general education tests have been extensively written about (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 1999; Green, 1998; Messick, 1993; Webb, Horton, & O'Neal, 2002), Yovanoff and Tindal (2007) cite only seven published studies of alternate assessment validation, with much of the research based on surveys.

There is no consensus on what evidence should be collected in alternate assessments or how much of it is sufficient to ensure technical adequacy. The most significant finding in the literature is that alternate assessments for students with the most significant cognitive disabilities typically lack adequate psychometric properties, with shortcomings in validity, reliability, and usability.

Validity. Validity is the extent to which a test measures what it was designed to measure. According to standard 1.1 in the *Standards for educational and psychological testing* (American Educational Research Association et al., 1999), validity evidence should be collected for every intended interpretation and use of the scores that a measurement

There is no consensus on what evidence should be collected in alternate assessments or how much of it is sufficient to ensure technical adequacy

instrument yields (U.S. Department of Education, Office of Elementary & Secondary Education, 2004). Alternate assessments are intended to serve several purposes:

- To provide a measure of student proficiency to inform parents, to help teachers plan instruction for the following year (so that the evaluation of instructional programs informs ongoing instruction), and to comply with federal mandates—especially adequate yearly progress requirements. Browder et al. (2005), Snell and Brown (2006), and Towles-Reeves et al. (in press) indicate that results from alternate assessments should help teachers determine functioning at the time of testing and identify specific skills acquired, those requiring continued instruction, and support needed, including assistive technologies. Ideally, this process will inform the student’s individualized education program and support a plan for instruction for the following year (Towles-Reeves et al., in press).
- To hold teachers, schools, and districts accountable for implementing standards-based curricula and using assessment results to improve student learning. The annual development and administration process helps to focus educators on the development, instruction, and assessment of performance goals aligned with state performance standards (Towles-Reeves et al., in press). Since the inception of the No Child Left Behind Act, such results must be sufficiently valid to support adequate yearly progress decisions.
- To inform and support program evaluation at the classroom, school, and district levels, including the identification of resources that may further support instruction and provide topics for professional development (Towles-Reeves et al., in press).

Validity evidence traditionally takes several forms and comes from a variety of sources. But most evidence supporting the use of alternate assessments

for the indicated purposes consists of descriptions of how assessments were developed and the evidence submitted to peer review (R. Quenemoen, Technical Assistance Team Leader, National Center for Educational Outcomes, personal communication, October, 2006; M. Thurlow, Director, National Center for Educational Outcomes, personal communication, 2005). Although such evidence is important, other evidence is needed (U.S. Department of Education, Office of Elementary & Secondary Education, 2004; U.S. Department of Education, 2006a). Yet it has not always been collected.

Alternate assessments are individualized, making it difficult to collect traditional validity evidence.

A score can have different meanings for different students, depending on

their instructional goals and the characteristics of the assessments administered (Schafer, 2005; Towles-Reeves et al., in press). So, validity evidence must focus more on consequential aspects for alternate assessments than for traditional on-demand assessments—which exist primarily to measure what has occurred, not to influence how instruction should take place and be supported.

Some forms of validity evidence adapted to alternate assessment include:

- *Intrinsic rational validity evidence.* An artifact of the test development process, this evidence is intrinsic because it is built into the test, and rational because it is derived from rational inferences about the kinds of tasks that will best meet measurement goals (Ebel, 1983). In most states, including Southwest Region states, the evidence submitted to peer review on assessment development served as intrinsic rational validity evidence (U.S. Department of Education, 2006a). Some states in the Southwest Region, especially those redesigning their assessments, might benefit by reviewing how to summarize this evidence.

Alternate assessments are individualized, making it difficult to collect traditional validity evidence

- *Content-related or curricular validity evidence.*

This evidence addresses how adequately assessment tasks are aligned to an assessment's intended focus (material or standards). Several features of the annual development process, such as teacher training and a test administration manual, can provide evidence that assessment results measure intended content-standard objectives (Towles-Reeves et al., in press).

- *Face validity.* Addressing whether an assessment appears to measure what it is supposed to measure, this evidence is important to any assessment program (though no respected technical source considers it a substitute for either conceptual or data-derived validity evidence). Face validity can help teachers, parents, and community members accept the results of an assessment. If they do not see it as relevant or understand its purpose, they are less likely to give it their attention and support. A test's face validity is typically gauged by how stakeholders respond to the use of test results to inform instruction and monitor accountability. One can learn this by periodically surveying parents, teachers, and other groups of interest. Some Southwest Region states took another approach: reviewers made a frequency count of the skills that teachers selected on portfolios or performance tasks and computed correlations (National Alternate Assessment Center, 2005; R. Quenemoen, Technical Assistance Team Leader, National Center for Educational Outcomes, personal communication, October, 2006).

- *Consequential validity evidence.* A test's appropriateness to a set of assessment goals is determined by evaluating the intended and unintended consequences of the assessment process and results (Messick, 1993). This is especially important for alternate assessments, where the development and administration processes can be unusually complex and labor intensive. Many states are developing monitoring protocols

and teacher professional-development support materials to ensure that the alternate assessment process examines how classroom practices change, both in administering the assessment and in incorporating the results into instruction. Research on training teachers to understand alternate assessment's positive aspects and to use results in the classroom is a part of consequential validity. States are beginning to analyze and collect evidence for this process (Browder, Spooner, Ahlgrim-Delzell, et al., 2003; Browder, Fallin, et al., 2003; D. Farley, Education consultant, Special Education Bureau at the New Mexico Public Education Department, personal communication, June 27, 2007; Towles-Reeves et al., in press).

Reliability. Reliability quantifies the consistency of measurement results. Consistent measurements are a prerequisite to interpreting scores appropriately. For traditional large-scale assessments, with standardized items, administration, and scoring procedures, computing reliability is fairly straightforward. Most state assessment programs use a measure of internal consistency that makes reliability above 0.8 (and often exceeding 0.9) readily attainable, given sufficient sample sizes (typically in the tens of thousands) and number of items (30 and above; Nitko, 1996).

Alternate assessments are far less standardized in their prompts (see appendix A) and their administration and scoring protocols. And—because of the 1 percent rule, the logistical support required, and the perceived burden on teachers—sample sizes for alternate assessment are usually well below those for more traditional programs. Other reliability challenges include the limited population of students and teachers to support training and the impossibility of certain statistical data analyses and checks that depend on sample size, such as bias review procedures and discrimination indices (Browder, Fallin, et al., 2003; Browder, Spooner, Algozzine, et al., 2003).

Reliability for alternate assessments can be conceptualized in several ways. One is the consistency

Reliability for alternate assessments can be conceptualized in several ways

of the observed outcomes associated with a given skill (S. Hall, Division of Special Education and Early Intervention Services, Maryland State Department of Education, personal communication, May 22, 2006; National Alternate Assessment Center, 2005). If a student has mastered a skill, proficiency should be evident over multiple settings and occasions. Inconsistencies suggest that mastery interpretations cannot be generalized beyond the conditions of the original assessment task or that the student's performance was scored incorrectly (Browder, 2006; Towles-Reeves et al., in press).

Consistent use of a specified scoring process is important to reliability. The procedures and materials for training scorers must be standardized, not only for a given year but also across administrations whenever possible. To eliminate ambiguity the scoring process and rules must be documented. Validity and scoring reports must be reviewed daily so that scoring supervisors can identify scorers who are starting to drift. Finally, there must be scoring agreement.

Usability. Unlike for validity or reliability, there are no general guidelines or statistical indices to determine the usability of a test or assessment program (S. Hall, personal communication, May 22, 2006; Schafer, 2005). Many variables influence decisions about usability. A hotly debated question about the use of assessments for accountability is how useful test results are for teaching and learning. When students as a whole do poorly on a test, either the test is a poor measure of their learning or it accurately reflects the fact that they did not learn. Whether it is a poor measure, and thus not usable for instructional decisionmaking, depends primarily on alignment—that is, whether the test is a good (reliable and valid) measure of the curriculum or standards to be mastered. If a test is aligned with the curriculum, teachers can use results to evaluate learning and instruction.

A key premise underlying the use of alternate performance-based assessments for students with significant cognitive disabilities is that such

assessments can be curriculum-embedded (administered as part of regular classroom activities) and directly shape future instruction. Such consequential validity evidence is rarely collected. Another key usability issue is how the assessment results are communicated. Stating results in terms that most consumers—especially teachers—can understand helps teachers teach and helps parents and students understand student performance.

Alternate assessments are useful when they represent what students have been taught and when they yield consistent, accurate scores. If these conditions are met, states can make confident inferences about classroom performance from test scores. When academic standards influence classroom instruction (as some state content standards have), it is reasonable to use test scores in a content area as evidence for how much students have acquired the knowledge and skills specified by the standards.

To build the technical adequacy of alternate assessments, Kleinert, Browder, and Towles-Reeves (2005) are applying and generalizing the cognitive psychology-based assessment triangle (Pellegrino, Chudowsky, & Glaser, 2001)—originally developed for general education students—to students with significant cognitive disabilities (see appendix A). Although this framework holds much promise, more research is needed to understand its full applicability (E. Towles-Reeves, Research Coordinator, National Alternate Assessment Center, personal communication, May 2, 2006).

How useful are test results for teaching and learning?

Creating reliable and valid alternate assessments

Because traditional paper-and-pencil tests are inappropriate for students with significant cognitive disabilities, states have had to consider alternative approaches and to build more valid instruments. Significantly cognitively disabled students tend to have limited communication skills—some being nonverbal—and extremely low academic achievement levels. They need highly specialized

instruction and support, such as augmented communication systems (Almond & Bechard, 2005). These needs are often complicated by English language learner or low socioeconomic status. The broad heterogeneity of this population requires a broad and flexible assessment approach.

Although researchers have made progress determining the technical requirements of alternate assessments (see, for example, Rabinowitz & Sato, 2005), their adequacy continues to lag behind that of their general education counterparts—primarily assessments with multiple-choice and short and extended constructed response questions. Given the range of student needs, one size of alternate assessment will not fit all.

Three alternate assessment approaches are seen as most promising (for a fuller description of each, see Yovanoff & Tindal, 2007, p. 185; Roeber, 2002; and Towles-Reeves, et al., in press):

- *Checklists/rating scales based on observations.* Teachers are asked to rate whether students can perform selected behaviors. Scoring is based on the number of skills the student can perform successfully.
- *Portfolios/bodies of evidence.* Teachers systematically collect student work samples (see

appendix A), which are evaluated or judged against predetermined scoring criteria. Some states select the standards to be assessed. Others allow teachers on IEP teams to select what is assessed.

- *Performance assessments (performance events).* A series of tasks are administered and scored in terms of exemplars of proficient performance.

Researchers believe that these methods can be tailored to the needs of students with significant cognitive disabilities and provide substantively more access than traditional multiple-choice assessments. They have been undergoing psychometric evaluation for some time (Bennet, 1993; Messick, 1996; Traub & Fisher, 1977; Thissen, Wainer, & Wang, 1994, as cited in Yovanoff & Tindal, 2007).

Table 1 summarizes the assessment approaches commonly used nationally and how the prevalence of each approach has changed since the No Child Left Behind Act took effect.

The National Center on Educational Outcomes reported that in 2005 nearly half the states across the country were using either portfolio or performance-based assessments (Thompson et al., 2005). Performance and portfolio assessments are

TABLE 1

Alternate assessment approaches commonly used by states (number and percent of states), 1999–2005

Year	Portfolio, performance, or body of evidence assessments		Rating scale or checklist assessments		Individualized education program analysis		Other assessments		Assessments in development or revision	
	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent
1999	28	56	4	8	5	10	6	12	7	14
2001	24	48	9	18	3	6	12	24	2	4
2003	23	46	15	30	4	8	5	10	3	6
2005 ^a	25	50 ^b	7	14 ^c	2	4	7	14	8	16

a. One state had not developed any statewide alternate assessment approaches.

b. Of these 25 states, 13 use a standardized set of performances, events, tasks, or skills.

c. Of these seven states, three require the submission of student work.

Source: Thompson et al., 2005.

appealing because they can provide rich descriptions of students' real-life knowledge and skills (Elliott & Fuchs, 1997). But Browder, Spooner, Algozzine, et al. (2003) have expressed concerns with performance-based alternate assessments, suggesting that their technical characteristics and limitations might have led to suspect outcome scores for students and schools. In addition, the same study argues (based on initial data from Kentucky's efforts) that state portfolio-based alternate assessments might face challenges to the reliability of their scores. Finally, the study has questioned how well versed special educators are in linking functional skills to content standards. Its authors assert that much training is needed in this area—both to improve instruction and to meet federal mandates (Browder, Spooner, Algozzine, et al., 2003).

Because portfolios are work samples collected over time, many states employ a rubric to score portfolio-based alternate assessments. Some states permit the students' teachers to score the work, for efficiency and in the belief that knowledge of the student and the context in which the work was collected is essential for valid scoring. Other states maintain a more rigorous process using more than one teacher to score student work to protect against bias and to measure interrater agreement. Although external scoring might seem advantageous, the administrative process and the artifacts for this student population make off-site scoring challenging. In either approach, allowing for personal judgment jeopardizes the scoring's reliability and validity (Kleinert, Farmer-Kearns, & Kennedy, 1997; North Central Regional Education Laboratory, 2007; Thompson & Thurlow, 2003).

The reliability of ratings was also challenged in states attempting to use portfolios and performance assessments as part of their general large-scale assessments, such as Arizona and Vermont. Indeed, these states were prevented from publicly reporting their assessment results (Koertz, McCaffrey, Klein, Bell, & Stecher, 1993; Tindal et al., 2003). Performance assessments may need to include numerous tasks and work samples if they are

to demonstrate adequate coverage of state standards and provide generalizable results. Because such processes are extensive and time-consuming, they are contraindicated for this population—one in which students must be assessed on a one-on-one basis, tend to tire easily, and lack long attention spans (Tindal et al., 2003; Browder, Fallin, et al., 2003; Almond & Bechar, 2005).

Demonstrating reliability in performance-based alternate assessments can be challenging

Demonstrating reliability in performance-based alternate assessments can be challenging. The primary determinant of reliability is the number of items or tasks on a test. Alternate assessments typically consist of a few larger tasks, rather than 40–60 discrete multiple-choice items. Because reliability depends partly on the number of discrete tasks that make up an assessment, it is possible to increase an alternate assessment's reliability by breaking larger tasks down into smaller subtasks for analysis. Still, the wide range of skills and tasks targeted by performance-based alternate assessments creates further challenges for comparability and for determinations of across-the-board technical adequacy. Browder, Spooner, Algozzine, et al. (2003) also identify student risk factors (such as unstable student behavior or health status) as possible influences on alternate assessment results. For on-demand performance tasks, fluctuations in student behavior or physical well-being could yield invalid results.

Performance-based alternate assessments typically employ some local "human" scoring. States can address this additional source of unreliability by carefully training scorers, monitoring implementation, and moderating locally derived scores. Allowing teachers a choice of performance tasks raises additional validity and reliability questions. Several states are using a combination of methods (such as portfolios with a checklist). This multi-method approach is designed both to increase the technical quality of the results and to meet the diverse access needs of a heterogeneous student population (see appendix B for demographic

information). All approaches need further study and articulation, both individually and in combination (Thompson et al., 2005; Towles-Reeves et al., in press).

Defining proficient performance (setting standards)

In December 2003 the U.S. Department of Education proposed a change in policy for the No Child Left Behind Act and students with significant disabilities. It recommended that states be permitted “to define alternative achievement standards for students with the most significant cognitive disabilities. Such students will take an alternate assessment. These alternate achievement standards must be aligned with the state’s academic content standards and reflect professional judgment of the highest learning standards possible for those students” (U.S. Department of Education, 2003b). States now had to decide what content to include on alternate assessments and how to define content mastery. Several states—including all five in the Southwest Region—defined mastery by setting standards for their alternate assessments (Lewis, Mitzel, & Green, 1996; Roach & Elliott, 2004; see table C7 in appendix C).

To set alternate achievement standards is to establish cutscores corresponding to the knowledge, skills, and competencies that constitute proficiency at each level of performance. As important as the cutscores themselves is a performance descriptor (see appendix A) that indicates typical student knowledge, skills, and abilities for a given score. The descriptor helps the teacher identify skills and abilities that a given student cannot yet perform consistently, communicate with others about the student’s progress, and determine next year’s instructional goals. Finally, it indicates the

student’s status relative to state learning standards (Roach & Elliott, 2004).

The most common standard setting method for general assessment, the bookmark method (Lewis et al., 1996), is better

sued to assessments with multiple items. (For more information on the bookmark method, see Kiplinger, 1997 and Olsen, Mead, & Payne, 2002.) More and more, states are using the body of knowledge method for alternate assessments, since it is better suited to assessments that include direct student performance (such as work samples; S. Bechard, Director, Office of Inclusive Educational Assessment at Measured Progress, personal communication, December 2006).

Using bookmark, reviewers must make fine distinctions between performance levels on adjacent items with slightly different item response theory difficulty values (Hambleton & Swaminathan, 1985; Tindal et al., 2003). Item response theory is the study of test and item scores based on assumptions concerning the mathematical relationship between abilities (or other hypothesized traits) and item responses (Baker, 2001). When assessments are more task-driven (Tindal et al., 2003), the grain size makes item-by-item comparisons more difficult to differentiate. Body of knowledge takes a more holistic view of performance, requiring standard setters to review and differentiate between profiles of performance rather than individual items.

Several issues specific to setting performance standards on alternate assessments for students with the most significant cognitive disabilities remain unsettled (Almond & Bechard, 2005; Thompson et al., 2005). First, given the diversity of this population, applying one standard for all is problematic—especially since many, because of their disabilities alone, may never be able to meet the state-determined standard. Equally challenging is setting standards on performance assessments, given the range of student evidence obtained and the ability of teachers to develop and administer tasks of their own.

The relatively small number of students in some states makes obtaining sufficient datasets for complex standard setting procedures difficult. Finally, a lack of clarity about the targeted population and content standards often yields vague performance

More and more, states are using the body of knowledge method for setting alternate assessment performance standards

descriptors. Three of the five states in this study received negative comments related to their setting of standards from a national peer review process to determine whether state standards meet No Child Left Behind requirements (U.S. Department of Education, 2006a).

A REVIEW OF ALTERNATE ASSESSMENTS ACROSS THE SOUTHWEST REGION STATES

This section focuses on alternate assessments across the Southwest Region states. It addresses the last five of the six research questions (2–6) that guided this study, namely:

2. What do alternate assessments across the Southwest Region states look like?
3. What training or professional development is provided for educators on alternate assessments?
4. How are results collected and used at the state, district, school, and student levels?
5. To what extent do states' alternate assessments capture the same or similar skills as state tests designed for the general student population?
6. What technical issues are states facing in developing and implementing reliable and valid alternate assessments?

In asking these questions about each Southwest Region state, the authors found that all five states were making efforts to develop policies and practices to include students with significant cognitive disabilities—though all five states still faced significant challenges. The Southwest Region states report including more students with significant cognitive disabilities in state assessments, increasing student exposure to the general curriculum, and reforming special education curricula and instruction (see appendix C). Yet the five states differ in their definitions of significant

cognitive disability, their guidelines for student participation, and the technical quality of their assessments.

Arkansas and Oklahoma, which have achieved full approval status from federal peer review, report that they are continuing to improve programming and instruction by refining content achievement standards to reflect a blend of functional and academic skills linked to content standards. Louisiana has changed from the Louisiana Educational Assessment Program Alternate Assessment (LAA), a checklist, to a performance-based assessment (LAA 1). It considered moving to portfolio assessment but opted for a traditional test (Jeanne Johnson, Education Consultant, Louisiana Department of Education, personal communication, January 30, 2008). New Mexico has switched from a checklist based wholly on functional achievement to performance-based tasks that are linked to alternate achievement standards. Texas has created alternate achievement standards based on state general content standards and is transitioning from locally selected alternate assessments and an optional state-developed alternate assessment to a uniform state-developed portfolio system.

What do alternate assessments across the Southwest Region states look like?

Demographic context. To facilitate meaningful comparisons table 2 summarizes state demographic information for the Southwest Region states. (For more specific information about state characteristics, see appendix B.)

Texas has by far the largest enrollment, with six times as many students enrolled in K–12 as the state with the second-largest enrollment (Louisiana) and significantly more public schools and public school districts (local education agencies) than the other four states. While it takes similar efforts to design, plan,

All five states were making efforts to develop policies and practices to include students with significant cognitive disabilities—though they still faced significant challenges

TABLE 2

Summary of state demographic statistics for the Southwest Region states

Statistic	New				
	Arkansas	Louisiana	Mexico	Oklahoma	Texas
K–12 enrollment	463,115	724,281	326,102	629,426	4,405,215
Public schools (number)	1,158	1,541	842	1,747	8,746
Local school districts (number)	252	68	89	540	1,227
Students receiving free or reduced-price lunch (percent)	52	62	58	54	48
Students receiving special education or related services (percent)	12	14	20	15	12
Students receiving English language learner services (percent)	4	2	19	7	16

Source: State report cards for Arkansas (<http://normessasweb.uark.edu/reportcards/state05.php>), Louisiana (http://www.doe.state.la.us/lde/pair/StateReport0405/10-Student_Achievement.pdf), New Mexico (<http://www.ped.state.nm.us/div/acc.assess/accountability/dlRptCard2005/NMStateReportCard%20English.pdf>), Oklahoma (<http://title3.sde.state.ok.us/studentassessment/2005results/reportcard2005state.pdf>), and Texas (http://www.tea.state.tx.us/research/pdfs/2005_comp_annual.pdf).

and devise assessment systems, differences will arise in resources, cost, and logistics when materials have to be distributed, collected, retrieved, and scored for larger numbers of students. For instance, although development of the assessment itself is a roughly equivalent task in each state, state scoring of portfolios or performance events for 40,000 students in Texas requires vastly greater resources and planning than for 2,200 students in New Mexico—an issue that influenced Texas’s decision to have teachers score tests and submit scores themselves. In New Mexico, by contrast, the state will score the results for 2007/08 (D. Farley, Education Consultant, Special Education Bureau at the New Mexico Public Education Department, personal communication, June 27, 2007; C. Wieland, Director, Special Education Assessments, Texas Education Agency, personal communication, June 2007).

Special education students can be expected to be overrepresented in the group of students receiving free or reduced-price lunches (Individuals with Disabilities Education Act regulations, 34 CFR 300.8). National literature and state reports show that low socioeconomic levels (as represented by this indicator) affect academic performance for both general education students and their special education counterparts.

For K–12 students in the five Southwest Region states, race and ethnicity are another important

factor affecting assessments (see appendix B, table B1). In Louisiana, New Mexico, and Texas non-White students are a majority. In New Mexico Hispanic students are the majority (53 percent). Texas reports a 45 percent Hispanic population. Although not all Hispanic students speak Spanish as their primary language, such data underscore the concerns of state contacts who note that additional assessment and instruction issues arise for special education students who are English language learners—for example, they may require additional language-based accommodations such as access to dictionaries or translations (D. Farley, Education consultant, Special Education Bureau at the New Mexico Public Education Department, personal communication, June 27, 2007; C. Wieland, personal communication, 2007).

New Mexico (20 percent), Oklahoma (15 percent), and Louisiana (14 percent) have the highest percentages of students receiving special education and related services, all higher than the national average of 10–12 percent. New Mexico has the highest percentage of students receiving both special education and English language learner services, with Texas next. Both states report high numbers of students receiving special education and English language learner services. Such numbers exacerbate the difficulties of developing alternate assessments and training teachers to implement them and interpret the results properly. State contacts expressed a need for technical

assistance to develop better ways of assessing non-English-speaking students with significant cognitive disabilities (D. Farley, Education consultant, Special Education Bureau at the New Mexico Public Education Department personal communication, June 27, 2007; C. Wieland, personal communication, June, 2007).

Heterogeneity in the population of most significantly cognitively disabled students intensifies the technical and logistical challenges for alternate assessment, as states attempt to create and implement valid, reliable, and bias-free tests for a population with large numbers of poor students, minorities, or both.

Population definition. States are permitted to differ in their definitions, policies, and practices related to alternate assessments. This study found that the five Southwest Region states used several frameworks to identify students for participation, with eligibility parameters differing substantially across the states.

Each state defined *significantly cognitively disabled* differently, within the constraints of its own statutes and administrative codes as well as federal guidance (see table C3 in appendix C). And each state operationalized its definition differently (see table C2). Variability in screening criteria has contributed to state-to-state variations in state assessments and to varying and perhaps unclear student characteristics (Thurlow, 2004). The screening criteria are broad, general, and sometimes ambiguous. They are specific to the group of students with significant cognitive disabilities, but not to subgroups within this broadly defined group.

Arkansas had the most detailed definition, close to the American Association of Mental Retardation's description (2004). The Arkansas definition indicates that students in this population have difficulty functioning cognitively across settings, have limited academic skills, and require extensive supports. New Mexico uses a checklist that covers these points. Louisiana has a description that specifies a diagnosis (moderately severe or profound

mental disability) and describes having a significant cognitive disability as having severe mental disabilities, being multiply disabled, or having autism or traumatic brain injury. In Oklahoma the IEP teams and in Texas the admission, review, and dismissal teams define cognitive disability based on state-specified criteria.

Numbers participating in alternate assessments.

Through the survey conducted for this study, contacts from each state indicated that the number of students with significant cognitive disabilities taking alternate assessments was consistent with their percentage of the state's student population (table 3).

In each state the contacts asserted that the state guidelines and definitions were adequate and fit their state context. The Arkansas contact pointed to a passage from the federal nonregulatory guidance concerning the 1 percent rule: "The rule does not limit the number of students with the most significant cognitive disabilities who may take alternate assessments based on alternate achievement standards when that is appropriate. It addresses only the inclusion of proficient and advanced scores for alternate assessments based on alternate achievement standards in adequate yearly progress calculations" (U.S. Department of Education, 2005).

Texas has 40,000 students taking alternate assessments (C. Wieland, personal communication, June, 2007). The other states had far fewer students, but their numbers are consistent with the respective size of the states. Available state reports and interviews indicate that students currently participating in alternate assessment typically have a special education label of autism, mental retardation, or multiple disabilities. This finding is consistent with research by Almond and Bechard (2005) and the National Alternate Assessment Center (2005).

Purpose of testing. The five Southwest Region states have laws, statutes, regulations, codes, board policies, and other public

Additional assessment and instruction issues arise for special education students who are English language learners

TABLE 3

Students taking alternate assessments in Southwest Region states, by grade, 2005/06

	Arkansas	Louisiana	New Mexico	Oklahoma	Texas
Total enrollment K–12	463,115	724,281	326,102	629,426	4,405,215
Alternate assessment grades tested	3–8, 11	3–8, 10	3–10	3–8, 10–12	3–10
Students enrolled in grades tested	246,399	346,134	204,488	400,692	2,724,679
Students taking alternate assessment in grades tested	3,700	4,800	2,200	3,000	40,000
Percent of all students in grades tested taking alternate assessment	1.5	1.4	1.1	0.8	1.5
Percent of all students in grades K–12 taking alternate assessment	0.8	0.7	0.7	0.5	0.9

Source: For enrollments, U.S. Department of Education, National Center for Education Statistics, 2007 (<http://nces.ed.gov/ccd/>); for Arkansas, alternate assessment, C. Marvel, Math and Assessment Specialist, Arkansas Department of Education, June 2007 and Tom Hicks, Special Projects, Arkansas Department of Education, June 2007; for Louisiana alternate assessment numbers, J. Johnson, Education Consultant, Louisiana Department of Education, June 2007; for New Mexico alternate assessment numbers, D. Farley, Education Consultant, Special Education Bureau at the New Mexico Public Education Department, June 2007; for Oklahoma alternate assessment numbers, A. Daugherty, Coordinator, Compliance Activities and Assessment Special Education Department, Oklahoma Department of Education, June 2007; for Texas alternate assessment numbers, C. Wieland, Director, Special Education Assessments, Texas Education Agency, survey comment, 2007. For enrollments, state report cards for Arkansas (<http://normessasweb.uark.edu/reportcards/state05.php>), Louisiana (http://www.doe.state.la.us/lde/pair/StateReport0405/10-Student_Achievement.pdf), New Mexico (<http://www.ped.state.nm.us/div/acc.assess/accountability/dlRptCard2005/NMStateReportCard%20English.pdf>), Oklahoma (<http://title3.sde.state.ok.us/studentassessment/2005results/reportcard2005state.pdf>), and Texas (http://www.tea.state.tx.us/research/pdfs/2005_comp_annual.pdf).

documents describing the purposes of alternate assessments for students with significant cognitive disabilities (see table C2 in appendix C). All five states share the intent of the No Child Left Behind Act and the Individuals with Disabilities Education Act of 2004 to meet federal and state requirements and to provide accountability data for programs; to provide instructional information to teachers, parents, and students; and to improve student outcomes.

All five Southwest Region states have created definitions and guidelines for participation in alternate assessments, as required by the Individuals with Disabilities Education Act of 2004 (see tables C3 and C4 in appendix C). To meet the requirements of this act—and those of the No Child Left Behind Act—all five states have passed laws, regulations, rules, state board policies, or administrative code (nomenclature varies by state; see table C2 in appendix C). All five states describe their alternate assessments as intended for students who are unable to participate in state and district assessments, even with accommodations. This description matches the acts' requirements.

Where states vary is in their working definition of *significantly cognitively disabled*, their criteria for participation in alternate assessments, and the status of the peer review process—Arkansas and Oklahoma have received ratings of full approval, New Mexico has received approval expected, and Louisiana and Texas have received approval pending. States are addressing the second and third purposes, to improve programs and instruction, by shifting from a solely functional curriculum to a functional curriculum blended with academics and linked to academic content standards. The result: a notable shift of curricular philosophy, as teachers report spending more time on more academically based instruction because of alternate assessment (Browder et al., 2005).

Alternate assessment approach. When the Individuals with Disabilities Education Act of 1997 was passed, states moved quickly to develop alternate assessments. At that time Louisiana and New Mexico used checklists to assess students. Texas used locally selected alternate assessments and also produced the State-Developed Alternate Achievement I, a multiple-choice assessment.

TABLE 4

Southwest Region state alternate assessment approaches for 2007/08

	Arkansas	Louisiana	New Mexico	Oklahoma	Texas
Approach	Portfolio	Performance	Performance	Portfolio	Portfolio
Program title	Arkansas Alternate Portfolio Assessment System (AAPAS)	LEAP Alternate Assessment Program (LAA 1)	New Mexico Alternate Performance Assessment (NMAPA)	Oklahoma Alternate Assessment Program (OAAP)	Texas Assessment of Knowledge and Skills–Alternate (TAKS–ALT)
Grades	3–8, 9 (for math), 11	3–8, 10	3–10	3–8, 10–12	3–10
Subject	English language arts, math, science	English language arts, math, science, social studies	English language arts, math, science, writing	Reading, writing, math, science, social studies	Reading (3–9) math (3–10) Writing (4 and 7) English language arts (10)

Source: Arkansas Department of Education web sites (<http://www.arkansased.org/students/assessment.html>; <http://www.arkedu.state.ar.us/actaap/index.htm>; http://www.arkedu.state.ar.us/actaap/student_assessment/student_assessment_p1.htm) and survey and interviews with department contacts (C. Marvel and T. Hicks); Louisiana Department of Education web sites (<http://www.doe.state.la.us/lde/saa/2273.html>; <http://www.doe.state.la.us/lde/accountability/home.html>; <http://www.doe.state.la.us/lde/saa/2343.html>) and survey and interviews with department contact (J. Johnson); New Mexico Public Education Department web sites (<http://legis.state.nm.us>; <http://www.ped.state.nm.us/div/acc.assess/assess/info.update.corner.html>; <http://www.ped.state.nm.us/div/acc.assess/assess/index.html>) and survey and interviews with department contact (D. Farley); Oklahoma State Department of Education web sites (<http://www.sde.state.ok.us/home/defaultie.html>; <http://www.lsb.state.ok.us>; <http://title3.sde.state.ok.us/studentassessment>) and survey and interviews with department contact (A. Daugherty); Texas Education Agency web sites (<http://www.tea.state.tx.us/student.assessment>; <http://www.tea.state.tx.us/curriculum.html>; <http://www.legis.state.tx.us>; <http://www.sos.state.tx.us/tac/index.shtml>; <http://www.tea.state.tx.us/student.assessment/resources/taksalt/index.html>) and survey and interviews with agency contact (C. Wieland).

Arkansas and Oklahoma were using a body of evidence (portfolio) alternate assessment.

All five states in the Southwest Region reported that they were planning to use portfolios or performance tasks in the 2007/08 school year (see tables C1, C5, and C8 in appendix C). Arkansas, Oklahoma, and Texas were to use portfolios, and Louisiana and New Mexico were to use performance tasks (table 4). In addition, the states are focusing their efforts on standards-based assessment in reading, math, and science. Louisiana and Oklahoma assess social studies and science. New Mexico and Texas include writing (see table C1 in appendix C).

The Southwest Region states differ greatly in selecting tasks to be administered as part of the alternate assessment (see table C5 in appendix C). Arkansas and New Mexico predetermine standards and tasks to be assessed in all content areas. In Oklahoma and Texas the state predetermines two academic standards to be assessed and teachers choose three additional standards to be measured. In Louisiana teachers may determine

what is assessed based on targeted indicators, and required state tasks are combined with those developed locally. It should be remembered that validity and reliability questions arise when teachers are allowed a choice of performance tasks, as they are in some Southwest Region states (see table C7 in appendix C).

Each state reports clear evidence of a shift in state alternate achievement standards and performance descriptors—from the functional to the academically focused. But the academic focus remains narrow. The five states have indicated that they need technical assistance to define the right blend between functional and academic standards and instruction (including performance descriptors that reflect the blend) to train teachers.

What training or professional development is provided for teachers on alternate assessments?

State contacts reported on what the five states are doing to train teachers and on each state's perceived needs in this area.

Current training efforts. Each state has its own procedures for training teachers to administer and interpret the results of alternate assessments (see table C6 in appendix C). Arkansas uses a regional train-the-trainer system. State and regional staff provide training to district staff, who then train teachers in their districts and schools how to administer alternate assessments, how to teach students with cognitive disabilities, and how to interpret score reports. Texas also uses a regional train-the-trainer model. State education agency staff and a contractor train the trainers in 20 regional centers. Participants are trained on participation guidelines, the grade-level curriculum, how to record data, how to score, and how to use the online system. The trainers then train local education agency staff and special school staff. Texas offers the training modules online. Regional centers across Texas offer training on related topics such as assistive technology and instructional modules. The Texas Education Agency sponsors an annual assessment conference that offers educators additional training on alternate assessment. Louisiana also uses a train-the-trainer model, with topics including how to administer the assessment and how to interpret and use results.

In New Mexico Public Education Department staff members and professors of special education from the University of New Mexico provide an annual, intensive three-day training session in three modules (an overview, what is measured, and scoring) for alternate assessments. Teachers watch a video of students taking the assessment and then score case studies. Web-based training is offered for support. Teachers must pass a test to qualify to administer the alternate assessment. More than 1,200 teachers have been trained. Every other year teachers are trained on best practices of instruction.

In Oklahoma Department of Education staff and a contractor train teachers in the state's five regions to administer the assessment and interpret the score report.

Several training needs were commonly identified by state contacts in the Southwest Region

Training needs. The following training needs were among those most commonly identified by state contacts in the Southwest Region:

- *Training on definitions and guidelines for IEP teams.* Because individualized education programs are responsible for connecting present levels of performance to annual education goals and assessments, more comprehensive training would likely improve decisionmaking. According to state contacts, teams need training on how to understand participation guidelines (including how to integrate present levels of performance and adaptive skills into participation criteria), how to select accommodations, and how to fit alternate assessments into state assessment systems.
- *Training for teachers on how to link functional skills and academics to alternate achievement standards.* Some teachers lack background knowledge for aligning functional skills to the curriculum and standards. State contacts report a need for uniform training across districts. One needed skill is how to link the goals and objectives of an individualized education program to the alternate achievement standards. Another is how to identify and develop functional and academic skills that emphasize a broader generalization of life skills and that are related to meaningful outcomes (such as inclusive learning, postsecondary experiences, vocational and employment opportunities, and community participation).
- *Training for teachers on how to access the general education curriculum.* State contacts suggest a need for training in effective classroom practices (explicit instruction, differentiated instruction, peer-mediated instruction, classroom management, and universal design) and curriculum enhancements (curriculum modifications, curriculum accommodations, graphic organizers, text enhancements, and computer-based lesson enhancements).

- *Training for teachers and IEP teams on how to align assessment and instruction.* Because knowledge of assessment practices and of how to interpret results is uneven, training could improve individualized education programs and student outcomes. State contacts indicate a need for structured frameworks that link performance standards and indicators to the curriculum and that integrate assessment tasks into daily instruction—training teachers to map instructional tasks to content standards (to assess curriculum and instruction), to monitor curricular breadth and depth, and to understand alignment models as they apply to students with disabilities.
- *Training for all staff on how to use data to inform instruction.* Many states are struggling in this area. State contacts suggested that providing states with models or support (technical assistance or training) on how to employ data-driven activities would allow them to pass this information on to local practitioners and could potentially result in feedback on their instruction. For example, school staff could receive training on identifying, collecting, and evaluating the appropriate data to modify or develop instruction and to use data for reporting.

State contacts also identified the technical assistance they would like.

- *Help in developing fair, bias-free assessments for students with significant cognitive disabilities.* State contacts would like training in practices that have been successful in other states and in methods to demonstrate lack of bias and validity for alternate assessments.
- *Help in operationalizing definitions and alternate assessment participation guidelines.* State contacts suggest that a written process or step-by-step guide would help ensure a fair and uniform inclusion process.
- *Technical assistance on linkage and alignment between standards, assessment, and instruction.* States can receive technical assistance (for example, through dissemination from expert sources) to build their understanding of this linkage.
- *Help in writing descriptors for various levels of performance.* This part of the assessment process has proven challenging for many states. State contacts suggest that training (including discussion of content standards, performance levels, the test, and expectations for students) could be set in the context of standard setting and that models from other states and programs could be provided.
- *Help in identifying, evaluating, selecting, and implementing best practices for alignment, scoring, rubrics, and standard setting.* State contacts indicate that states need such help (or general training) to build their alternate assessments or improve current practices.
- *Help in collecting and reporting evidence for validity, reliability, and usability—including developing technical reports.* States must demonstrate the technical adequacy of their assessments for a variety of legal, technical, and ethical reasons, including passing the No Child Left Behind peer review process. State contacts indicate that of several possible training approaches, the most efficient might be to provide models from other more advanced programs. In addition, they suggest that states can be directed to web sites or organizations that provide technical assistance or to publications, manuals, and other sources of needed guidance.
- *Guidance on using assessment data to inform and improve instruction.* According to state contacts, both state and local education agencies would benefit from such guidance, which could include how to develop alternate assessments that best produce measurable results and how data drive accountability. They say that training or other dissemination resources on current accountability and school reform

issues would help drive the development of aligned alternate assessments.

All five state contacts highlighted the needs for teacher training and technical assistance. Many teachers and state department of education staff need to be trained on functional academics or on how to use the right blend of functional skills and academics in teaching (Almond & Bechard, 2005; Towles-Reeves et al., in press).

In addition to ascertaining the right blend of functional skills and academic skills, Browder, Fallin, et al. (2003) emphasize the need to train teachers in aligning instruction with assessment and in using results. Improving alternate assessment practices also requires understanding teachers' perspectives and considering how they have been trained for alternate assessments. Kim et al. (2006) found that teachers had a negative perception of their alternate assessment system—in part because, by their own report, they had only a limited general understanding of large-scale assessment systems and the interface between these systems and academic instruction. According to Kleinert, Kennedy, and Kearns (1999), although teachers felt positively about students with disabilities being included in large-scale assessment systems, they reacted negatively to the amount of time spent completing and implementing the portfolio, and they questioned the reliability of scores. Kampher, Horvath, Kleinert, and Kearns (2001) posit that

teachers who felt time constraints completing and implementing the portfolio reached their (self-reported) negative perceptions primarily because they had not been adequately trained and prepared to align the curriculum to the assessment.

Many teachers and staff need to be trained on functional academics or on how to use the right blend of functional skills and academics

All state contacts suggested that additional research focus on the following questions:

- Are teachers increasingly part of a long-term professional training program that improves instruction in math or reading?
- Are students increasingly meeting the standards because of focused instruction in math or reading?

How are results collected and used at the state, district, school, and student levels?

States are using the 1 percent rule in reporting alternate assessment results for adequate yearly progress. Because all scores must be included in adequate yearly progress calculations, any alternate assessment scores of “proficient” and “advanced” for students with the most significant cognitive disabilities that exceed the 1 percent cap must be counted as nonproficient (U.S. Department of Education, 2005, p. 7). States can request a slightly higher cap. When making such a request, states must adhere to strict eligibility criteria and address several issues, including (U.S. Department of Education, 2003b, p. 2):

- Incidence rates of students with the most significant cognitive disabilities.
- Circumstances in the state that would explain the higher incidence rates (such as specialized health programs or facilities).
- Documentation showing that the state has implemented safeguards to limit the inappropriate use of alternate assessments.

States may also grant districts an exception allowing them to exceed the 1 percent cap.

The authors of the present study verified the application of the 1 percent rule by looking at state assessment data submitted to the U.S. Department of Education, as reported by Thurlow, Moen, and Altman (2006). For 2003/04 Arkansas, New Mexico, and Oklahoma were within the 1 percent cap for reading and math. Louisiana exceeded the cap by 16 students for high school, reporting 1.36 percent for reading and 1.16 percent for math at proficient and above. Texas exceeded the cap significantly for reading (5.54 percent for elementary, 5.70 percent for middle school,

and 5.03 percent for high school) and math (4.61 percent for elementary, 4.87 percent for middle school, and 5.01 percent for high school). According to Texas’s reported percentages, the state exceeded the cap by an average of 13,500 students per grade.

States must identify the content to be included on alternate assessments and must decide how mastery of this content is defined. To define mastery the five states in the Southwest Region undertook standard setting procedures for their alternate assessments (see appendix A and table C7 in appendix C).

Because states define their own performance levels (such as advanced, basic, below basic), comparisons across states can be difficult. That said, this study’s findings replicate a national pattern of significant differences across states for the number and percentage of students scoring proficient or above on alternate assessments (National Center on Educational Outcomes, 2005). The authors of this study examined evidence from the 2005/06

school year because these were the data submitted for the peer review process. Table 5 summarizes the percentages of students at proficient or above for mathematics and reading.

In the state report cards for 2005/06 some states report large differences between the percentage of students scoring proficient or above on the 1 percent alternate assessment and on the general education criterion-referenced test (table 6; see also table C1 in appendix C).

Differences between proficient scores on general and alternate assessments cannot be attributed solely or primarily to instructional practices. States define their own achievement levels. And the peer review results—as well as ongoing revisions to assessment approaches in each state—indicate that the technical quality of states’ alternate assessments has not been fully demonstrated. States must monitor alternate assessment administration and scoring more closely to ensure that procedures are followed uniformly and objectively. Further, states must explain any differences in

TABLE 5
Students scoring at proficient or above on alternate assessments in math and reading by grade and state, 2005/06 (percent)

Grade	Arkansas		Louisiana		New Mexico		Oklahoma		Texas	
	Math	Reading	Math	Reading	Math	Reading	Math	Reading	Math	Reading
3	67	57	50	64	30	67	88	90	98	96
4	60	61	53	75	41	67	87	88	95	91
5	50	56	60	75	36	66	83	88	94	91
6	57	59	68	80	46	70	84	88	88	88
7	50	53	64	76	53	77	71	74	83	84
8	44	66	67	80	48	73	79	81	83	86
9	—	—	—	—	49	78	—	—	74	80
10	—	—	68	77	50	74	75	76	78	80
11	—	—	—	—	—	—	66	66	—	—
12	—	—	—	—	—	—	63	40	—	—

— is not available or no assessment at this grade.

Source: For math percentages: in Arkansas, Arkansas Department of Education, 2006c; in Louisiana, Louisiana Department of Education, 2007; in New Mexico, New Mexico Public Education Department 2007a; in Oklahoma, Oklahoma State Department of Education, 2006; and in Texas, Texas Education Agency, 2007. For reading percentages: in Arkansas, Arkansas Department of Education, 2006c; in Louisiana, Louisiana Department of Education, 2007; in New Mexico, New Mexico Public Education Department 2007b; in Oklahoma, Oklahoma State Department of Education, 2006; for Texas, Texas Education Agency, 2007.

TABLE 6

Students scoring proficient or above for general and alternate assessments in math and reading in grades 4 and 5 in the Southwest Region states, 2005/06

Assessment and grade	Arkansas	Louisiana	New Mexico	Oklahoma	Texas
<i>Math</i>					
General assessment					
4	50	61	36	—	—
5	—	—	—	74	81
Alternate assessment					
4	—	53	41	—	—
5	50	—	—	83	94
<i>Reading</i>					
General assessment					
4	—	—	—	—	—
5	50	64	55	77	81
Alternate assessment					
4	—	—	—	—	—
5	56	75	66	88	91

— is not available or no assessment at this grade.

Source: State report cards for Arkansas (<http://normessasweb.uark.edu/reportcards/state05.php>), Louisiana (http://www.doe.state.la.us/lde/pair/StateReport0405/10-Student_Achievement.pdf), New Mexico (<http://www.ped.state.nm.us/div/acc.assess/accountability/dlRptCard2005/NMStateReportCard%20English.pdf>), Oklahoma (<http://title3.sde.state.ok.us/studentassessment/2005results/reportcard2005state.pdf>), and Texas (http://www.tea.state.tx.us/research/pdfs/2005_comp_annual.pdf)

proficiency rates in scores on general and alternate assessments.

Finally, the No Child Left Behind Act and the Individuals with Disabilities Education Act of 2004 require that assessment systems collect data and provide reports at the student, class, school, district, and state levels. A 2005 report on the status of the states in special education by the National Center on Educational Outcomes found that states were complying with these requirements for accountability purposes (Thompson et al., 2005). The authors of the present study, however, have noted that states differ greatly in the data they report, in how they report it, and in how clearly they present information across the required levels (student, class, school, and local education agency). Different states use different cells and different language. For example, New Mexico's student reports show a performance level for each standard measured. But the other states report chiefly at the content-strand level.

To what extent do state alternate assessments capture the same or similar skills as state tests designed for the general student population?

State systems of standards and assessment provide useful information for valid accountability decisions and education improvement only to the degree that all components are aligned or linked. State assessments must be linked to state standards and—ultimately—to instruction. The federal peer review process requires each state to present evidence that its assessment system is aligned or linked to its standards and to submit the evidence from alignment studies and plans to fill any coverage gaps (National Alternate Assessment Center, 2005; U.S. Department of Education, Office of Elementary & Secondary Education, 2004).

Some states generate alignment evidence in the test development process and document the steps that they have taken to ensure linkage (such as

mapping items to standards). Some states have prepared documents and training to link assessment with instruction. But studies of alternate assessment's linkage to general education content standards and of their alignment with alternate content standards are needed (Lewis et al., 1996; Roach & Elliott, 2004). Three of the Southwest Region states have to submit alignment studies for the 2005/06 year, identify any gaps, and outline a plan to fill them (U.S. Department of Education, 2006a).

Four states used, or are using, the Webb methodology for alignment. The Webb alignment model uses content experts to measure the degree of correspondence between state-level standards and assessments (Ananda, 2003). The criteria for measuring linkage are (Wisconsin Center for Educational Research, 2007):

- Categorical concurrence—the extent to which the same, or consistent, content categories appear in standards and assessments.
- Range of knowledge correspondence—whether the span of knowledge expected of students on the basis of a standard corresponds to the span that they need to answer the corresponding assessment items or activities correctly.
- Balance of representation—whether objectives that fall under a specific standard are given relatively equal emphasis on the assessment.
- Depth of knowledge consistency—the extent to which the knowledge elicited from students on the assessment is as complex within the content area as what students are expected to know and do.

Louisiana used the WestEd methodology (WestEd, 2004), which uses the Webb review categories (see table C7 in appendix C).

All Southwest Region states reported that they are just beginning to validate the linkage of their

alternate assessments with instruction. Louisiana, New Mexico, and Texas needed to submit reports on alignment studies and to show how they would fill any gap. Texas had to submit its standards, which it has now completed. Accurate data about the consequences of the assessment system take time to collect, so states such as Louisiana and Texas will not have these data available at least until 2008. New Mexico was expected to have study results in the fall of 2007, but data were not available at the time of writing (D. Farley, Education Consultant, Special Education Bureau at the New Mexico Public Education Department, personal communication, June 27, 2007; C. Wieland, personal communication, 2007).

Representatives from all five states recognize the need for empirical evidence and guidance to demonstrate the relationships among standards, assessment, and instruction. State contacts in Arkansas and New Mexico pointed out that this is an ongoing process. If a content area is added—as Arkansas has added high school science—the process must be repeated. If standards change—as they have in New Mexico and Oklahoma—the linkage must be reexamined.

Four of the five Southwest Region states (all but Oklahoma) participate in the Council of Chief State School Officers Assessing Special Education Students State Collaborative on Assessment and Student Standards, where some of these issues are discussed.

All states can participate in the Inclusive Assessment and Accountability Community of Practice, sponsored by the National Alternate Assessment Center. This community can help states working on their technical adequacy learn about the types of evidence to collect, the questions to ask, and the resources available to support objective linkage studies.

Representatives from all five states recognize a need for empirical evidence to demonstrate the relationships among standards, assessment, and instruction

What technical issues are states facing in developing and implementing reliable and valid alternate assessments?

The No Child Left Behind Act says that the “requirements for high technical quality . . . including validity, reliability . . . and consistency with nationally recognized professional and technical standards apply to alternate assessments as well as to regular State assessments” (Rules and Regulations, 68, p. 68609. Fed. Reg. 236, December 9, 2003). States and researchers are wrestling with how to demonstrate the validity, reliability, and usability of alternate assessment systems. Louisiana, New Mexico, and Texas are working to address this question. Arkansas and Oklahoma will become increasingly engaged in doing so as they add high school science (Arkansas) and redefine their alternate academic standards (Oklahoma; see table C8 in appendix C.)

In the survey and interviews that inform this study, representatives of all five Southwest Region states openly discussed technical challenges (see appendix C). As states conduct new alignment studies and standard setting activities, they are required to demonstrate the alignment or linkage of alternate assessment achievement standards with grade-level content standards and alternate assessments—a challenging task (especially when states allow teachers to select the content standards being measured or the tasks for assessment). Because of changes to content standards, achievement standards, and assessment approaches in the past several years, three states lack continuous, year-to-year trend data that could be used to measure progress and growth. All three are exploring ways to conduct consequential validity studies to explore policies’ positive and negative effects.

Some states are attempting to increase the technical adequacy of their alternate assessments with strategies that go beyond traditional statistical analyses

All five Southwest Region states report combinations of raw scores, transformed scores, and scores based on achievement levels. A few states outside the region (such as Kentucky, North Carolina, Oregon, and South Carolina) have used the assessment triangle (see appendix A) for validating

alternate assessments for students with significant cognitive disabilities (Kleinert, Browder, & Towles-Reeves, 2005; see also Pellegrino et al., 2001). Three states—Louisiana, New Mexico, and Texas—still had to submit evidence on the technical adequacy, alignment, inclusion, and reporting of their full alternate assessment system to federal peer review (U.S. Department of Education, 2006a). Arkansas submitted additional evidence and was given full approval in December 2006 (U.S. Department of Education, 2006a). The information presented in appendix C suggests that peer review categories are the critical consideration in developing alternate assessments that meet the challenges of the No Child Left Behind Act and assessment standards.

A review of technical documents indicates that all states need better documentation of their efforts to demonstrate validity, reliability, and usability. Much work is needed to ensure that alternate assessments reflect adequate psychometric properties and instructional relevance for students with significant cognitive disabilities. And more research is needed to validate the alternate approaches being used for students with significant cognitive disabilities.

Because states in the Southwest Region have not used the same approach or standards for three years in a row, researchers cannot look at data trends to show student progress. Further, three states—Louisiana, New Mexico, and Texas—have fallen short in providing evidence for content, response process, and various types of validity. Contacts in all five Southwest Region states said that since each state needs to continually address questions of validity, reliability, and usability, technical assistance from researchers or advisory committees looking at technical adequacy and reviewing technical manuals would help.

FAR-REACHING QUESTIONS, EXISTING APPROACHES, FURTHER RESEARCH

The states in the Southwest Region are grappling with the same basic technical challenges and

tradeoffs as states in other regions. All states confront two far-reaching questions about alternate assessments:

- Do the advantages of flexibility (tailoring assessments to the wide range of needs of students with significant cognitive disabilities) outweigh the technical challenges of such a nonuniform approach?
- Does the possibility of attaining greater consequential validity through alternate assessments balance the difficulty of obtaining evidence for sufficient technical adequacy (at least as adequacy is traditionally defined)?

These questions are researchable.

Some states are attempting to increase the technical adequacy of their alternate assessments with a range of strategies that go beyond traditional statistical analyses. Such broad-based approaches supplement technical analysis with careful explanation of the assessments' purposes and procedures, mandatory uniform training for all administrators and scorers, and active monitoring of live administrations (McGregor, 2007). Another promising approach involves combining data across several years to increase sample sizes and better measure reliability and validity using existing methodologies.

Several states have joined an assessment consortium to increase available resources, expertise, and sample sizes for assessments of English language proficiency. Such collaboration requires upfront planning and—possibly—compromising on an assessment's content and the range of purposes for which it may be used. Otherwise, the test may not be valid (aligned and accessible) for each state in the consortium.

Several questions merit further research attention:

- What are the characteristics of students with significant cognitive disabilities, and is it possible to validate the characteristics and

criteria for their placement? Much current research points to a need to better identify who these students are—to prevent states from overidentifying them for the alternate assessment, but also to enable states to develop and tailor alternate assessments to them.

- Can appropriate guidelines be developed to justify IEP team decisions on alternate assessment? More data should be collected on the types of guidelines that the teams must follow in determining whether a student qualifies. Since guidelines and criteria vary from state to state, research is needed to get a better sense of which students are being served, the numbers being served, and the efficacy of eligibility requirements.

For some of these questions the answers might ultimately lie in the policy and values arenas rather than in the technical realm

- Should assessment approaches (portfolio and performance based) be redefined for the Southwest Region states? For states across the country? Although states have had to use some form of alternate assessment for seven years, comprehensive research into the relationships between assessment formats and student outcomes remains sparse.
- How expert are teachers in linking functional skills to content standards? Some researchers have asserted that once teachers are well versed in this area, more checks and balances will be added to the accountability picture (Browder et al., 2004). More must be learned about the alignment of functional skills and content standards and teacher training, as very little is known about teachers' perceived skill levels in this area.
- How is the impact of alternate assessment policies and practices best measured? Because of national variations, more research is needed here.

- Are states validly, reliably, and accurately measuring student performance?
- Are teachers increasingly part of long-term professional development programs that improve instruction in reading? In math? In science? Some states already provide training in these areas.
- Are students increasingly meeting the standards because of focused instruction in reading? In math? In science?
- Is the assessment triangle usable and beneficial for alternate assessments? What is its efficacy?
- Do alternate assessments reflect robust psychometric properties, instructional relevance, and technical adequacy? Exploring these three areas is a longstanding challenge. Each area needs to be examined separately to meet federal guidelines and to improve alternate assessments and outcomes for students with significant cognitive disabilities.

For some of these questions the answers might ultimately lie in the policy and values arenas rather than in the technical realm—though new approaches to validation must be developed and studied.

APPENDIX A

GLOSSARY

1 percent rule. When measuring adequate yearly progress, states and school districts have the flexibility to count the “proficient” and “advanced” scores of students with the most significant cognitive disabilities who take alternate assessments based on alternate achievement standards—as long as the number of proficient and advanced scores so counted does not exceed 1 percent of all students in the grades tested (about 9 percent of students with disabilities). The 1 percent cap is based on current incidence rates of students with the most significant cognitive disabilities, allowing for reasonable local variation in prevalence. (U.S. Department of Education, 2003a)

2 percent rule. Under the final regulations on modified academic achievement standards, when measuring adequate yearly progress, states and local education agencies have the flexibility to count the “proficient” and “advanced” scores of certain students—a small group who are identified as disabled and take alternate assessments based on modified academic achievement standards, but are not identified as having the most significant cognitive disabilities—so long as the number of proficient and advanced scores so counted does not exceed 2 percent of all students in the grades assessed (or about 20 percent of students with disabilities). The 2 percent cap, in conjunction with the requirements for state guidelines, is meant to discourage the inappropriate assessment of students based on modified academic achievement standards. (U.S. Department of Education, 2007)

Accommodations. A change in the administration of an assessment (such as setting, scheduling, timing, presentation format, response mode, or others, including any combination of these) that does not change the construct intended to be measured by the assessment or the meaning of the resulting scores. Accommodations are used for equity, not advantage, and serve to level the playing field. To be appropriate, assessment accommodations must be identified in the student’s individualized

education program or an accommodation plan under Section 504 of the Rehabilitation Act of 1973 and used regularly during instruction and classroom assessment. (Policy to Practice Study Group, 2003)

Accountability. The use of assessment results and other data to ensure that schools are moving in desired directions. Common elements include standards, indicators of progress toward meeting those standards, data analysis, reporting procedures, and rewards or sanctions. (Policy to Practice Study Group, 2003)

Adaptations. A general term that describes a change in the presentation, setting, response, timing, or scheduling of an assessment that may or may not change the construct of the assessment. (Policy to Practice Study Group, 2003)

Adequate yearly progress. A provision of the federal No Child Left Behind legislation requiring schools, districts, and states to demonstrate, on the basis of test scores, that students are making academic progress. Each state was required to submit by January 31, 2003, a specific plan for monitoring adequate yearly progress. (Policy to Practice Study Group, 2003)

Alignment. Alignment refers to the degree to which the content (such as skills and concepts) in two sets of standards or in an assessment and set of standards concurs. Alignment relationships tend to be direct relationships (skill and content matches) and are typically observed between standards and assessments for a single student population (such as general education, special education, or English language learners). (WestEd, 2004)

Alternate assessment. An instrument used to gather information on the standards-based performance and progress of certain students, such as those whose disabilities preclude their valid and reliable participation in general assessments. Alternate assessments measure the performance of a relatively small population of students who are unable to participate in the general assessment

system, with or without accommodations, as determined by the individualized education program team. (Policy to Practice Study Group, 2003)

Alternate achievement standard. An alternate achievement standard is an expectation of performance that differs in complexity from a general achievement standard. Alternate achievement standards must be aligned with a state's academic content standards, promote access to the general curriculum, and reflect professional judgment of the highest achievement standards possible (see No Child Left Behind Act of 2001, §200.1(d)). These standards will be considered during the peer review of each state's standards and assessment system under the No Child Left Behind Act. (U.S. Department of Education, 2003a)

Assessment triangle. An assessment framework, based on the premise that three foundational elements—cognition, observation, and interpretation—influence all functions of an assessment's design and use and must work in synchrony to be effective (Pellegrino et al., 2001).

The framework triangulates the balance and evidence needed among three areas: student cognition (what do we know about how students learn?), observation (measurement, or how do we create situations that allow students to demonstrate what they have learned?), and interpretation (how do we draw inferences from the performance?). For instance, instead of raw scores, transformed scores, such as scaled scores (see "Scores" below), should be used for interpretation and decisionmaking (American Educational Research Association et al., 1999). The performance of students receiving transformed scores can be compared with that of other students, using percentiles for example.

The assessment triangle does not allow measurement of differentiated learning, a key concept in large-scale testing.

Bias (test bias). In a statistical context bias is a systematic error in a test score. In discussing test fairness, bias is created by not allowing certain groups

into the sample, not designing the test to allow all groups to participate equitably, selecting discriminatory material, testing content that has not been taught, and so on. Bias usually favors one group of test takers over another, resulting in discrimination. (Policy to Practice Study Group, 2003)

Body of evidence. Information or data establishing that a student can perform a particular skill or has mastered a specific content standard. The information or data were either produced by the student or collected by someone knowledgeable of the student. (Policy to Practice Study Group, 2003)

Bookmark. An approach to standard setting where reviewers establish cut scores for specified levels of proficiency. (Olsen et al., 2002)

Checklist/rating scale. In this alternate assessment approach teachers evaluate whether students can perform certain behaviors or have mastered certain skills. Scoring is based on the number of skills the student is able to perform successfully.

Cutscore. A specified point on a score scale. Scores at or above that point are interpreted differently from scores below that point. (Policy to Practice Study Group, 2003)

Functional academics. Cognitive abilities and skills learned at school. The school subjects that directly apply to and teach the skills needed in one's everyday environment. The idea behind functional academics is to implement and teach academic skills in a way that students can generalize from one setting to the next, outside the school context. (Turnbull, Turnbull, Shank, Smith, & Leal, 2002)

Functional skills. Functional skills are daily living skills that provide the essential knowledge, skills, and understanding to operate successfully, effectively, and independently in life and at work. The premise behind functional skills is that they allow an individual with a disability more access to opportunities for participation in the community. (Turnbull et al., 2002)

Individualized education program. A document that reflects the decisions made by an interdisciplinary team, including the parent and the student when appropriate. During an individualized education program meeting for a student with a disability, the individualized education program team (IEP team) will identify the student's abilities and disabilities. (Policy to Practice Study Group, 2003)

Large-scale assessment. A test administered simultaneously to large groups of students within the district or state. (Policy to Practice Study Group, 2003)

Linkage. Relationships that tend to be developmental, foundational, or proximal and are typically observed between standards and assessments developed for different populations (such as general education standards and alternate standards). (WestEd, 2004)

Peer review. The use of state experts to review a state's standards and assessment system to determine whether it meets No Child Left Behind requirements.

Performance descriptors. Statements that describe what students at each performance level should know and be able to do.

Performance levels. Performance levels provide a determination of the extent to which a student has met the content standards. They distinguish a proficient or adequate performance from a novice or expert performance. (Policy to Practice Study Group, 2003)

Portfolio. A collection of student-generated or student-focused evidence that provides the basis for demonstrating the student's mastery of a range of skills, performance level, or improvement in these skills over time. The portfolio evidence may include student work samples, photographs, videotapes, interviews, anecdotal records, interviews, and observations. (Policy to Practice Study Group, 2003)

Prompt. A picture or text (for example, a word) to stimulate a response to an item on an assessment. (Salvia & Ysseldyke, 2001)

Reliability. The consistency of the test instrument; the extent to which it is possible to generalize a specific behavior observed at a specific time by a specific person to observations of similar behavior at different times or by different behaviors. (Policy to Practice Study Group, 2003)

Rubric. A scoring tool based on criteria used to evaluate a student's test performance. The criteria contain a description of the requirements for varying degrees of success in responding to the question or performing the task. Rubrics can be diagnostic or analytic (providing ratings of multiple criteria), or they can be holistic (describing a single global trait). (Policy to Practice Study Group, 2003)

Scores (raw and scale). Raw scores are traditionally converted to scale scores (such as 200–600, with 400 the mean average) for various purposes. Using scale scores can support a testing program's validity, simplify reporting, and compensate for variation in task or item difficulty and each item's weighted importance during scoring. By using raw scores, testing programs limit their ability to compare student performance or show individual growth from one testing window to the next. Scale scores are often linked to cutscores (set during standard setting) and performance descriptors. Qualitative descriptions can be attached readily to ranges of scores to enhance interpretation. (American Educational Research Association et al., 1999)

Standard setting. Determining appropriate cutscores that correspond to specified levels of performance (such as below basic, basic, proficient, and advanced). In addition, during standard setting, descriptors are written indicating what students at each performance level should know and be able to do. Standard setting has an important relationship to instruction because this information—which accompanies assessment

results—helps inform instructional planning for each student. (Roach & Elliott, 2004)

Standardized. An established procedure that ensures that a test is administered with the same directions, under the same conditions (time limits and so on), and is scored in the same manner for all students to ensure reliable and valid comparison among students taking the test. (Policy to Practice Study Group, 2003)

Technical adequacy. The extent to which an assessment meets the requirements for validity, reliability, accessibility, objectivity, and consistency with nationally recognized professional and technical standards. Evidence for technical adequacy can include information on administration, scoring, interpretation, and technical data. (U.S. Department of Education, 2007)

Technical assistance. Technical assistance services are timely, specialized guidance and customized supports that help states, districts, schools, and educators solve specific problems, and increase their capacity to improve student learning. Technical Assistance can be short- or long-term (Delaware Department of Education, 2008).

Validity. The extent to which a test measures what it was designed to measure (Policy to Practice Study Group, 2003). Common types of validity include:

- *Construct validity.* The extent to which the characteristic to be measured relates to test scores measuring the behavior in situations in which the construct is thought to be an important variable.
 - *Content validity.* The extent to which the stimulus materials or situations composing the test call for a range of responses that represent the entire domain of skills, understandings, or behaviors that the test is intended to measure.
 - *Convergent validity.* The extent to which the assessment results positively correlate with the results of other measures designed to assess the same or similar constructs.
 - *Criterion-related validity.* The extent to which test scores of a group or subgroup are compared with other criterion measures (ratings, classifications, other tests) assigned to the examinees.
 - *Face validity.* A concept based on a judgment about how relevant the test items appear to be; it relates more to what a test appears to measure than to what the test actually measures.
- Work sample.* Work samples, as found in *portfolios*, are examples of student work collected over time. To allow scoring after a portfolio is completed the results of work samples must be stored as artifacts. Examples of artifacts in portfolios are photographs or videotapes of the student performing a task (such as placing pictures in sequential order), audiotapes (such as a student reading), writing samples, drawings, and tests.

APPENDIX B

SOUTHWEST REGION STATE

DEMOGRAPHIC INFORMATION

TABLE B1

State demographic statistics

Statistic	Arkansas	Louisiana	New Mexico	Oklahoma	Texas
	Arkansas Department of Education	Louisiana Department of Education	New Mexico Public Education Department	Oklahoma State Department of Education	Texas Education Agency
K–12 student enrollment	463,115	724,281	326,102	629,476	4,405,215
Number of school districts	318	68	89	541	1,039
Approximate number of students taking alternate assessment	3,700	4,500	2,200	3,100	40,000
Number of public schools	1,158	1,541	842	1,787	8,746
Elementary schools	600	831	449	1,020	4,224
Middle schools	201	245	152	294	1,576
High schools	311	303	119	483	1,687
Multilevel schools	—	156	—	—	469
Alternative schools	5	36	27	—	714
Career/tech schools	—	—	—	54	—
Charter schools	17	17	44	12	321
Student race/ethnicity (percent)					
Asian	1	1	1	2	3
Black	23	48	3	11	14
Hispanic	6	2	53	8	45
American Indian	1	1	11	19	<1
White	69	48	32	61	38
Students receiving free or reduced-price lunch (percent)	52	62	58	54	48
Students receiving special education services (percent)	12	14	20	15	12
Students receiving English language learner services (percent)	4	2	19	7	16
Graduation rate (percent)	81	83	84	86	84

— is not available.

Source: State report cards for Arkansas (<http://normessasweb.uark.edu/reportcards/state05.php>), Louisiana (http://www.doe.state.la.us/lde/pair/StateReport0405/10-Student_Achievement.pdf), New Mexico (<http://www.ped.state.nm.us/div/acc.assess/accountability/dlRptCard2005/NMStateReportCard%20English.pdf>), Oklahoma (<http://title3.sde.state.ok.us/studentassessment/2005results/reportcard2005state.pdf>), and Texas (http://www.tea.state.tx.us/research/pdfs/2005_comp_annual.pdf); state department of education contacts (for Arkansas, Charlotte Marvel, Math and Assessment Specialist, Arkansas Department of Education; for Louisiana, Jeanne Johnson, Education Consultant, Louisiana Department of Education; for New Mexico, Dan Farley, Education Consultant, Special Education Bureau at the New Mexico Public Education Department; for Oklahoma, Amy Daugherty, Coordinator, Compliance Activities and Assessment, Special Education Department Oklahoma Department of Education; for Texas, Cari Wieland, Director, Special Education Assessments, Texas Education Agency).

APPENDIX C

SIDE-BY-SIDE COMPARISON OF SOUTHWEST REGION STATES' ALTERNATE ASSESSMENT POLICIES AND PRACTICES

TABLE C1

State assessments by grade and subject

	Arkansas	Louisiana
Comprehensive No Child Left Behind Title I assessment system	<p>State tests:</p> <ul style="list-style-type: none"> Arkansas Comprehensive Testing and Accountability Program (ACTAAP). Benchmark exams (criterion-referenced tests, grades 3–8). End-of-course exams (algebra, geometry, literacy). Iowa Tests of Basic Skills (ITBS; grades 3–8) and Iowa Test of Education Development (ITED) at grade 9. Arkansas Alternate Portfolio Assessment System (AAPAS) for students with significant cognitive disabilities. 	<p>Comprehensive assessment system:</p> <ul style="list-style-type: none"> iLeap (norm-referenced test augmented criterion-referenced test in grades 3, 5–7, 9). Louisiana Educational Assessment Program (LEAP; criterion-referenced test in grades 4 and 8). Graduation exit exam (GEE). Louisiana alternate assessments (LAA 1 and LAA 2).
Alternate assessment system for 2007/08	Arkansas Alternate Portfolio Assessment System (AAPAS)	LEAP Alternate Assessment (LAA 1)
Grades and subjects tested	3–8, 9 (for math), and 11 English language arts, math, science	3–8, and 10 English language arts, math, science, social studies

New Mexico	Oklahoma	Texas
<p>New Mexico Achievement Assessment Program (NMAAP):</p> <ul style="list-style-type: none"> • New Mexico Standards Based Assessment (NMSBA, grade 3–9 and 11). • New Mexico High School Competency Exam (NMHSCE). • New Mexico English Language Proficiency Assessment (NMELPA). • New Mexico Alternate Performance Assessment (NMAPA). 	<p>Oklahoma School Testing Program (OSTP) based on core curriculum (Priority Academic Student Skills—PASS):</p> <ul style="list-style-type: none"> • Oklahoma Core Curriculum Tests (OCCT, grades 3–8). • End-of-course tests (algebra, geometry, literacy). 	<p>Texas Assessment Program (TAP):</p> <ul style="list-style-type: none"> • Texas Assessment of Knowledge and Skills (TAKS and TAKS–ALT; grades 3–12 in English and language arts, math, science, and social studies). • State-developed Alternative Assessment (SDAA II), the Texas English Language Proficiency Assessment System (TELPAS). • Texas Assessment of Academic Skills (TAAS, being phased out as exit exam).
<p>New Mexico Alternate Performance Assessment (NMAPA)</p>	<p>Oklahoma Alternate Assessment Program (OAAP)</p>	<p>Texas Assessment of Knowledge and Skills–Alternate (TAKS–ALT)</p>
<p>3–10</p> <p>English language arts, math, science, writing</p>	<p>3–8 and 10–12</p> <p>reading, writing, math, science, social studies</p>	<p>3–9 reading</p> <p>3–10 math</p> <p>4 and 7 writing</p> <p>10 English language arts</p>

Source: Arkansas Department of Education web sites (<http://www.arkansased.org/students/assessment.html>; arkedu.state.ar.us/actaap/index.htm; http://www.arkedu.state.ar.us/actaap/student_assessment/student_assessment_p1.htm) and survey and interviews with department contacts (C. Marvel and T. Hicks); Louisiana Department of Education web sites (<http://www.doe.state.la.us/lde/saa/2273.html>; <http://www.doe.state.la.us/lde/accountability/home.html>; www.doe.state.la.us/lde/saa/2343.html) and survey and interviews with department contact (J. Johnson); New Mexico Public Education Department web sites (<http://legis.state.nm.us>; <http://www.ped.state.nm.us/div/acc.assess/assess/info.update.corner.html>; <http://www.ped.state.nm.us/div/acc.assess/assess/index.html>) and survey and interviews with department contact (D. Farley); Oklahoma State Department of Education web sites (<http://www.sde.state.ok.us/home/defaultie.html>; <http://www.lsb.state.ok.us>; title3.sde.state.ok.us/studentassessment) and survey and interviews with department contact (A. Daugherty); Texas Education Agency web sites (<http://www.tea.state.tx.us/student.assessment>; <http://www.tea.state.tx.us/curriculum.html>; <http://www.legis.state.tx.us>; <http://www.sos.state.tx.us/tac/index.shtml>; <http://www.tea.state.tx.us/student.assessment/resources/taksalt/index.html>) and survey and interviews with agency contact (C. Wieland).

TABLE C2

Southwest Region state laws, regulations, rules, and administrative code for alternate assessment

Arkansas	Louisiana
<p>Arkansas Department of Education Rules and Regulations Act 999 of 1999 established the Arkansas Comprehensive Testing and Accountability Program.</p> <p>Arkansas Code 6-41-217 governs individualized education programs, alternate assessment options b (3) (b) (3) (A) iii.</p> <p>Board rules and regulations: Arkansas ADC 005 19 006 5.00–5.02, 5.02.5, 5.04–5.08.</p> <p>Arkansas Department of Education has adapted 34 CFR 300.138, and</p> <ul style="list-style-type: none"> • Has developed guidelines for the participation of children with disabilities in alternate assessments for children who cannot participate in state and districtwide assessment programs. • Has developed an alternate assessment system consisting of a portfolio assessment methodology, in accordance with the above guidelines, which was field tested during the spring semester of the 1999/2000 school year. 	<p>Louisiana Administrative Code, Bulletin 118 establishes a comprehensive assessment system:</p> <p>LEAP Alternate Assessment, Level 1 (LAA 1) has been specially designed to evaluate the progress of students with significant disabilities.</p> <p>R.S.17:24(F)(4) mandates the assessment of all students in Louisiana public schools.</p>

New Mexico	Oklahoma	Texas
<p>Per Title 6, chapter 31, part 2, §§6.31.2.11(E)(3)(a)-(c) of the New Mexico Administrative Code, students with disabilities for whom alternate assessments are appropriate under the department's established participation criteria may participate in alternate assessments; the individualized education program team must agree and document that the student is eligible for participation in an alternate assessment based on alternate achievement standards according to 34 CFR §300.320(a)(6).</p>	<p>Oklahoma Stat. tit. 70 Sec. 1210-10-508, Okla. Admin. Code §210:10-13-2</p> <p>Board policy (210:10-13-2) requires that districts include all students in state assessments, with appropriate accommodations when necessary. Alternate assessments are offered to students with the most significant cognitive disabilities.</p>	<p>Section 39.023 of the Texas Education Code was amended by the 75th Texas Legislature to address the assessment of students receiving special education services:</p> <p>Texas Education Code, subtitle H, chapter 39, subchapter B.</p>

Source: Arkansas Department of Education web sites (<http://www.arkansased.org/students/assessment.html>; <http://www.arkedu.state.ar.us/actaap/index.htm>; http://www.arkedu.state.ar.us/actaap/student_assessment/student_assessment_p1.htm) and survey and interviews with department contacts (C. Marvel and T. Hicks); Louisiana Department of Education web sites (<http://www.doe.state.la.us/lde/saa/2273.html>; <http://www.doe.state.la.us/lde/accountability/home.html>; <http://www.doe.state.la.us/lde/saa/2343.html>) and survey and interviews with department contact (J. Johnson); New Mexico Public Education Department web sites (<http://legis.state.nm.us>; <http://www.ped.state.nm.us/div/acc.assess/assess/info.update.corner.html>; <http://www.ped.state.nm.us/div/acc.assess/assess/index.html>) and survey and interviews with department contact (D. Farley); Oklahoma State Department of Education web sites (<http://www.sde.state.ok.us/home/defaultie.html>; <http://www.lsb.state.ok.us>; title3.sde.state.ok.us/studentassessment) and survey and interviews with department contact (A. Daugherty); Texas Education Agency web sites (<http://www.tea.state.tx.us/student.assessment>; <http://www.tea.state.tx.us/curriculum.html>; <http://www.legis.state.tx.us>; <http://www.sos.state.tx.us/tac/index.shtml>; <http://www.tea.state.tx.us/student.assessment/resources/taksalt/index.html>) and survey and interviews with agency contact (C. Wieland).

TABLE C3

Southwest Region state definition of significant cognitive disability

Arkansas	Louisiana
<p>The student's demonstrated cognitive functioning and adaptive behavior in the home, school, and community environments are significantly below age expectations, even with program modifications, adaptations, and accommodations.</p> <p>The student's course of study is primarily functional and life-skills oriented.</p> <p>The student requires extensive direct instruction or extensive supports in multiple settings to acquire, maintain, and generalize academic and functional skills necessary for application in school, work, home, and community environments.</p> <p>The student demonstrates severe and complex disabilities and poor adaptive skill levels (determined to be significantly below age expectations by that student's comprehensive assessment) that essentially prevent the student from meaningful participation in the standard academic core curriculum or achievement of the academic content standards established at grade level.</p> <p>The student's disability causes dependence on others for many, if not all, daily living needs, and the student is expected to require extensive ongoing support in adulthood.</p> <p>The student's inability to complete the standard academic curriculum at grade level is not primarily the result of the following:</p> <ul style="list-style-type: none"> • Excessive or extended absences, poor attendance, or lack of instruction. • Sensory (visual or auditory) or physical disabilities, emotional-behavioral disabilities, or a specific learning disability. • Social, cultural, linguistic, or economic differences. • Below average reading level. • Low achievement in general. • Expectations of poor performance. • Disruptive behavior. • The student's IQ. • The anticipated impact of the student's performance on school or district performance scores. • The student's disability category, educational placement, type of instruction, or amount of time receiving special education services. 	<p>To be eligible for participation in LEAP Alternate Assessment, the student shall:</p> <ol style="list-style-type: none"> 1. have a current multidisciplinary evaluation of the following exceptionalities: <ul style="list-style-type: none"> • Moderate mental disability. • Severe mental disability. • Profound mental disability. <p>Or have an assessed level of intellectual functioning and adaptive behavior three or more standard deviations below the mean and the following exceptionalities:</p> <ul style="list-style-type: none"> • Multiple disabilities. • Traumatic brain injury. • Autism.

New Mexico	Oklahoma	Texas
<p>The working definition of “significant cognitive disability” is supplied by the criteria for participation on a New Mexico alternate assessment:</p> <ul style="list-style-type: none"> • Does the student’s past and present performance in multiple settings (home, school, community) indicate that a significant cognitive disability is present? • Does the student need intensive, pervasive, or extensive levels of support in school, home, and community settings? • Do the student’s current cognitive and adaptive skills and performance levels require direct instruction to accomplish the acquisition, maintenance, and generalization of skills in multiple settings (home, school, community)? 	<p>A student is defined as having a significant cognitive disability through the completion of the Alternate Assessment Participation Checklist.</p>	<p>The determination of significant cognitive disability is made by the Admission, Review, and Dismissal (ARD) Committee based on state education agency guidelines (see table C4).</p>

Source: Thompson et al., 2005.

TABLE C4

Southwest Region state education agency guidelines for participation in alternate assessment

Arkansas	Louisiana
<p>Participation in the alternate assessments is determined by:</p> <ul style="list-style-type: none"> The student's individual education program (IEP) team, as documented in the student's IEP program. <p>Or</p> <ul style="list-style-type: none"> The IEP team determines whether participation in the standard state assessment program is appropriate for students with IEPs. Students with disabilities for whom it is deemed inappropriate to take the standard state assessments (benchmarks and end-of-course) with the established accommodations participate in the Arkansas Alternate Portfolio Assessment System (AAPAS) following the guidelines established by the board. 	<p>Students for whom the general statewide assessment is not appropriate. The Louisiana Educational Assessment Program (LEAP) Alternate Assessment, Level 1, is designed for students whose IEPs reflect significant modifications of the general education curriculum and have an emphasis on functional academic and life skills. A student participating in LEAP Alternate Assessment, Level 1, is progressing toward a certificate of achievement rather than a high school diploma.</p>

New Mexico	Oklahoma	Texas
<p>The participation criteria for the New Mexico Alternate Assessments have become rule in the state of New Mexico. Sections 6.31.2.11(E)(3)(a)-(c) of the New Mexico Administrative Code now require that IEP teams agree and document that the student is eligible for participation in an alternate assessment according to the following criteria:</p> <ul style="list-style-type: none"> • The student’s past and present levels of performance in multiple settings (home, school, community) indicate that a significant cognitive disability is present. • The student needs intensive, pervasive, or extensive levels of support in school, home, and community settings. • The student’s current cognitive and adaptive skills and performance levels require direct instruction to accomplish the acquisition, maintenance, and generalization of skills in multiple settings (home, school, community). 	<ul style="list-style-type: none"> • Does the student’s disability result in substantial academic difficulties? • Is the student’s difficulty with general curriculum demands primarily due to the student’s disability and not due to excessive absences unrelated to the disability, or social, cultural, environmental, or economic factors? • Does the student’s IEP reflect curriculum and daily instruction that focus on modified or alternate standards? • Does the student have a significant cognitive disability? • Does the student’s demonstrated cognitive ability and adaptive behavior require substantial adjustments (alternate achievement standards) to the general education curriculum? • Do the student’s learning objectives and expected outcomes focus on functional application of skills as illustrated in the student’s IEP goals, benchmarks, and objectives? • Does the student require direct and extensive instruction to acquire, maintain, generalize, and transfer new knowledge and skills? 	<p>The Admission Review and Dismissal Process (ARD) Committee may decide if the student can take Texas Assessment of Knowledge and Skills–Alternate (TAKS–ALT), if the student:</p> <ul style="list-style-type: none"> • Requires support to access the general curriculum. Support may include assistance involving communication, response style, physical access, or daily living skills. • Requires direct, intensive, individualized instruction in a variety of settings to accomplish the acquisition, maintenance and generalization of skills. • Participates in the grade-level Texas Essential Knowledge and Skills (TEKS) through activities that focus on prerequisite skills. • Demonstrates knowledge and skills routinely in class by methods other than paper and pencil tasks. • Demonstrates performance objectives that may include real-life applications of the grade-level TEKS as appropriate to the student’s abilities and needs.

Source: Arkansas Department of Education web sites (<http://www.arkansased.org/students/assessment.html>; <http://www.arkedu.state.ar.us/actaap/index.htm>; http://www.arkedu.state.ar.us/actaap/student_assessment/student_assessment_p1.htm) and survey and interviews with department contacts (C. Marvel and T. Hicks); Louisiana Department of Education web sites (<http://www.doe.state.la.us/lde/saa/2273.html>; <http://www.doe.state.la.us/lde/accountability/home.html>; <http://www.doe.state.la.us/lde/saa/2343.html>) and survey and interviews with department contact (J. Johnson); New Mexico Public Education Department web sites (<http://legis.state.nm.us>; <http://www.ped.state.nm.us/div/acc.assess/assess/info.update.corner.html>; <http://www.ped.state.nm.us/div/acc.assess/assess/index.html>) and survey and interviews with department contact (D. Farley); Oklahoma State Department of Education web sites (<http://www.sde.state.ok.us/home/defaultie.html>; <http://www.lsb.state.ok.us>; title3.sde.state.ok.us/studentassessment) and survey and interviews with department contact (A. Daugherty); Texas Education Agency web sites (<http://www.tea.state.tx.us/student.assessment>; <http://www.tea.state.tx.us/curriculum.html>; <http://www.legis.state.tx.us>; <http://www.sos.state.tx.us/tac/index.shtml>; <http://www.tea.state.tx.us/student.assessment/resources/taksalt/index.html>) and survey and interviews with agency contact (C. Wieland).

TABLE C5

Assessment approaches and tasks

	Arkansas	Louisiana	New Mexico
Approach to alternate assessment	Portfolio	Performance tasks	Performance tasks
Description of tasks	<p>The state selects the standards and student learning expectations to be measured, and teachers choose specific tasks.</p> <p>The portfolio can consist of three types of evidence related to the student learning expectations. Each piece of evidence should show students' performance on specific tasks that indicate progress in the general curriculum. Evidence can include: work sample or permanent products, captioned photographs, and videotape with a brief script that provides an objective and clear measure of what the student can perform.</p>	<p>The state specifies target indicators, and teachers select activities.</p> <p>Louisiana Alternate Assessment (LAA 1) enables test administrators to assess students while they are engaged in their daily activities. For purposes of LAA 1, "activities" are defined as organized educational procedures designed to stimulate performance of the skills that will be assessed. Examples of activities include lunchtime, field experiences (such as a trip to a store or museum), or a math lesson.</p> <p>Target indicators from different content areas can be assessed during one activity.</p>	<p>The state defines performance events of tasks to be used.</p> <p>The New Mexico Alternate Assessments are similarly constructed in terms of test design. They are on-demand assessments, meaning that the student has to perform the required elements at one particular point in time or during an event specifically developed for the purpose of administering the test. It is a direct observation assessment, which means that the test administrator is involved solely in observation and scoring the student's behaviors against the performance indicators as the student completes the activities that compose the assessment. The IEP team designs the activities, thus they are structured events. The checklist is the portion of the assessment on which the test administrator scores each indicator, with a range of 0–4, based upon the demonstrated ability. Once a behavior is observed and scored, it cannot be revisited. If the administrator is unable to observe a particular indicator being demonstrated, activities or portions of activities can be repeated to directly target that particular indicator.</p>

Oklahoma	Texas
Portfolio	Hybrid portfolio (portfolio and checklist)
<p>The state mandates two areas per grade per content area, and teachers select three areas to be measured.</p>	<p>The state mandates three essence statements to be measured, and teachers choose two. Teachers decide on activities to measure all five tasks.</p>
<p>The Oklahoma Alternate Assessment Program (OAAP) is a performance-driven assessment. It is based on the Extended Academic Standards, which consist of specific domains, outcomes, standards, and benchmarks extended from Priority Academic Student Skills (PASS). Grades 3–8, and 10 outline the required subject areas to be assessed for the 2005/06 school year. The subject areas differ from grade to grade in order to meet No Child Left Behind requirements and Oklahoma state law governing the Oklahoma School Testing Program. Each subject area must have five pieces of evidence with support described.</p>	<p>Teachers use the state resources to develop assessment activities for students that reflect the instruction they have received on prerequisite skills linked to grade-level expectations. Students being assessed with TAKS–ALT can have whatever accommodations or supports the teacher feels are necessary for the student to be as independent as possible during the activity. The state provides suggestions and hints on how to develop a good assessment.</p>
<p>The teacher should answer and document the following questions about student tasks:</p>	
<ul style="list-style-type: none"> • What is the student’s performance or functional level? • What subjects or tasks can the student do with little or no difficulty? • Is the student mobile (ambulatory or nonambulatory)? • What is the student’s communication ability and communication system used? • What is the student’s attendance history? • What are the student’s learning strengths (visual, auditory, tactile)? • What modifications, accommodations (including assistive devices and technology) and supports are provided? • What behavior interventions and positive behavior supports are used with this student? • What general supports and prompt hierarchy does the student need throughout the day? • How does the student interact with others in the school environment (natural supports)? • What IEP objectives or benchmarks address functional academics skills? • How does the student participate in different school environments? 	
<p>Points are awarded for each answer up to a total of 25.</p>	

Source: Arkansas Department of Education web sites (<http://www.arkansased.org/students/assessment.html>; <http://www.arkedu.state.ar.us/actaap/index.htm>; http://www.arkedu.state.ar.us/actaap/student_assessment/student_assessment_p1.htm) and survey and interviews with department contacts (C. Marvel and T. Hicks); Louisiana Department of Education web sites (<http://www.doe.state.la.us/lde/saa/2343.html>; <http://www.doe.state.la.us/lde/accountability/home.html>; <http://www.doe.state.la.us/lde/saa/2343.html>) and survey and interviews with department contact (J. Johnson); New Mexico Public Education Department web sites (<http://legis.state.nm.us>; <http://www.ped.state.nm.us/div/acc.assess/assess/info.update.corner.html>; <http://www.ped.state.nm.us/div/acc.assess/assess/index.html>) and survey and interviews with department contact (D. Farley); Oklahoma State Department of Education web sites (<http://www.sde.state.ok.us/home/defaultie.html>; <http://www.lsb.state.ok.us/title3.sde.state.ok.us/studentassessment>) and survey and interviews with department contact (A. Daugherty); Texas Education Agency web sites (<http://www.tea.state.tx.us/student.assessment>; <http://www.tea.state.tx.us/curriculum.html>; <http://www.legis.state.tx.us>; <http://www.sos.state.tx.us/tac/index.shtml>; <http://www.tea.state.tx.us/student.assessment/resources/taksalt/index.html>) and survey and interviews with agency contact (C. Wieland).

TABLE C6

Southwest Region state-provided training on administration and use of results

Arkansas	Louisiana	New Mexico
<p>Intensive regional training.</p> <p>State and regional staff provide training to district and school staff.</p>	<p>Regional training.</p> <p>Louisiana State Department of Education provides training on administration and use of results.</p>	<p>State staff and University of New Mexico professors provide training on alternate assessment. Training is a multiday model, including videotaped administrations in which the participants score case studies.</p> <p>Includes three training modules:</p> <ol style="list-style-type: none"> 1. Overview. 2. Format and overview of clusters (what is measured). 3. Scoring. <p>Web-based training is also provided.</p> <p>The state and University of New Mexico have, in previous years, provided training on instruction for the severely disabled.</p>

Oklahoma	Texas
<p>Provided annually in five regions. A technical assistance document is provided annually.</p>	<p>The state and 20 regional centers provide training where modules and presentations are provided that discuss instructional and assessment decisions for students with the most significant cognitive disabilities. Modules are available online at http://pearson.learn.com/taksalt.</p> <p>There are four modules:</p> <ul style="list-style-type: none"> • Topics include defining and explaining Texas Assessment of Knowledge and Skills–Alternate (TAKS–ALT) participation guidelines, defining access to grade-level curriculum, and a step-by-step process to access grade-level content and standards. • Topics include recording anecdotal notes and samples of student work, making fair observations, time management strategies, and effective planning for focused classroom observations. • Topics include TAKS-ALT scoring rubric, rating and expectations of students, evidence or data to be collected for the observation evaluation, and how to document observations. • Topics include descriptions of how to use the actual TAKS-ALT online assessment with system training simulations.

Source: Arkansas Department of Education web sites (<http://www.arkansased.org/students/assessment.html>; [arkedu.state.ar.us/actaap/index.htm](http://www.arkedu.state.ar.us/actaap/index.htm); http://www.arkedu.state.ar.us/actaap/student_assessment/student_assessment_p1.htm) and survey and interviews with department contacts (C. Marvel and T. Hicks); Louisiana Department of Education web sites (<http://www.doe.state.la.us/lde/saa/2273.html>; <http://www.doe.state.la.us/lde/accountability/home.html>; www.doe.state.la.us/lde/saa/2343.html) and survey and interviews with department contact (J. Johnson); New Mexico Public Education Department web sites (<http://legis.state.nm.us>; <http://www.ped.state.nm.us/div/acc.assess/assess/info.update.corner.html>; <http://www.ped.state.nm.us/div/acc.assess/assess/index.html>) and survey and interviews with department contact (D. Farley); Oklahoma State Department of Education web sites (<http://www.sde.state.ok.us/home/defaultie.html>; <http://www.lsb.state.ok.us>; title3.sde.state.ok.us/studentassessment) and survey and interviews with department contact (A. Daugherty); Texas Education Agency web sites (<http://www.tea.state.tx.us/student.assessment>; <http://www.tea.state.tx.us/curriculum.html>; <http://www.legis.state.tx.us>; <http://www.sos.state.tx.us/tac/index.shtml>; <http://www.tea.state.tx.us/student.assessment/resources/taksalt/index.html>) and survey and interviews with agency contact (C. Wieland).

TABLE C7

Alignment and standard setting in the Southwest Region states

Component	Arkansas	Louisiana
Alignment of alternate standards to alternate assessments	Built content standards and achievement standards using horizontal and vertical alignment approach. Then, the assessment was linked to the standards.	Built content standards and achievement standards using horizontal and vertical alignment approach by grade bands. Then, the assessment was linked to the standards.
Overall alignment methodology	Webb	WestEd
Standard setting methodology	Bookmarking (state-declared) reviewers mark the spot in a specially constructed test booklet, arranged in order of item difficulty, where a desired percentage of minimally proficient or advanced students would pass the item, or standard setters mark where the difference between proficient and advanced performance on an exercise is a desired minimum percentage of students; reasoned judgment (a score scale, such as 32 points, is divided into a desired number of categories, such as 4, in some way, such as equally, larger in the middle, and so on; the categories are determined by a group of experts, policymakers, or others); and judgmental policy capturing (reviewers determine which of the various components of an overall assessment are more important than others, so that components or types of evidence are weighted).	Bookmarking (state-declared).
Summary of scoring	<p>State scores.</p> <p>Two readers score each portfolio on a 4-point scale. If the scores are not adjacent (such as 2 and 4), a third reader scores for resolution.</p> <p>Prior to scoring, a range-finding committee meets to establish scoring decisions and pulls exemplar papers in grades 3–8 and 11.</p> <p>Teachers must submit three entries for three English language arts strands, two entries for five strands in math and science, and two entries for three strands in social studies.</p> <p>The rubric is weighted by domain (performance, context, and level of assistance settings). English language arts has 540 total points, math has 600 points, and science has 360 points.</p> <p>Scores for students with disabilities or English language learners participating in the Arkansas Alternate Assessment Program are reported by the state, district, and school in separate reports at all levels.</p>	<p>State scores on scoring rubric.</p> <p>Two test administrators observe and rate the student. The state specifies two target indicators for each participation level. The test administrator selects an additional three target indicators. The rubric has a scale of 1 to 6.</p> <p>Prior to scoring, a range-finding committee meets to establish scoring decisions and pulls exemplar papers.</p> <p>Teachers submit five entries for each strand in each subject area. There were a total of 20 target indicators in 2005/06: five in English language arts, five in math, six in social studies, and four in science.</p> <p>Scaled score ranges were used in the reporting of achievement levels. All content areas had a maximum of 340 scaled scores. Cutscores vary by content area and grade.</p>

New Mexico	Oklahoma	Texas
Built content standards and achievement standards using horizontal and vertical alignment approach. Then, the assessment was linked to the standards.	Built content standards and achievement standards using horizontal and vertical alignment approach. Then, the assessment was linked to the standards.	Built content standards and achievement standards using horizontal and vertical alignment approach. Then, the assessment was linked to the standards.
Webb	Webb	Webb
Body of work (state-declared) reviewers use a student's data to place the student in one of the overall performance levels; standard setters receive a set of papers that demonstrate the complete range of possible scores from low to high and reasoned judgment.	Body of work (state-declared) and reasoned judgment.	Body of work, reasoned judgment, and judgmental policy capturing.
<p>One person scores the performance of each student on a scale of 0–6. Each of four performance tasks in each content area is rated separately.</p> <p>The scoring rubric is summarized below:</p> <p>0 = Unable to perform.</p> <p>1 = Acquisition—student can perform 20 percent of task.</p> <p>2 = Fluency building—student performs 60 percent or more of task with or without minimal assistance or prompting.</p> <p>3 = Maintenance—student performs 80 percent or more of the task, with very minimal assistance or prompting on a regular basis.</p> <p>4 = Generalization—student performs 90 percent or more of task without prompting on a regular basis.</p>	<p>A team of three to four special educators score the portfolio through the state. Each rubric uses a 4-point criterion. Ten percent of the portfolios are scored by a second team to establish inter-rater reliability.</p> <p>Five pieces of evidence with support proof (such as videotapes and work evidence) are required for each content area.</p> <p>The content rubric consists of a possible 100 points:</p> <p>25 points for portfolio content.</p> <p>75 points for the evidence content rubric.</p>	<p>The teacher scores the student portfolio using a rubric of 1–3 for demonstration of skill and 1–3 for level of support. If the student receives at least 2s or higher, then the student can be rated on the generalization of skill aspect of the rubric, receiving a yes (rating of 1) or no (rating of 0).</p> <p>Students are scored on three state-selected essence statements per content area being assessed. The teachers observe the students completing teacher-designed activities that link to the TAKS curriculum. The teacher also selects two additional essence statements and designs activities that are to be used.</p> <p>Raw scores are converted to proficiency levels. The assessment system was field tested in 2006/07. An alignment study and standard setting were scheduled for June 2007.</p>

Source: Arkansas Department of Education web sites (<http://www.arkansased.org/students/assessment.html>; arkedu.state.ar.us/actaap/index.htm; http://www.arkedu.state.ar.us/actaap/student_assessment/student_assessment_p1.htm) and survey and interviews with department contacts (C. Marvel and T. Hicks); Louisiana Department of Education web sites (<http://www.doe.state.la.us/lde/saa/2273.html>; <http://www.doe.state.la.us/lde/accountability/home.html>; www.doe.state.la.us/lde/saa/2343.html) and survey and interviews with department contact (J. Johnson); New Mexico Public Education Department web sites (<http://legis.state.nm.us>; <http://www.ped.state.nm.us/div/acc.assess/assess/info.update.corner.html>; <http://www.ped.state.nm.us/div/acc.assess/assess/index.html>) and survey and interviews with department contact (D. Farley); Oklahoma State Department of Education web sites (<http://www.sde.state.ok.us/home/defaultie.html>; <http://www.lsb.state.ok.us/title3.sde.state.ok.us/studentassessment>) and survey and interviews with department contact (A. Daugherty); Texas Education Agency web sites (<http://www.tea.state.tx.us/student.assessment>; <http://www.tea.state.tx.us/curriculum.html>; <http://www.legis.state.tx.us>; <http://www.sos.state.tx.us/tac/index.shtml>; <http://www.tea.state.tx.us/student.assessment/resources/taksalt/index.html>) and survey and interviews with agency contact (C. Wieland).

TABLE C8

Review procedures and recent changes

Procedure	Arkansas	Louisiana
Special education advisory council/ committees	Meets annually. Reviews policies and practices of alternate assessment and gives general advice on training needs.	Meets quarterly. Reviews policies and practices of alternate assessment and gives general advice on training needs.
Changes in alternate assessment for 2007/08 and beyond	Adding high school science in 2009/10. Alternate content and achievement standards are in progress.	Changing standards from functional to academic. Building a new assessment system.

New Mexico	Oklahoma	Texas
<p>Meets on alternate months with additional meetings as needed.</p> <p>Reviews all aspects of alternate assessment, including items and scoring procedures.</p>	<p>Meets semiannually.</p> <p>Reviews policies and practices of alternate assessment and gives general advice on training needs.</p>	<p>Several separate advisory committees meet, but there is a core group on all committees (such as the testing overview committee meets monthly).</p> <p>Reviews all aspects of alternate assessments, including items.</p>
<p>2005/06 Checklist</p> <p>2006/07 Checklist being phased out and performance standards being phased in. Completed data review and standard setting in June 2007.</p> <p>2007/08 New system operational.</p>	<p>Standard setting. Teachers are shifting from functional skills to grade-level academic skills.</p>	<p>Phasing out local choices and SDAA II. Moving to a portfolio system. In process of changing from functional skills to academic standards.</p>

Source: Arkansas Department of Education web sites (<http://www.arkansased.org/students/assessment.html>; <http://www.arkedu.state.ar.us/actaap/index.htm>; http://www.arkedu.state.ar.us/actaap/student_assessment/student_assessment_p1.htm) and survey and interviews with department contacts (C. Marvel and T. Hicks); Louisiana Department of Education web sites (<http://www.doe.state.la.us/lde/saa/2273.html>; <http://www.doe.state.la.us/lde/accountability/home.html>; <http://www.doe.state.la.us/lde/saa/2343.html>) and survey and interviews with department contact (J. Johnson); New Mexico Public Education Department web sites (<http://legis.state.nm.us>; <http://www.ped.state.nm.us/div/acc.assess/assess/info.update.corner.html>; <http://www.ped.state.nm.us/div/acc.assess/assess/index.html>) and survey and interviews with department contact (D. Farley); Oklahoma State Department of Education web sites (<http://www.sde.state.ok.us/home/defaultie.html>; <http://www.lsb.state.ok.us>; title3.sde.state.ok.us/studentassessment) and survey and interviews with department contact (A. Daugherty); Texas Education Agency web sites (<http://www.tea.state.tx.us/student.assessment>; <http://www.tea.state.tx.us/curriculum.html>; <http://www.legis.state.tx.us>; <http://www.sos.state.tx.us/tac/index.shtml>; <http://www.tea.state.tx.us/student.assessment/resources/taksalt/index.html>) and survey and interviews with agency contact (C. Wieland).

APPENDIX D

STUDY METHODS AND LIMITATIONS

The selection of methods for this study was informed by researchers' experience studying alternate assessments (Sato, Rabinowitz, & Wilkins, 2007), by national studies investigating state policies and practices on alternate assessments (such as those by the National Center on Educational Outcomes; Thompson & Thurlow, 2001, 2003; Thompson et al., 2005), and by studies on alternate assessment issues and practices in one or more states (such as Browder et al., 2002, and Browder et al., 2005; Thompson, Case, & Thurlow, 2000; and Thurlow & Case, 2004).

The researchers investigated the challenges to designing and implementing alternate assessments and developed six questions to guide their study, using different methods to answer each:

1. What challenges are states encountering when implementing new alternate assessment policies and practices? (literature review, publicly available state documents, survey, interview)
2. What do alternate assessments across the Southwest Region states look like? (publicly available state documents, survey, interview)
3. What training or professional development is provided for teachers on alternate assessments? (literature review, publicly available state documents, survey)
4. How are results collected and used at the state, district, school, and student levels? (literature review, publicly available state documents, interview)
5. To what extent do state alternate assessments capture the same or similar skills as state tests designed for the general student population? (literature review, interview)
6. What technical issues are states facing in developing and implementing reliable and

valid alternate assessments? (literature review, survey, interview)

Researchers looked for data that were reliable, valid, and targeted the research questions.

Data collection

Quantitative data collection involved state material audits, surveys, and interviews using descriptive techniques. Qualitative data collection consisted of semistructured surveys and a review of documents, which researchers analyzed with coding to develop themes and categories.

Step one was to collect state-specific materials on alternate assessments, including test administrator manuals; descriptions of laws, regulations, and policies; and descriptions of tasks. Step two was to ask states to send copies of nonsecure materials that could not be located on the web or that were being revised for 2007/08. Step three was to develop and administer a survey for staff in the five state education agencies. The survey was sent to the person in charge of alternate assessment for each state. In some instances states chose to have more than one person respond. Alternate assessment contacts from each state were asked by email whether they would prefer to fill out their responses independently or have a telephone interview. All opted to respond to the questions independently. (See appendix E for the survey instrument.) Step four was to interview representatives from all five states by email or telephone to clarify information about each state's alternate assessment system.

Document analysis

Before beginning the analysis, researchers read the available literature and state documents to refamiliarize themselves with each state's alternate assessment systems. They organized the information into a matrix to compare state policies, practices, and procedures (see appendix C). They compared multiple data sources to

increase reliability. All findings had to be verified by additional primary or secondary sources. For example, to verify the status of peer review, researchers used the Peer Review Status Letters sent to states by the U.S. Department of Education (2006a) and verified the information through interviews, information gained from professional meetings, or survey data.

The study replicated the procedures used by Browder et al. (2005), a mixed-methods approach to research first discussed in relation to this population by Creswell (2002). The approach includes qualitative and quantitative methods that, when systematically combined, provide rigorous, methodologically sound investigations in a range of fields (Creswell, Fetters, & Ivankova, 2004). Triangulation and data transformation models add rigor. Browder (2006) and Ahlgrim-Delzell, Flowers, Browder, and Wakeman (2006) have successfully used the method for more than 20 studies. Other researchers have been able to replicate their work.

Following Browder et al. (2005), researchers began analyzing each state's alternate assessment from an *emic* perspective—that of one who participates in the administration and development of a state's alternate assessment—and retained the language of individual states to describe their alternate assessment systems. The researchers then shifted to an *etic* perspective—a more analytical perspective where researchers compare phenomena—and developed common terminology for describing commonalities and differences (Creswell, 2002). Researchers identified criteria from the literature to analyze specific aspects of alternate assessments.

The method used takes information gleaned from quantitative and qualitative data and merges the information to best understand the research topic.

Study limitations

Whenever evidence is collected and reviewed by a third party, ensuring the reliability of data collection and the validity of findings poses a special challenge. Researchers made every effort to conduct comprehensive searches for print and online information and to verify and clarify their findings with survey and interview methods (Creswell, 2002). Using an iterative process, researchers reviewed print and electronic resources. They clarified and augmented this information with the survey results and by communicating through telephone and email. Despite these efforts, or perhaps because of them, researchers often found inconsistencies between required policies and their understanding and implementation in state practice.

Future studies would benefit from additional data verifying procedures, such as site visits, interviews with local education agency staff, or interviews with a broader group of state education agency staff—sources that were not available for this study—to obtain a more robust picture of states' alternate assessment policies, practices, programs, and needs. Interviews with a formal protocol might be used to collect data more independently, rather than just to clarify survey findings. Full access to state technical reports would have been invaluable. The researchers did not explore in-depth questions about scoring techniques or the alignment of alternate assessment scoring linkages to general assessment.

This study's findings represent only a snapshot of states' alternate assessment policies, practices, and programs at a particular time. Because the findings are based on a small sample of states linked only by geographic proximity, caution is warranted in drawing comparisons across states that are dissimilar in other meaningful ways and in generalizing beyond the sample.

APPENDIX E

ALTERNATE ASSESSMENT INTERVIEW FOR THE SOUTHWEST REGION STATES

Thank you for agreeing to complete these questions. As I explained to you on the phone, this research study is not for accountability purposes, but for your regional laboratory (REL SOUTHWEST/Edvance) to use as a “state of the states” for alternate assessment systems in your region. This information will go to your regional laboratory and will be shared with the states in your region to support or improve alternate assessment policies and practices.

About the study

What: This study focuses on examining state practices and policies for the most severely cognitively disabled student population (i.e., the lowest 1 percent of students with disabilities) across five states in your region.

How: The interview protocol includes 9 questions concerning the alternate assessment policies and practices of your state. It should take about 30 minutes to complete.

Outcomes: Your answers and those of the four other states will be collected and reviewed so that we can gather critical information on:

- a. common/effective elements of state policy related to alternate assessments for the most severely cognitively disabled student population;
- b. critical elements affecting state policy (such as student demographics, teacher qualifications, stakeholder interests, resource availability);
- c. common/promising alternate assessment practices for this population of students;
- d. potential challenges for policy and practice implementation; and
- e. strategies for addressing such challenges.

These outcomes will inform other states in your region that are facing similar influences affecting the implementation of state alternate assessment policy and practice.

Interview questions

We have looked at publicly available information, through the state web site, but need additional information, which we hope to glean through the following questions:

1. What is the format of your state’s alternate assessment? (such as portfolio, checklist, rating scale)
 - a. How did the state decide on this format?
2. How congruent is the description of the intended population of students with disabilities and the actual population assessed?
 - a. Does the local education agency or state education agency monitor this issue?
3. As mentioned, we looked at web sites for information on programs, practices, and products related to alternate assessments:
 - a. Have any polices or practices been added in the past six months?
 - b. Are there other alternate assessment policies or initiatives being considered by the state at this time?
4. What are the particular alternate assessment priorities on which the state is focusing?

For example, priorities might include:

- access to general curriculum
- alignment to curriculum or content standards
- raising expectations

- educator accountability
- student demographics
- resources
- stakeholder interests

5. What support is the state providing local education agencies?

For example: Is training or professional development provided to teachers on alternate assessments?

6. How is the impact of the policies and/or practices being measured? That is, the outcomes or effects of the policies/practices upon which you base success.

Examples might include:

- increased assessment scores
- increased student attendance
- access to general curriculum
- inclusion
- changes to SPED curricula & instruction

7. How are data being used and monitored at the state, district, and school levels?

For example:

Used	Monitored
• feedback purposes	• type of assessment
• program level	• student participation (who, where, etc.)
• instruction for teachers	• quality of assessment
• teacher training	• administration and scoring procedures
	• overall progress

8. Is there technical evidence that supports a link between the state’s alternate assessments and a comprehensive curriculum (i.e., alignment to instruction and content standards)?

For example (per peer review guide): validity, reliability, fairness/accessibility, comparability of results, procedures for test administration, scoring, data analysis, and reporting, interpretation and use of results.

Thank you for your participation. If you have any questions, please feel free to e-mail or call WestEd directly.

REFERENCES

- Ahlgrim-Delzell, L., Flowers, C., Browder D., & Wakeman, S. (2006). *School administrators' perceptions of the impact of alternate assessments*. Manuscript submitted for publication.
- Almond, P., & Bechard, S. (2005). *In-depth look at students who take alternate assessments: What do we know now?* Retrieved June 13, 2007, from <http://www.measuredprogress.org/resources/inclusive/articlespapers/StudentsTakingAltAssess.pdf>
- Almond, P. J., & Case, B. J. (2004). *Alternate assessments for students with significant cognitive disabilities*. San Antonio, TX: Harcourt Assessment, Inc. Available from <http://harcourtassessment.com>
- American Association on Mental Retardation. (2004). *Mental retardation: Definitions, classification, and systems of supports*. Retrieved June 13, 2007, from http://www.aamr.org/Bookstore/MR/mental_definition.shtml
- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: Author.
- Ananda, S. (2003). *Rethinking issues of alignment under No Child Left Behind* (Knowledge Brief). San Francisco, CA: WestEd.
- Arkansas Department of Education. (2006a). *Administrative manual 2006–07: Arkansas alternate portfolio assessment system for ninth grade mathematics for students with disabilities*. Retrieved June 14, 2007, from http://arkansased.org/testing/pdf/Gr9_Math_Admin_FINAL.pdf
- Arkansas Department of Education. (2006b). *Administrative manual 2006–07: Arkansas alternate portfolio assessment system for students with disabilities, grades 3–8, 11*. Retrieved June 14, 2007, from http://arkansased.org/testing/pdf/SWD_AdminManual_FINAL.pdf
- Arkansas Department of Education. (2006c). *Statewide scores 2005–06*. Retrieved August 29, 2007, from http://arkansased.org/testing/test_scores_archive.html#2006
- Baker, F. (2001). *The basics of item response theory*. ERIC Clearinghouse in Assessment and Evaluation, University of Maryland, College Park, MD.
- Bechard, S. (2006). Personal Communication Dated December, 2006.
- Browder, D. (2006, May). *Links for academic learning: The UNC Charlotte alignment method for AA-AAS*. Presentation at ASES SCASS of the Council of Chief State School Officers in Washington, DC.
- Browder, D., Ahlgrim-Delzell, L., Flowers, C., Karvonen, M., Spooner, F., & Algozzine, R. (2002). How states define alternate assessments. *Research and Policy in Developmental Disabilities, 15*(4).
- Browder, D., Ahlgrim-Delzell, L., Flowers, C., Karvonen, M., Spooner, F., & Algozzine, R. (2005). How states implement alternate assessments for students with disabilities and recommendations for national policy. *Journal of Disability Policy Studies, 15*, 209–220.
- Browder, D. M., Fallin, K., Davis, S., & Karvonen, M. (2003). Consideration of what may influence student outcomes on alternate assessment. *Education and Training in Developmental Disabilities, 38*, 255–270.
- Browder, D. M., Flowers, C., Ahlgrim-Delzell, L., Karvonen, M., Spooner, F., & Algozzine, B. (2004). The alignment of alternate assessment content with academic and functional curricula. *Journal of Special Education, 37*, 211–223.
- Browder, D. M., Spooner, F., Ahlgrim-Delzell, L., Flowers, C., Karvonen, M., & Algozzine, R. (2003). A content analysis of the curricular philosophies reflected in states' alternate assessment performance indicators. *Research and Practice for Persons with Severe Disabilities, 28*, 165–181.
- Browder, D. M., Spooner, F., Algozzine, B., Ahlgrim-Delzell, L., Flowers, C., & Karvonen, M. (2003). What we

- know and need to know about alternate assessment. *Exceptional Children*, 70(1), 45–62.
- Case, B. J. (2007). *Alternate assessments for students with significant cognitive disabilities: New rules of measurement*. San Francisco, CA: WestEd.
- Case, B. J., Thompson, S. J., & Thurlow M. L. (2001). *Alternate assessment: State of the states 2001*. Paper presented at the Council of Chief State School Officers Large Scale Assessment Conference, Snowbird, UT.
- Chicago Public Schools (2007). *Northside Learning Center Curriculum*. Retrieved October 2, 2007, from http://www.nlc.cps.k12.il.us/NLC10_courses.html
- Crawford, L., & Tindal, G. (2006). Policy and practice: Knowledge and beliefs of educational professionals related to the inclusion of students with disabilities in a state assessment. *Remedial and Special Education*, 27(4), 208–217.
- Creswell, J. W. (2002). *Qualitative inquiry and research design: Choosing among five traditions*. Thousand Oaks, CA: SAGE.
- Creswell, J. W., Fetters, M. D., & Ivankova N. V. (2004). Designing a mixed methods study in primary care. *Annals of Family Medicine* 2, 7–12.
- Data Recognition Corporation & WestEd. (2006). *LEAP alternate assessment, level 2: Standard setting report*. Minneapolis, MN: Author.
- Delaware Department of Education (2008). *Technical assistance*. Retrieved on January 30, 2008, from <http://www.doe.state.de.us/default.shtml>
- Ebel, R. L. (1983). The practical validation of tests of ability. *Educational Measurement: Issues and Practices*, 2(2), 7–10.
- Elementary and Secondary Education Act, 20 U.S.C. §6311(b)(3)(C)(ii) (2000).
- Elliott, S. N., & Fuchs, L. S. (1997). The utility of curriculum-based measurement and performance assessment as alternatives to traditional intelligence and achievement tests. *School Psychology Review*, 26(2), 224–233.
- Ford, A., Davern, L., & Schnorr, R. (2001). Learners with significant disabilities: Curricular relevance in an era of standards-based reform. *Remedial and Special Education*, 22(4), 214–222.
- Forte Fast, E. (2004). *Developing a framework for examining validity in state accountability systems*. Paper presented at the annual conference of the American Educational Research Association in Minneapolis, MN.
- Fuhrman, S. H., & Elmore, R. E. (2004). *Redesigning accountability systems for education*. New York, NY: Teachers College Press.
- Gong, B., & Marion, S. (2006). *Dealing with flexibility in assessments for students with significant cognitive disabilities* (Synthesis Report 60). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes. Retrieved June 14, 2007, from <http://education.umn.edu/NCEO/OnlinePubs/Synthesis60.html>
- Green, D. R. (1998). Consequential aspects of the validity of achievement tests: A publisher's point of view. *Educational Measurement: Issues and Practices*, 17(2), 16–19.
- Hambleton, K., & Swaminathan, H. (1985). *Item response theory: Principles and applications*. Norwell, MA: Kluwer Academic Publishers.
- Individuals with Disabilities Education Act Amendments of 1997, Pub. L. No. 105–17, 111 Stat. 37 (codified as amended at 20 U.S.C. §1400 *et seq.*).
- Individuals with Disabilities Education Act Amendments of 2004, Pub. L. No. 108–446, 111 Stat. 2647 (codified as amended at 20 U.S.C. §1400 *et seq.*).
- Kampher, S. H., Horvath, L. S., Kleinert, H. L., & Farmer-Kearns, J. (2001). Teachers' perceptions of one's state alternate assessment: Implications for practice and preparation. *Exceptional Children*, 67(3), 671–674.

- Kane, M. (2002). Validating high-stakes testing programs. *Educational Measurement: Issues and Practice*, 21(2), 31–41.
- Karvonen, M., Flowers, C., Browder, D. M., Wakeman, S. Y., & Algozzine, B. (2006). Case study of the influences on alternate assessment outcomes for students with disabilities. *Education and Training in Developmental Disabilities*, 41, 95–110.
- Kim, Y., Angell, M. E., O'Brian, M., Strand, K. H., Fulk, B. M., & Watts, E. H. (2006). Relationships among teachers' perspectives, self-reported practices, and concerns related to alternate assessment system. *Teacher Education and Special Education*, 29(2), 83–97.
- Kiplinger, V. L. (1997). *Standard setting procedures for the specification of performance levels on standards-based assessment*. Denver, CO: Colorado Department of Education.
- Kleinert, H. L., & Kearns, J. F. (1999). A validation study of the performance indicators and learner outcomes of Kentucky's alternate assessment for students with significant disabilities. *The Journal of the Association for Persons with Severe Handicaps*, 24, 100–110.
- Kleinert, H., Browder, D., & Towles-Reeves, E. (2005). *The assessment triangle and students with significant cognitive disabilities: Models of student cognition*. National Alternate Assessment Center, Human Development Institute, University of Lexington, KY. Retrieved June 14, 2007, from <http://www.naacpartners.org/products/Files/NAAC%20Assmt%20Triangle%20White%20Paper%20-%20FINAL%20for%20Website.pdf>
- Kleinert, H. L., Farmer-Kears, J., & Kennedy, S. (1997). Accountability for all students: Kentucky's alternate portfolio assessment for students with moderate and severe cognitive disabilities. *The Journal of the Association for Persons with Severe Handicaps*, 22, 88–108.
- Kleinert, H. L., Kennedy, S., & Farmer-Kears, J. (1999). The impact of alternate assessments: A statewide teacher survey. *The Journal of Special Education*, 33(2) 93–102.
- Koertz, D., McCaffrey, D., Klein, S., Bell, R., & Stecher, B. (1993). *The reliability of scores from the 1992 Vermont portfolio assessment program* (CSE Technical Report 355). Los Angeles, CA: National Center for Research on Evaluation, Standards, and Student Testing.
- Lewis, D. M., Mitzel, H. C., & Green, D. R. (1996, June). Standard setting: A bookmark approach. In D. R. Green (Chair), *IRT-based standard-setting procedures utilizing behavioral anchoring*. Symposium conducted at the meeting of the Council of Chief State School Officers National Conference on Large Scale Assessment in Phoenix, AZ.
- Louisiana Department of Education. (2006). *Annual report 2006*. Retrieved August 29, 2007, from <http://www.doe.state.la.us/lde/saa/759.html>
- Louisiana Department of Education. (2007). *Students with disabilities on state assessments: Submission to the Office of Special Education Programs*. Baton Rouge, LA: Author.
- Manasevit, L. M., & Maginnis, A. M. (2005). *IDEA: New expectations for schools and students*. Washington, DC: Thompson Publishing Group.
- McDonnell, J. J., Hardman, M. L., & McDonnell, A. P. (2003). *Introduction to persons with moderate and severe disabilities: education in social issues*. (2nd ed). Boston, MA: Allyn and Bacon.
- McDonnell, L. M., McLaughlin, M. J., & Morison, P. (Eds.). (1997). *Educating one and all: Students with disabilities and standards-based reform*. Washington, DC: National Academy Press.
- McGregor, G. (2007). *Examining the interrater reliability of Montana's CRT-alternative* (Technical Report). Helena, MT: NCLB Peer Review Process, Montana Office of Public Instruction.
- Mechling, L. (2006, Spring). *Functional academics lab: Course syllabus*. Retrieved August 29, 2007, from <http://people.uncw.edu/MechlingL/documents/SU.S.DepartmentofEducationL365lab.doc>

- Messick, S. (1993). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed.). New York, NY: Macmillan.
- Missouri Department of Elementary and Secondary Education. (2007). *Functional skills support materials guide*. Retrieved August 24, 2007, from http://www.ssd.k12.mo.us/Staff/instructional_tools/assets/FUNCTIONAL%20SKILLS.pdf
- National Alternate Assessment Center. (2005). *Designing from the ground floor: Alternate assessment on alternate achievement standards*. Lexington, KY: Author.
- National Center on Educational Outcomes. (2003). *Accountability for assessment results in the No Child Left Behind Act: What it means for children with disabilities*. Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes. Retrieved November 6, 2007, from <http://cehd.umn.edu/nceo/OnlinePubs/NCLBdisabilities.html>
- National Center on Learning Disabilities. (2007). *IDEA Terms to Know*. Retrieved August 23, 2007, from <http://www.ncl.org/content/view/921/456099/>
- New Mexico Public Education Department. (2007a). *New Mexico standards-based and alternative assessments—Math state level results 2004–05*. Retrieved June 6, 2007, from <http://www.ped.state.nm.us/div/acc.assess/assess/index.html>
- New Mexico Public Education Department. (2007b). *New Mexico standards-based and alternative assessments—Reading state level results 2004–05*. Retrieved June 15, 2007, from http://www.ped.state.nm.us/div/acc.assess/assess/NMSBA_Statistics_and_Data/downloads/NMSBA%20State%20Level%20Proficiency%20Results%20_%20Reading,%20Alternate%20Assessment%20as%20Percentage%20of%20All%20Students%20Tested%202004-2005.pdf
- Nitko, A. (1996). *Educational assessments of students* (2nd ed.). Englewood Cliffs, NJ: Merrill Prentice Hall.
- No Child Left Behind Act of 2001, Public Law 107–110.
- North Central Regional Education Laboratory. (2007). *Reliability, validity, and fairness of classroom assessments*. Retrieved August 29, 2007, from <http://www.ncrel.org/sdrs/areas/issues/methods/assment/as5relia.htm>
- Oklahoma State Department of Education. (2006). *Oklahoma alternate assessment program portfolio 2006*. Retrieved June 8, 2007, from <http://se.sde.state.ok.us/ses/assessment/OAAPManual2006.pdf>
- Olsen, B., Mead, R. & Payne, D. (2002). *A report of standard setting methods for alternate assessment for students with significant cognitive disabilities*. (NCEO Synthesis Report 47). Minneapolis MN: University of Minnesota National Center on Educational Outcomes.
- Pellegrino, J., Chudowsky, N., & Glaser, P. (2001). *Knowing what students know*. Washington, DC: National Research Council, National Academy Press.
- Policy to Practice Study Group of the Assessing Special Education Students. (2003). *Glossary of assessment terms and acronyms used in assessing special education students: A report from the Assessing Special Education Students (ASES) State Collaborative on Assessment and Student Standards (SCASS)*. Washington, DC: Council of Chief State School Officers. Retrieved November 6, 2007, from <http://www.ccsso.org/content/pdfs/ASESSCASSGlossary.pdf>
- Provincial Outreach Program for Autism and Related Disorders. (2007). *Functional curriculum*. Retrieved August 24, 2007, from <http://www.autismoutreach.ca/?action=view&type=page&id=86&loc=functional-curriculum>
- Pugach, M. C., & Warger, C. L. (2001). Curriculum matters: Raising expectations for students with disabilities. *Remedial and Special Education*, 22(4), 194–196.
- Quenemoen, R., Rigney, S., & Thurlow, M. (2002). *Use of alternate assessment results in reporting and accountability systems: Conditions for use based on research and practice* (Synthesis Report 43). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes. <http://www.education.umn.edu/NCEO/OnlinePubs/Synthesis43.html>

- Rabinowitz, S., & Sato, E. (2005). *The technical adequacy of assessment for alternate student populations: Guidelines for consumers and developers*. San Francisco, CA: WestEd.
- Roach, A. T., & Elliott, S. N. (2004). *Alignment analysis and standard-setting procedures for alternate assessment* (Working Paper No. 2004-1). Wisconsin Center for Education Research. Retrieved November 6, 2007, from http://www.wcer.wisc.edu/publications/workingPapers/Working_Paper_No_2004_1.php
- Roach, A. T., & Elliott, S. N. (2006). The influence of access to the general education curriculum on the alternate assessment performances of students with significant cognitive disabilities. *Education Evaluation and Policy Analysis*, 20(2), 181–194.
- Roach, A. T., Elliott, S. N., & Webb N. L. (2005). Alignment of an alternate assessment with state academic standards: “Evidence for the content validity of the Wisconsin alternate assessment.” *Journal of Special Education*, 38(4), 218–231.
- Roeber, E. (2002). *Setting standards on alternate assessments* (Synthesis Report 42). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes. <http://www.education.umn.edu/NCEO/OnlinePubs/Synthesis42.html>
- Salvia, T. & Ysseldyke, J. E. (2001). *Assessment* (8th ed.). Boston, MA: Houghton Mifflin.
- Sato, E., Rabinowitz, S., & Wilkins, C. (2007, April). *Alternate assessment policy and practice: An examination of implementation and impact across southwestern states*. Paper presented at the annual meeting of the American Educational Research Association in Chicago, IL.
- Schafer, W. D. (2005). Technical documentation for alternate assessments. *Practical Assessment, Research, & Evaluation*, 10(10).
- Snell, M. E., & Brown, F. (Eds.). (2006). *Instruction of students with severe disabilities* (6th ed.). Upper Saddle River, NJ: Merrill.
- Texas Education Agency. (2005). *Technical digest 2004–2005*. Retrieved June 7, 2007, from <http://www.tea.state.tx.us/student.assessment/resources/techdig05/index.html>
- Texas Education Agency. (2006a). *TEKS curriculum framework for TAKS-Alt: Grade 3 reading*. Retrieved June 6, 2007, from http://www.tea.state.tx.us/student.assessment/resources/taksalt/curriculum/TKA07_AltCurrFramework_G3_R.pdf
- Texas Education Agency. (2006b). *TEKS curriculum framework for TAKS-Alt grade 4 reading*. Retrieved June 6, 2007, from http://www.tea.state.tx.us/student.assessment/resources/taksalt/curriculum/TK06_AltCurrFramework_G4_R_final_tagged.pdf
- Texas Education Agency. (2006c). *TEKS curriculum framework for TAKS-Alt grade 5 reading*. Retrieved June 6, 2007, from http://www.tea.state.tx.us/student.assessment/resources/taksalt/curriculum/TK06_AltCurrFramework_G5_R_final_tagged.pdf
- Texas Education Agency. (2006d). *TEKS curriculum framework for TAKS-Alt grade 6 reading*. Retrieved June 6, 2007, from http://www.tea.state.tx.us/student.assessment/resources/taksalt/curriculum/TK06_AltCurrFramework_G6_R_final_tagged.pdf
- Texas Education Agency. (2006e). *TEKS curriculum framework for TAKS-Alt grade 7 reading*. Retrieved June 6, 2007, from http://www.tea.state.tx.us/student.assessment/resources/taksalt/curriculum/TK06_AltCurrFramework_G7_R_final_tagged.pdf
- Texas Education Agency. (2006f). *TEKS curriculum framework for TAKS-Alt grade 8 reading*. Retrieved June 6, 2007, from http://www.tea.state.tx.us/student.assessment/resources/taksalt/curriculum/TK06_AltCurrFramework_G8_R_final_tagged.pdf
- Texas Education Agency. (2006g). *Participation requirements for TAKS-Alt*. Retrieved June 5, 2007, from http://www.tea.state.tx.us/student.assessment/resources/taksalt/training/participation_requirements.pdf

- Texas Education Agency. (2006h). *State-required essence statements for TAKS-Alt*. Retrieved June 5, 2007, from <http://www.tea.state.tx.us/student.assessment/resources/taksalt/essence/index.html>
- Texas Education Agency. (2006i). *Samples of completed data collection sheets*. Retrieved June 5, 2007, from <http://www.tea.state.tx.us/student.assessment/resources/taksalt/samples/index.html>
- Texas Education Agency. (2006j). *TAKS-Alt rubric*. Retrieved June 5, 2007, from http://www.tea.state.tx.us/student.assessment/resources/taksalt/rubric/taksalt_rubric.pdf
- Texas Education Agency. (2006k). *TEKS vertical alignment documents for TAKS-Alt*. Retrieved June 5, 2007, from <http://www.tea.state.tx.us/student.assessment/resources/taksalt/index.html>
- Texas Education Agency. (2007). *Texas state-developed alternate assessment II summary report English language arts—all students*. Retrieved June 6, 2007, from http://www.tea.state.tx.us/student.assessment/reporting/results/summary/sum07/sdaa/sdaa2_ela_tagged.pdf
- Thompson, S. J., & Thurlow, M. L. (2001). *2001 state special education outcomes: A report on state activities at the beginning of a new decade*. Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.
- Thompson, S., & Thurlow, M. (2003). *2003 state special education outcomes: Marching on*. Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes. Retrieved November 16, 2007, from <http://cehd.umn.edu/NCEO/OnlinePubs/2003StateReport.htm>
- Thompson, S., Case, B., Thurlow, M. (2000). *Alternate Assessments: The state of the states—a comparison of state versus local district viewpoints*. Paper presented at the national meeting of the Council of Chief State School Officers in Snowbird, UT.
- Thompson, S. J., Johnstone, C. J., Thurlow, M. L., & Altman, J. R. (2005). *State special education outcomes: Steps forward in a decade of change*. Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes. Retrieved November 16, 2007, from <http://cehd.umn.edu/NCEO/OnlinePubs/2005StateReport.htm>
- Thurlow, M. (2004). *How State Policies and Practices for Alternate Assessment Impact Who is Included in NAEP for State Assessments*. Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes. Retrieved August 24, 2007, from www.nagb.org/pubs/conferences/thurlow.doc
- Thurlow, M., & Case, B. J. (2004). *Alternate assessments: State of the states*. Presentation at the Council of Chief State School Officers Large Scale Conference in Boston, MA.
- Thurlow, M., Moen, R., & Altman, J. (2006, June). *Annual performance reports: 2003–2004 state assessment data*. Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes. Retrieved on July 15, 2007, from <http://www.education.umn.edu/NCEO/>
- Tindal, G., McDonald, M., Tedesco, M., Glasgow, A., Almond, P., Crawford, L., & Hollenbeck, K. (2003). Alternate assessments in reading and math: Development and validation for students with significant disabilities. *Exceptional Children*, 69, 481–494.
- Towles-Reeves, E., Muhomba, M., & Kleinert, H. (in press). Alternate assessment: Have we learned anything new? *Exceptional Children*.
- Turnbull, R., Turnbull, A., Shank, M., Smith, S., & Leal, D. (2002). *Exceptional Lives: Special Education in Today's Schools* (3rd ed.). Columbus, OH: Merrill Prentice Hall.
- U.S. Department of Education. (2002). *Title I—Improving the academic achievement of the disadvantaged: Final Rule*. Washington, DC: Author.
- U.S. Department of Education. (2003a). *Title I Regulation on Alternate Achievement Standards: Questions and Answers*. Washington, DC: Author.
- U.S. Department of Education. (2003b). *No Child Left Behind provision gives schools increased flexibility*

- while ensuring that all children count, including those with disabilities* (press release). Retrieved April 1, 2007, from <http://www.ed.gov/news/pressreleases/2003/12/12092003.html>
- U.S. Department of Education. (2004). *A–Z index: Glossary of terms*. Retrieved November 5, 2007, from <http://www.ed.gov/nclb/index/az/glossary.html>
- U.S. Department of Education. (2005). *No Child Left Behind alternate achievement standards for students with the most significant cognitive disabilities: Non-regulatory guidance*. Washington, DC: Author.
- U.S. Department of Education. (2006a). *No Child Left Behind standards and assessments: Peer review status letters*. Retrieved November 6, 2007, from <http://www.ed.gov/admins/lead/account/nclbfinalassess/index.html>
- U.S. Department of Education. (2006b). *Regulations for the Individuals with Disabilities Education Act 2004*. 71 Fed. Reg. 156 (Aug. 14, 2006).
- U.S. Department of Education. (2007). *No Child Left Behind modified achievement standards: Non-regulatory guidance*. Washington, DC: Author.
- U.S. Department of Education, National Center for Education Statistics. (2007). *Common Core of Data 2005/06*. Author.
- U.S. Department of Education, Office of Elementary & Secondary Education. (2004). *Standards and assessments peer review guidance: Information and examples for meeting requirements of the No Child Left Behind Act of 2001*. Washington, DC: Author.
- Webb, N. L., Horton, M., & O’Neal, S. (2002, April). *An analysis of the alignment between language arts standards and assessment in four states*. Paper presented at the annual meeting of the American Educational Research Association, New Orleans.
- WestEd. (2004). *Alignment study: Louisiana content standards and the LEAP alternate assessment (LAA) final report*. San Francisco: WestEd.
- Wisconsin Center for Educational Research. (2007). *Aligning assessments and standards*. Retrieved September 6, 2007, from http://www.wcer.wisc.edu/news/coversStories/aligning_assessments_and_standards.php
- Yovanoff, P., & Tindal, G. (2007). Scaling early reading alternate assessments with statewide measures. *Exceptional Children*, 73(2), 184–201.
- Ysseldyke, J., & Olsen, K. (1999). Putting alternate assessment into practice: What to measure and possible sources of data. *Exceptional Children*, 65(2), 175–185.
- Ysseldyke, J., Dennison, A., & Nelson, R. (2004). *Large scale assessment and accountability systems: Positive consequences for students with disabilities* (Synthesis Report 51). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.