

A Brief History of Alternate Assessments Based on Alternate Achievement Standards



**NATIONAL
CENTER ON
EDUCATIONAL
OUTCOMES**

In collaboration with:

Council of Chief State School Officers (CCSSO)

National Association of State Directors of Special Education (NASDSE)

Synthesis Report 68

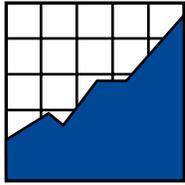
A Brief History of Alternate Assessments Based on Alternate Achievement Standards

Rachel Quenemoen

September 2008

All rights reserved. Any or all portions of this document may be reproduced and distributed without prior permission, provided the source is cited as:

Quenemoen, R. (2008). *A brief history of alternate assessments based on alternate achievement standards* (Synthesis Report 68). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.



**N A T I O N A L
C E N T E R O N
E D U C A T I O N A L
O U T C O M E S**

The Center is supported through a Cooperative Agreement (#H326G050007) with the Research to Practice Division, Office of Special Education Programs, U.S. Department of Education. The Center is affiliated with the Institute on Community Integration at the College of Education and Human Development, University of Minnesota. Opinions expressed herein do not necessarily reflect those of the U.S. Department of Education or Offices within it.



NCEO Core Staff

Martha L. Thurlow, Director	Ross E. Moen
Deb A. Albus	Michael L. Moore
Jason R. Altman	Rachel F. Quenemoen
Manuel T. Barrera	Christopher Rogers
Laurene L. Christensen	Dorene L. Scott
Christopher J. Johnstone	Vitaliy Shyyan
Jane L. Krentz	Miong Vang
Sheryl S. Lazarus	Yi-Chen Wu
Kristi K. Liu	

National Center on Educational Outcomes
University of Minnesota • 207 Pattee Hall
150 Pillsbury Dr. SE • Minneapolis, MN 55455
Phone 612/626-1530 • Fax 612/624-0879
<http://www.nceo.info>

The University of Minnesota is committed to the policy that all persons shall have equal access to its programs, facilities, and employment without regard to race, color, creed, religion, national origin, sex, age, marital status, disability, public assistance status, veteran status, or sexual orientation.

This document is available in alternative formats upon request.

Executive Summary

This report provides a historical look back over the past 15 years of alternate assessment development, from the early 1990s through the mid 2000s, as reported by state directors of special education on the National Center on Educational Outcomes (NCEO) state surveys, and augmented by other research and policy reports published by NCEO and related organizations during that time frame. It is meant to be a resource to state and federal policymakers and staff, researchers, test companies, and the public to help us understand why and where we have come from and where we may be going in the challenging work of alternate assessment for students with significant cognitive disabilities.

The early work on alternate assessments in Kentucky and Maryland was a lens through which early alternate assessments required by the Individuals with Disabilities Education Act Amendments of 1997 were viewed, but states immediately began to tailor these new tests to their own views of education reform for all students, as well as to historical state perspectives on teaching and learning for students with the most significant disabilities. Shifting state perspectives over the time span are documented here. There are six alternate assessment topics covered more or less throughout the span of NCEO survey and research reports, including stakeholder expectations and principles; content coverage (linkage to content standards); approaches (test format); scoring criteria and procedures; performance/achievement level descriptors and standard setting; and reporting and accountability. In the years since the passage of the No Child Left Behind Act of 2001, the focus of alternate assessment work has been on technical defense of state approaches. The work of the National Alternate Assessment Center and related projects and centers has focused on a validity framework as a heuristic for state practice, and that work is described here.

The report ends with four recommendations to guide state practices at this point. Because of the number of uncertainties still in play, we need:

1. **Transparency.** We need to know what varying practices and targets yield for student outcomes, and the only way to build that knowledge base is to ensure that assessment development, implementation, and results are transparent and open to scrutiny.
2. **Integrity.** Building on the need for transparency is the need for integrity. The amount of flexibility needed to ensure that all students can demonstrate what they know and can do is higher in alternate assessments for this group of students than in more typical student populations. Flexibility can mask issues of teaching and learning unless it is carefully structured and controlled. Similarly, standardization as a solution risks reducing the integrity of the assessment results when the methods do not match the population being assessed and how that population demonstrates competence in the academic domains.

3. **Validity studies.** Building on the issues of transparency and integrity, we have an obligation to monitor carefully the effects of alternate assessments over time, as well as to ensure the claims we are making for the use of the results are defensible.

4. **Planned improvement over time.** In building a validity argument, we study whether the interpretations and uses of the test are defensible, and whether consequences that are hoped for and those that are to be avoided are in fact falling into their respective places.

An important part of validity studies is the ongoing day-to-day oversight of the assessment development, implementation, and use of testing results, and high quality data collection and continuous improvement based on the data are absolutely necessary for these assessments.

Acknowledgements

This report is an adaptation from a paper first presented at the Maryland Assessment Research Center for Education Success (MARCES) conference in College Park Maryland, October 2007. The conference Web site has numerous other presentations at the conference, and is a valuable resource: http://www.education.umd.edu/EDMS/MARCES/conference/alt_assessment/agenda.htm.

Special thanks are due to two of NCEO's Research to Practice Panelists, Dr. Claudia Flowers of the University of North Carolina Charlotte and Dr. Harold Kleinert of the University of Kentucky, for their thoughtful substantive reviews and comments on earlier drafts of this paper. In addition, Dr. Marianne Perie from the National Center for the Improvement of Educational Assessment and Dr. Jacqui Kearns from the National Alternate Assessment Center provided invaluable feedback. Finally, NCEO's Director Dr. Martha Thurlow provided ongoing feedback and guidance during conceptualization and development of the report. Each of these researchers improved this report substantially; any errors remaining are the author's alone.

Table of Contents

Overview	1
Purpose of This Report	1
Documentation of Alternate Assessments: State Surveys and Policy Research	2
Early Thinking that Shaped Alternate Assessment	3
Federal Policy Historical Context for Alternate Assessments	4
1997: The Initiation of Federal Requirements for Alternate Assessment	5
Stakeholders, Expectations, and Principles	5
Content Coverage (Linkage to Content Standards)	8
Alternate Assessment Approaches (Format)	10
Scoring Criteria and Procedures	11
Performance/Achievement Level Descriptors and Standard Setting	14
Reporting and Accountability	15
The Transition to New Thinking	16
Current Status of Alternate Assessments based on Alternate Achievement Standards	18
Stakeholders, Expectations, and Principles	18
Content Coverage (Linkage to Content Standards)	20
Alternate Assessment Approach (Format)	21
Scoring Criteria and Procedures	21
Performance/Achievement Level Descriptors and Standard Setting	22
Reporting and Accountability	23
Considerations for State Practice	23
Conclusion	26
References	27

Overview

The standards-based educational reform efforts that began in the late 1980s resulted in a renewed focus on the participation and performance of all students on state-defined academic standards and assessments. In the early 1990s, most states included 10% or fewer of their students with disabilities in state assessments (Shriner & Thurlow, 1993). Negative consequences of excluding students with disabilities were documented, including increased rates of referral to special education, exclusion from the curriculum, and no information on the educational results of students with disabilities (Ysseldyke, Thurlow, McGrew, & Shriner, 1994). Participation rates in state assessments grew into the 2000s, pushed along by Congressional action through the reauthorizations of the Title I and special education legislation. As documented through public peer review of state assessment systems under the No Child Left Behind Act, by 2008 all states have built assessment systems with the goal of at least the federally required 95% participation rates by all students and subgroups including students with disabilities.

In these state systems, “all students” means all students, including those students with significant cognitive disabilities (cognitive disabilities generally defined for this purpose as mental retardation). In 1990, large-scale academic assessment of these students did not exist, and only a few policymakers were contemplating the necessity of doing so. This report documents the relatively brief history of alternate assessments for students with significant cognitive disabilities, a history that reflects a cornerstone effort to support a truly inclusive and accountable public education system.

Purpose of This Report

This synthesis report provides a historical look back over the past 15 years of alternate assessment, from the early 1990s through the mid 2000s, as reported by state directors of special education on the National Center on Educational Outcomes (NCEO) state surveys, and augmented by other research and policy reports published by NCEO and related organizations during that time frame. It is meant to be a resource to state and federal policymakers and staff, researchers, test companies, and the public to help us understand why and where we have come from and where we may be going in the challenging work of alternate assessment for students with significant cognitive disabilities.

Documentation of Alternate Assessments: State Surveys and Policy Research

NCEO has conducted biennial surveys on state assessment practices related to students with disabilities since the early 1990s. Through 2005, state directors of special education participated in the survey, with 100% response rate by regular states over that time span, and more varied participation by unique entities (i.e., entities beyond the 50 states receiving Federal special education funding). These surveys covered a full range of issues related to inclusive assessment practices in states, including accommodations, alternate assessments, universal design of assessments, and emerging trends. The focus was on state assessments designed for the purpose of public reporting and accountability. In 2007, survey items related to alternate assessment were eliminated because the National Alternate Assessment Center (NAAC) at the University of Kentucky took over the role of research into and documentation of state practices in alternate assessment. (For NCEO's reports based on these surveys, see <http://www.nceo.info/OnlinePubs/statereports.html>.)

NCEO has also documented alternate assessment practices through periodic research and policy publications, beginning with the earliest development of alternate assessments in Kentucky and Maryland in the early 1990s, and continuing in collaboration with special education and measurement organizations and researchers through the first decade of 2000.

There are six alternate assessment topics covered more or less throughout the span of these survey and research reports, including:

- stakeholder expectations and principles;
- content coverage (linkage to content standards);
- approaches (test format);
- scoring criteria and procedures;
- performance/achievement level descriptors and standard setting; and
- reporting and accountability.

Not all topics were covered equally in NCEO surveys and research reports throughout the time span. Initially the topics of stakeholder expectations and principles, content coverage (linkage to content standards), and approaches (test format) were the focus from the 1999 survey forward. Recently the topics of scoring criteria and procedures, performance/achievement level descriptors and standard setting, and reporting and accountability emerged as the No Child Left Behind Act of 2001 (NCLB) required that states demonstrate the technical defensibility of their alternate assessments for use in their accountability systems. This evolution of topics illustrates the challenges states faced during initial conceptualization of alternate assessments, and also

how these assessments changed to meet new professional understanding as well as new state and federal requirements.

Early Thinking that Shaped Alternate Assessment

In the early 1990s, Maryland and Kentucky were states that initiated school accountability systems based on student achievement, required by the Maryland legislature, and in Kentucky the state courts provided impetus to change followed by legislative action. The states shared a common policy imperative that all students must be included in school accountability analyses, including students who could not participate in the general assessments, even with accommodations, adaptations, or other supports (Kleinert, Haigh, Kearns, & Kennedy, 2000; Ysseldyke, Thurlow, Erickson, Gabrys, Haigh, Trimble, & Gong, 1996). These students were identified primarily as students who had what were considered the most severe and complex disabilities, students served under varying labels like “severe-profound disabilities” and “trainable mentally handicapped (TMH).” Experts in severe disabilities weighed in on these new assessments, and based on research done in Kentucky and Maryland—and on the literature in severe disabilities—four assumptions were posed by Ysseldyke and Olsen (1997) that reflected early beliefs and practice in the development of alternate assessments. These assumptions shaped the early efforts in alternate assessment and continue to be reflected in many state alternate assessments today. The four foundational assumptions identified in that important report, and excerpts from the rationale for each, are included below:

1. Focus on authentic skills and on assessing experiences in community/real life environments. Artificial assessment tasks will not provide an indication of how well the system is preparing the students; however, “community” means different things at primary, middle and secondary levels. For a third grader, community might be the school, the playground and home, whereas community for an exiting senior would have to mean the store, bank, and workplace, for example.
2. Measure integrated skills across domains. [E]ducation, especially for students with moderate to severe cognitive disabilities, requires integration of skills. So should the assessments. For example, assessing personal and social skills separately from assessing independence and responsibility would result in redundant effort and possibly result in reinforcing a focus on isolated skills. A generic rubric that encompasses multiple skills would be more appropriate.
3. Use continuous documentation methods if at all possible. Using assessment methods that involve multiple measures over time will result in more accurate and reliable information. Students with severe challenges have greater variability in their skills from

day-to-day than do students without disabilities or even students with milder disabilities. Therefore, a skill that cannot be observed on one day might be fully in place the next day. Milestones for students with severe disabilities are much farther apart than for other students, and methods that capture change rather than status will better reflect success of the educational system.

4. Include, as critical criteria, the extent to which the system provides the needed supports and adaptations and trains the student to use them. If the purpose is to hold the educational system accountable, the only way to assess the extent to which a school system is providing the needed education is to include, as one of the criteria for success, the extent to which the school system provides the needed assistive devices, people and other supports to allow the students to function as independently as possible. There is more variability in the skill levels and needs of this 1% of the students than there is in the rest of the total student population. Kentucky has shown that including this criterion has the added benefit of driving effective school and classroom practice (Kleinert, Kennedy, & Kearns, [in press at time of the 1997 report] 1999). (Ysseldyke & Olsen, 1997, pp. 16-17).

These assumptions were shaped in a context of state standards-based reform prior to Federal laws that later shifted focus to accountability for academic achievement for all students. Since then, some of the underlying beliefs and practices from that time have been augmented by new understanding of how these students with complex disabilities access and demonstrate skills and knowledge in the academic standards-based curriculum. Even with our new understanding of how this small group of students learns in the academic domains, these assumptions from the late 1990s reflect the teaching and learning literature of severe disabilities prior to the addition of a standards-based curriculum for these students. A review of state survey data suggests that many states still see these assumptions as important to consider in development of alternate assessments, although states have had to raise the bar on expectations for these students and for the alternate assessments that tell us how well these students are achieving in a standards-based academic context.

Federal Policy Historical Context for Alternate Assessments

The Individuals with Disabilities Education Act (IDEA) Amendments in 1997 redefined what students with disabilities should know and be able to do. IDEA 1997 also included the first Federal requirement of alternate assessments. In the preamble to IDEA 1997, Congress noted that historically, “the implementation of this Act has been impeded by low expectations, and an insufficient focus on applying replicable research on proven methods of teaching and learning

for children with disabilities. Over 20 years of research and experience has demonstrated that the education of children with disabilities can be made more effective by having high expectations for such children and ensuring their access in the general curriculum to the maximum extent possible.”

IDEA previously had required that students with disabilities have access to the school building, but now these students were to have access to and show progress in the same challenging curriculum as their peers. Although not everyone recognized the magnitude of the shift at the time, the states that responded to the requirements with increased expectations started redefining what “the maximum extent possible” described in the IDEA preamble really meant for students with disabilities, including those with the most severe disabilities. The history of alternate assessments reflects this shift in thinking, predating Federal law, but gathering momentum with the passage of IDEA 1997.

1997: The Initiation of Federal Requirements for Alternate Assessment

IDEA 1997 first required alternate assessments, and in the 1999 NCEO survey of state special education directors, 20 states indicated they were developing some type of alternate assessment. Still, only Kentucky and Maryland reported they had the alternate assessment in place. Most state systems were still in development as reported on the 1999 and 2001 surveys, but by 2003, nearly all states had at least one alternate assessment in place. Eight states had two alternate assessments for students with varying needs, and three states had three or more different alternate assessment options in place. During this time of rapid change, the surveys addressed early steps in the creation of alternate assessments, including identification of stakeholders involved in development, as well as core principles guiding development, the content assessed, and the approach or format used by each state.

Stakeholders, Expectations, and Principles

The early years of alternate assessments reflect what later became a dramatic shift in the field. While severe disability experts were beginning to see the value in academic instruction for students with significant cognitive disabilities, most states’ alternate assessments still reflected a predominantly functional curricular approach (Kleinert & Kearns, 1999).

Most state agencies and researchers began working on alternate assessment by tapping into key stakeholders who were well trained in a functional approach. They built on a research base that had almost no mention of academic content as desirable or even attainable for these students.

For example, in the 1999 NCEO survey, a question asked state special education directors to estimate “the percent of students whose exposure to content was too limited for them to participate in regular assessment.” The question was meant to reflect the percentage of the entire student population, not just those with the disabilities, and respondent comments corroborate that was how the question was interpreted. Table 1 shows that of the 36 state directors who responded to the question, 8 (22%) estimated that for more than 4% of the total population of students exposure to content was too limited for them to participate in regular assessment, and almost the same number (n=7, 19%) estimated that less than 1% of the total student population had limited exposure to the content to participate in the regular assessment. The remainder of state special education directors estimated between 1 and 4% of the student population had such limited exposure to the content that they could not participate in the regular assessment. This was different from the IDEA 1997 definition of the students who require alternate assessments, which was that they cannot take regular assessments, even with accommodations. These responses probably reflect accurately the status of these students’ access to the general curriculum.

Table 1. Estimated Percentages of All Students Whose Exposure to Content is Too Limited for Them to Participate in Regular Assessment

< 1 – 1%	> 1 – 2%	> 2 – 4%	> 4%
Delaware*	California	Arkansas*	Mississippi
Kansas	Colorado	Connecticut	Ohio
Kentucky	Hawaii	Massachusetts	South Dakota
Maryland	Idaho	Missouri	Tennessee
Minnesota	Indiana	New Hampshire	Texas*
Nebraska	Florida*	New Mexico	West Virginia
Vermont	Louisiana	Utah	
	Nevada	Washington	
	Oregon	Wisconsin	
	Rhode Island		
	Virginia		

*State provided percentage of students with disabilities was transformed to a percentage of all students using the special education rate.

Note. From *1999 State Special Education Outcomes: A Report on State Activities at the End of the Century*, by S. Thompson & M. Thurlow, 1999, Minneapolis: National Center on Educational Outcomes. Reprinted with permission.

Since students with disabilities are roughly 10% of average state total population, the survey results above can be translated to suggest that from under 10% up to 90% of students with disabilities in their states were not being taught the content that was covered by the regular assessment. For example, one state reported that for 9% of the entire student population their exposure to content was too limited for them to participate in regular assessment, which would

then translate to almost the entire estimated 10% of students who may have disabilities not having access to the content on the regular assessment. That would be in conflict with the requirement that all students with disabilities have access to the general curriculum. By contrast, the two states with an alternate assessment in place at that time, Kentucky and Maryland, were among the “less than 1% group,” which would be less than 10% of students with disabilities.

In addition to the limited access these students had to academic content up until that time, states were faced with almost no practice or research on the inclusion of these students in large-scale assessments. For many states, the starting point for building an alternate assessment was to identify principles to guide development, defining expectations in a general way. States varied dramatically in how they defined these principles (Thompson & Thurlow, 2000). Compare and contrast the principles below, taken from three states:

State #1

- Expectations for all students should be high, regardless of the existence of any disability.
- The goals for an educated student must be applicable to all students, regardless of disability.
- Special education programs must be an extension and adaptation of general education programs rather than an alternate or separate system.

State #2

- All children have value, can learn and are expected to be full participants in the school experience.
- School personnel, parents, local, and state policymakers, and the students themselves are responsible for ensuring this full participation.
- The Standard Course of Study is the foundation for all students, including students with unique learning needs.

State #3

- Meet the law.
- Nonabusive to students, staff, parents.
- Inexpensive.
- Easy to do and takes little time.

(Thompson & Thurlow, 2000, pp. 2-3)

Thompson and Thurlow (2000) identified several trends that affected alternate assessment development throughout the time period. First, most states developed the overall approach and format of the alternate assessment in partnership with stakeholders, given the dearth of experience on alternate assessments in the literature or in practice. Stakeholders typically included general

and special educators, often joined by parent representatives from the state special education advisory committees or parent organizations, but it was clear that in a small number of states, alternate assessment was perceived as a problem to be resolved by and for special education (see also Kohl, McLaughlin, & Nagle, 2006). Second, even at the very beginning of alternate assessment work, functional content versus academic content was emerging as a tension in design of alternate assessments; debates on what to measure have been ongoing since that time. Finally, that report identified the emerging challenge of understanding in state assessment offices how these “odd” large-scale tests could be scored and reported with integrity.

Content Coverage (Linkage to Content Standards)

The changing understanding of the nature of content coverage, in the context of the IDEA 1997 mandate of access to and progress in the general curriculum, is reflected in shifts over the time period. The field moved from a focus on functional skills in the early years to a focus on academics in the most recent years. The belief systems in some states were challenged early on by the 1997 IDEA requirements, and their alternate assessments reflected that shift. This shift in content has continued throughout the time period, with more states refocusing on academic content, particularly after implementation of NCLB requirements. Table 2 shows this trend across all state survey reports.

Table 2: Content Addressed by Alternate Assessments: Change Over Time

Year	Functional Skills No Link to SCS	Functional Skills Link to SCS	SCS Plus Functional Skills	Expand Extend SCS	Grade Level SCS	IEP Team IDs Content	Other	Revising
1999	16	---	1	19	---	---	24	---
2000	9	3	7	28	---	---	3	---
2001	4	15	9	19	---	---	3	---
2003	2	---	4	36	---	3	3	2
2005	---	---	1	21	10	1	7	10

SCS=State Content Standards

Note: Data taken from 1999-2005 NCEO State Survey Reports, S. Thompson & Martha Thurlow, 1999, 2001, 2003; S.J. Thompson, C.J. Johnstone, M.L. Thurlow, & J.R. Altman, 2005, Minneapolis: National Center on Educational Outcomes. Adapted with permission.

Note that in 2005 there were still states revising the content covered by the alternate assessment, and in 2005 NCEO added a response category called “grade level standards.” States that had implemented the IDEA 1997 emphasis on access to and progress in the general curriculum were beginning to collect evidence that these students could learn and achieve in the academic content in ways that surprised even long time researchers in the area (Browder, Ahlgrim-Delzell, Courtade, Gibbs, & Flowers, in press).

From the beginning, state leaders and stakeholders in Massachusetts built their alternate assessment based on the assumption that all students should have access to the same challenging academic skills and knowledge, and be able to demonstrate their achievement (Wiener, 2005). Soon, Massachusetts and a few other early pioneering states shared student work evidencing academic content and skills that had never before been taught to these students. That evidence resulted in increasing pressure from federal policy and from advocates that all states shift to higher expectations for these students. Increasing academic expectations for students with severe disabilities is arguably the most dramatic result of development of alternate assessments in the wake of IDEA 1997.

Changing curricular content for students with significant cognitive disabilities. A brief summary of the series of changes in curricular content for students with significant cognitive disabilities is included here to provide context for the shifting content coverage of alternate assessments. The field of education for students with severe disabilities has been in a state of constant rediscovery since the early and mid 1970s, and has been documented by many researchers (e.g., Browder & Spooner, 2006; National Alternate Assessment Center training materials, 2005).

In the early 1970s, the field of severe disabilities focused on adapting infant/early childhood curriculum for students with the most significant disabilities of all ages. However, severe disability experts began to question the validity of this approach (see Brown, Nietupski, & Hamre-Nietupski, 1976), in part because of the disconnect between the learning progressions assumed by the infant/early childhood curriculum and the actual observations of what these students could achieve in spite of not having developed earlier skills. By the 1980s, the field had moved to a functional skills model. As the evidence for this approach mounted, the field refocused on age-appropriate skills and knowledge performed in authentic settings, and the functional life skills curriculum became “best practice.” The functional, age-appropriate curricular focus resulted in these students demonstrating skills and knowledge not thought possible earlier (Browder & Spooner, 2006).

In the 1990s, additional important new practices were identified as best practice in teaching and learning for students with severe disabilities. The practice of including students with severe disabilities with typical peers in classroom settings for purposes of social inclusion, along with a new focus on self determination skills, reflected a new acceptance of the students, and an

understanding of values related to social development (Browder & Spooner, 2006). The advent of more sophisticated assistive technology opened the world of communication for the first time for some students, and enhanced the ability of teachers and students to interact. The next major shift was that of general curriculum access, as required by IDEA 1997, and clarified by NCLB 2001 and IDEA 2004. Academics joined earlier priorities (functional, social inclusion, self determination) in the curriculum for students with severe disabilities across the country in principle, if not in practice, in all schools.

IDEA 1997 required that all children who receive special education services are to have access to and make progress in the general curriculum, but NCLB and IDEA 2004 and subsequent regulatory language for both laws clarified that the general curriculum was defined as based on the same academic standards and expectations that applied to all other students in a given state. Alternate assessments are to be aligned to (or “linked to” in later terminology related to peer review) the state content standards in each grade.

Alternate Assessment Approaches (Format)

In states’ early development of alternate assessments, most had some type of body of evidence collected over time. Table 3 shows alternate assessment approaches and changes over the time from 2000-2005.

Table 3. Alternate Assessment Approaches 2000-2005

Year	Portfolio or Body of Evidence	Rating Scale or Checklist	IEP Analysis	Other	In Development/ Revision
Regular States					
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*	25 (50%)**	7 (14%)***	2 (4%)	7 (14%)	8 (16%)
Unique States					
2003	4 (44%)	0 (0%)	1 (11%)	1 (11%)	3 (33%)
2005	1 (11%)	1 (11%)	1 (11%)	0 (0%)	1 (11%)

*One state has not developed any statewide alternate assessment approaches.

**Of these 25 states, 13 use a standardized set of performance/events/tasks/skills.

***Of these 7 states, 3 require the submission of student work.

Note. From 2005 state special education outcomes: Steps forward in a decade of change, by S.J. Thompson, C.J. Johnstone, M.L. Thurlow, & J.R. Altman, 2005, Minneapolis: National Center on Educational Outcomes. Reprinted with permission.

State special education directors may have categorized their approaches in varying ways over the years, particularly where there is great overlap in methodology across the nominal types. For example, in 1999 the category “other” specifically included performance assessments. In later years, the category choices became more descriptive, for example, portfolio or body of evidence with or without a standardized set of performance/events/tasks/skills, or a checklist/rating scale with or without a required submission of student work. Some of the changes in categories across the years may reflect changes in how the directors described their assessment, as opposed to real changes in format.

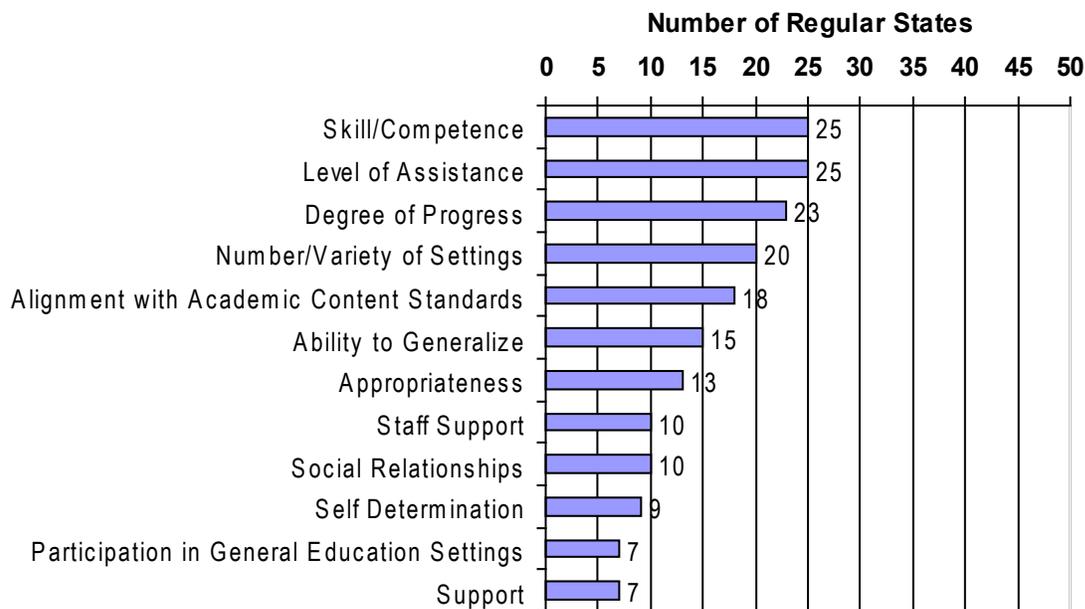
A few trends are very clear. States that formerly required linkages of state alternate assessments to student Individualized Education Programs (IEPs) have shifted away from individualized IEP definitions of assessment targets; numbers of states with alternate assessments in revision or development fell briefly, but rebounded in 2005; and there is a tendency for blurring of format boundaries as portfolios and bodies of evidence add more standardization and checklists/rating scales add more collected evidence of student achievement. The latter tendency relates to issues of scoring, reporting, and accountability that emerged as major issues around the technical defense of alternate assessment as NCLB-required peer review of assessment systems commenced in 2005.

Scoring Criteria and Procedures

By July of 2000, IDEA 1997 required that alternate assessments should be in place. Most states had an initial version of their alternate assessment in place when NCLB was passed. NCLB increased the accountability stakes for schools, districts, and states based on assessment results. The scoring and reporting issues in alternate assessment that states had identified earlier (e.g., Thompson & Thurlow, 2000) became extremely important to solve. At that time, based on what was considered best practices in the 1997 Ysseldyke and Olsen paper, many states still incorporated both student and system performance measures in their scoring rubrics or procedures. Figure 1 shows the use of these student and system measures still in place in 2005.

The first criterion in the list on Figure 1 represents the only criterion that has been without controversy among measurement experts, with lesser agreement on the second and third criteria. These experts believe that since achievement results traditionally reflect independent student performance on content skills and knowledge, all the other criteria are seen as system measures. All of the other criteria reflect research-based understanding of effective teaching for students with severe disabilities, and each can be defended on some level for some purposes. Whether or not these defenses are sustainable for purposes of system accountability is another question that has not been fully answered.

Figure 1. Outcomes Measured by Rubrics on Alternate Assessments - 2005



Note. From 2005 state special education outcomes: Steps forward in a decade of change, by S.J. Thompson, C.J. Johnstone, M.L. Thurlow, & J.R. Altman, 2005, Minneapolis: National Center on Educational Outcomes. Reprinted with permission.

NCEO case studies of five states with varying approaches to alternate assessment, completed in 2003, show a very complex picture of how system performance versus student performance measures were used in scoring state alternate assessments (Quenemoen, Thompson, & Thurlow, 2003). Although the scoring criteria used by the states appeared to be very different, when underlying assumptions and procedures for assessment instrument development were examined—including blueprints—and when analysis of training procedures for gathering evidence or for scoring were reviewed, there were striking similarities in how the varying scoring criteria played out.

The definitions and examples and the side by side examination of the criteria, the scoring elaborations, and the assumed criteria in the design of training materials and assessment format yield a surprising degree of commonality in the way these states define success for students with significant cognitive disabilities. Six criteria are included in all of the five states' approaches in some way, either articulated or assumed. They include "content standards linkage," "independence," "generalization," "appropriateness," "IEP linkage," and "performance." Three scoring criteria are very different across the five states' approaches. They include "system vs. student emphasis," "mastery," and "progress." (Quenemoen, Thompson, & Thurlow, p.iii, 2003)

The notion of “defining success” through rubric construction points to the very real challenge faced by developers of alternate assessments for students with the most significant cognitive disabilities. The scoring criteria that differed in these five states included system versus student emphasis, but the line between the two was difficult to draw. In some states, teachers would provide varying levels of prompting to ensure a student response, and that was viewed in some states as a system measure—the degree to which supports were provided for student learning. In other states levels of prompting were viewed as a student measure—the degree to which the student performed independently. The distinctions between the two were not as clear as the language suggests.

The other scoring criteria that varied among the five states included mastery and progress. The term “progress” is used to define the amount of progress in learning new skills and knowledge from student baseline within the testing year, as opposed to grade-to-grade, or year-to-year progress assumed in growth models. Charting learning progress for students with severe disabilities has been an important long-time teaching and assessment tool. Ysseldyke and Olsen had identified this as an essential challenge in their 1997 assumptions. States continue to grapple with this issue, and the definition of success continues to play out in scoring procedures, and as importantly, in the complexities of defining performance level descriptors and alternate achievement standards for these assessments.

During this time period, states began rethinking who should score the alternate assessments. The requirements for alternate assessments in IDEA 1997 stated that test results for students with disabilities should be publicly reported in the same frequency and format as all other student results, and the Improving America’s School Act (IASA) of 1994 required public reporting of achievement results for all students. Some states built assessment scoring procedures to ensure that common scoring protocols would apply to all assessments, setting up regional or statewide scoring institutes, or contracting with a test publisher for scoring out of state. Other states had teachers score their own students, sometimes on a skills checklist with no evidence required, and other times administering state-developed items or tasks and scoring according to a protocol. Between the 2001 and 2003 NCEO state surveys, state special education directors reported a slight shift from teacher scoring of their own students to centralized scoring (Thompson & Thurlow, 2001, 2003). Other states moved toward more oversight of teacher scoring, including increased requirements for evidence of student work to support ratings or checklist scores, random sampling for verification of the evidence, or videotaping of assessment processes for later review by a neutral trained second scorer. The push for these scoring enhancements was related to increased pressure from NCLB peer review processes, with the expectation that these strategies would result in increased confidence in the accuracy and reliability of scoring processes.

Performance/Achievement Level Descriptors and Standard Setting

Beginning in 2003, the NCEO survey included questions about state plans for setting achievement standards. Regulations allowing states to set alternate achievement standards on alternate assessments designed for students “with the most significant cognitive disabilities” were released in 2003 (U.S. Department of Education, Office of Elementary and Secondary Education, 2003). Although there were a few pioneering states that had already set achievement standards unique to these assessments, NCLB statutory requirements did not permit different content or achievement standards for any students. This new regulation added the option to develop alternate achievement standards using a validated and documented method. These standards had to reflect high expectations for this group of students and align with state content standards. Up to 1% of the total student population in tested grades could be categorized as Proficient using these alternate achievement standards.

Special education directors generally had no experience with the concept or procedures of standard setting, and in states where the special education section was in control of the alternate assessment, the learning curve was very steep. They had just come through a similar steep learning curve as they had grappled with the notion of the general curriculum based on the content frameworks for the state. In many states, special educators assumed that “alternate achievement standards” was a new name for extended content standards of some type. One state assessment coordinator reported that the state special education director had just explained to him that alternate achievement standards in the regulation really meant extended content standards. He wanted an explanation of why the new regulations used the same term for extended content standards that was always associated with performance standards in the regular assessment. The confusion of content standards and achievement standards slowed the field down in the progress on alternate assessments, and many states had false starts before it was all sorted out.

The pattern of responses in 2003 and 2005 to a question of whether states had a standard-setting process in place for their alternate assessment may reflect this confusion. In 2003, 52% of the regular states responded they did, and only 14% said they did not, with 10% saying they didn’t know, along with some reporting an informal process. In 2005, 55% said they did, and were able to name the process. Given the intensive work being done in states in preparation for peer review at that time, we can speculate that perhaps in the 2003 survey, state directors responded “yes” while thinking of their work on extending or expanding the content standards, and the 55% saying “yes” in 2005 actually reflected a larger increase than what the data suggest.

A few states were pioneers in this area. Early standard-setting approaches in states reflected the necessity of adapting existing methods to these new assessments. This early work resulted in three synthesis reports documenting initial efforts (Arnold, 2003; Olson, Mead, & Payne, 2002; Wiener, 2002), and one summarizing the standard-setting approaches that could be tailored to

alternate assessments (Roeber, 2002). The 2003 regulation and the release of Peer Review Guidance in 2004 began a new phase in alternate assessment, as all states began to struggle with the very real challenges of developing “real” large-scale assessments for this small group of students with varying communication requirements and varying learning characteristics. This redoubling of efforts to build technically defensible assessments was also in response to another related key demand: use of the assessments in NCLB required reporting and accountability systems.

By 2005, discussions about the format of the assessment approach had dropped from being the primary focus of change, and more states were looking at enhancing the approach through working on refining content targets, better understanding achievement standards, and ensuring integrity in scoring. Table 4 shows that twice as many states (17) were concerned about scoring criteria being improved than were identifying the format as their primary issue (8).

Table 4. Alternate Assessment Development/Revision: Focus of Change Efforts - 2005

Focus of Change Efforts on Alternate Assessment	Number of Regular States
Approach	8
Content	10
Standard-setting	13
Scoring Criteria	17

Note. From 2005 state special education outcomes: Steps forward in a decade of change, by S.J. Thompson, C.J. Johnstone, M.L. Thurlow, & J.R. Altman, 2005, Minneapolis: National Center on Educational Outcomes. Reprinted with permission.

Reporting and Accountability

Challenges in reporting of alternate assessment results had been identified in the 2000 Thompson and Thurlow report, and the NCLB requirements that all student results had to be included in system accountability measures intensified the challenges and raised the stakes. State work on the development of alternate achievement standards was an essential step in including all scores in accountability calculations. By 2001, stakeholders across the country were seeing positive consequences for students with disabilities related to their inclusion in accountability systems, although some challenges were identified (Quenemoen, Lehr, Thurlow, & Massanari, 2001). Quenemoen et al. (2001) summarized the conclusions of 135 stakeholders from 39 states (plus American Samoa and the Bureau of Indian Affairs) who participated in a structured discussion of issues related to implementation of alternate assessments. Among the findings was:

Technical and psychometric difficulties with existing assessment systems were perceived as a major issue, but fairness of use of results is a related and complicating issue. Some of the challenges identified by participants include: putting all students on the same scale versus accountability for all, a need for a balance between what makes sense for improvement planning versus psychometric soundness, and how to compare fairly across schools, districts, and states with so many uncontrolled variables. (Quenemoen et al., 2001, pp. 5-6)

Two synthesis reports dealt with issues and methods of reporting of alternate assessment scores just as NCLB was authorized (Bechard, 2001; Quenemoen, Rigney, & Thurlow, 2002), but the larger issue remained how to defend the technical adequacy of the assessment results for reporting and accountability purposes.

The Transition to New Thinking

As the field continued to struggle with the issues, it became clear that retrofitting alternate assessments for this group of students into existing measurement paradigms, using traditional statistical methods of documenting technical qualities, was not working well. At the 2004 American Educational Research Association Annual Meeting, a paper that described the chasm between traditional measurement tools and the challenges of alternate assessment for students with significant cognitive disabilities stimulated discussion across measurement, curriculum, and special education partners (Quenemoen, Thurlow, & Ryan, 2004). It resulted in the recognition that the challenges of alternate assessment were not going to be solved with the expertise and tools of one educational discipline alone. These challenges required collaboration that would yield educationally sound but technically defensible strategies.

In 2001, the National Research Council had sponsored a Committee on the Foundations of Assessments “to look at the advances in the cognitive and measurement sciences, as well as early work done in the intersection between the two disciplines, and to consider the implications for reshaping educational assessment” (National Research Council, p. xii, 2001). Large-scale assessment and special education colleagues around the country began investigating the application of the Committee’s work to state assessment systems. Through two Federal grant opportunities, a research collaborative was formed that consisted of experts in special education (including severe disabilities), curriculum, and measurement, and a dozen partner states. The New Hampshire Enhanced Assessment Initiative (NHEAI) and the National Alternate Assessment Center (NAAC) funding allowed this partnership the luxury of working as a team to identify key issues in developing technically defensible alternate assessments for use in NCLB required accountability systems.

By early 2008, 10 states had partnered with NHEAI and NAAC to apply this framework to their own alternate assessment. The framework has proven useful as a practical tool to identify recommendations for areas where states may need new approaches to document the structure and function of their assessments. In preliminary analyses of application of the framework by the NHEAI/NAAC expert panels to the second group of partner states (personal correspondence among the analysis and writing team of Rachel Quenemoen, Jacqui Kearns, and Scott Marion, June 2008), it appears that even when the approaches to alternate assessment (e.g., portfolio, checklist, performance assessment) vary dramatically, common issues arise even though the solutions may be somewhat different. For example, in all six states that were reviewed by the experts, teacher/administrators were identified as a source of measurement error that needs careful study. For performance assessments and checklists, there is a need to uncover response processes on the part of the teacher: that is, are teachers developing appropriate tasks and applying scoring procedures as the developers intended? In portfolio assessment, student work is provided as well as a description of the task, and scoring is generally done by someone other than the teacher. The needs in this case tend to be on the appropriateness of the content targets chosen for the student, and the implementation of the task.

It also appears that in initial analyses, these recommendations will contribute to new alternatives to some traditional methods of documenting the technical qualities of these assessments, building on the work of Kane (2002) and others (e.g., Cronbach, 1988). This is important given that the small numbers of students who participate in these assessments, and the heterogeneity of their learning characteristics, means that the underlying assumptions for use of some traditional methods are not met. Three organizations are partnering on developing these initial findings into white papers and articles. These works in progress can be found on the Web sites for the National Alternate Assessment Center (www.naacpartners.org), the National Center for the Improvement of Educational Assessment (www.nciea.org), and the National Center on Educational Outcomes (www.nceo.info).

Current Status of Alternate Assessments based on Alternate Achievement Standards

The current status of alternate assessments is reflected in the work done by NHEAI and NAAC and the states that are partnering with them to test their frameworks. The same themes that NCEO surveys have covered in the past are being addressed.

Stakeholders, Expectations, and Principles

The National Alternate Assessment Center has developed and validated a tool to capture the learning characteristics of students who participate in alternate assessment based on alternate

achievement standards, the Learner Characteristics Inventory (LCI) (Kearns, Towles-Reeves, Kleinert, & Kleinert, 2006). They have conducted this survey in multiple states, with extensive analyses for four states completed (Towles-Reeves, Kearns, Kleinert, & Kleinert, in press). The data are remarkably similar across states.

What is alarming in these data is that in most states for which there are data, there is no meaningful progression of skills from elementary to high school levels. While these data are cross-sectional and not longitudinal (and are thus not tracking the same students over time), Kearns and Towles-Reeves suggest this reflects the history of low expectations for this group of students, and a historical “gold standard” that holds sight words and use of calculators as the ultimate end-goal of academic instruction for these students (Kearns & Towles-Reeves, 2007). Even more alarming are the data that show that the percentage of students who do not have meaningful communication strategies does not change from elementary to high school levels (Towles-Reeves et al., 2008). Not only are these students not making progress in the academic content, they apparently are not even able to access the content through communication tools, high tech or low.

Some states have reported a sharp rise in use of assistive technology following implementation of alternate assessments. If that is followed by a decrease in the percentages of students who do not have a communication strategy, it would be a powerful endorsement of the positive consequences of alternate assessment on raising expectations and outcomes.

There are other data that suggest expectations have not as yet risen universally. In 1999, stakeholders estimated that from less than 1% to more than 9% of all students had such limited exposure to content that it would prevent them from participating in regular assessment (see Table 1 of this report). In 2007, with the advent of a second NCLB regulation allowing another separate achievement standard, the 2007 “2% regulation,” data from state public reports (e.g., IDEA required Annual Performance Reports) show that from less than 1% to as high as 9% of all students participate in various alternate assessments in states. These percentages are of the total student population, and depending on individual state incidence figures, that could represent as high as 90% of all students with disabilities. These data are based on the assumption that 10% of the entire student population has a disability, on average, and that states accurately reported the percentages as percent of all students, and assuming that all alternate assessment options are included in their estimates, including those on alternate, modified, and grade-level achievement standards.

States are exploring options for alternate assessments based on modified achievement standards, and several states have already developed them, but have not completed peer review (Lazarus, Thurlow, Christensen, & Cormier, 2007). Given national incidence figures showing that 85% of all students with disabilities ages 6 through 21 do not have cognitive disabilities (Cortiella,

2007), it is disheartening to see so many students being held to alternate and modified achievement standards.

Content Coverage (Linkage to Content Standards)

Since 2004, NAAC at the University of Kentucky included content issues in alternate assessment as one of three research foci. NAAC's University of Kentucky partners continue working to define what linkage to grade-level content means in practice. They have developed national training on tools that help states determine appropriate content targets, focusing on available student work as the field changed (National Alternate Assessment Center, 2005). "Is it reading? Is it math? Is it science?" training materials are posted on their Web site at <http://www.naacpartners.org/products.aspx>. As part of the NHEAI joint work with NAAC, Kleinert, Browder, and Towles-Reeves (in press) developed a white paper summarizing extant literature on a theory of learning for students with disabilities as compared and contrasted to the literature base on learning theory in the National Research Council's *Knowing What Students Know*. NAAC partners at the University of North Carolina Charlotte (UNCC) meanwhile developed and validated a procedure for alignment studies on alternate assessments for students with significant cognitive disabilities, called Links for Academic Learning (LAL) (Flowers, Wakeman, Browder & Karvonen, 2007; Flowers, Wakeman, Browder & Karvonen, in press).

Although these tools have been developed over the past few years, states were required to have their state systems ready for peer review under NCLB requirements prior to tool validation. Results from peer review to date suggests great variability of content coverage—what the UNCC researchers called near and far linkages—including several states that still included broken links. A few still reflect a one-size-fits-all functional or very low level academic curriculum reminiscent of the infant/early childhood curriculum of years ago, but most states are moving away from functional targets. Some states are still struggling with designing curriculum and assessments that do not extend the standards so far as to lose the integrity of the grade-level content standards, particularly for students with the most significant challenges, those at a pre-symbolic level of communication use (personal communication with Claudia Flowers, June, 2008). Even so, there is a clear and steady trend toward more challenging academic content as more states implement alternate assessments more strongly linked to grade-level academic content standards.

As work continues on instructional outcomes for these students, we are learning more about how to ensure appropriately challenging and accessible learning targets. The UNCC researchers are working on instructional issues as well as assessment issues, and are finding that these students can indeed learn challenging academic content the field did not think possible in the past (Browder, Gibbs, Ahlgrim-Delzell, Courtade, Mraz, & Flowers, in press). They propose a

conceptual foundation for early literacy instruction (literacy includes the early skills and components of reading) that includes “accessing books” through “story based lessons.” These and other research projects will help firm up our conceptions of the construct of reading for students with significant cognitive disabilities.

Alternate Assessment Approach (Format)

Several NCEO reports have called attention to the degree to which nominal categories of alternate assessment approach (e.g., portfolio, performance assessment) are not particularly useful descriptors (e.g., Gong & Marion, 2006; Quenemoen, Thompson, & Thurlow, 2003; Thompson & Thurlow, 2000). The Gong and Marion (2006) report is devoted to this topic, after the NHEAI and NAAC expert panel drew attention to the fact that nominal categories are not useful for characterizing the technical aspects of the assessment. The expert panel’s technical review of partner state alternate assessments demonstrated that the evaluation of technical adequacy interacts with the types of alternate assessments being employed, but the types were better described along a continuum of standardization and flexibility in design choices rather than as nominal types. Gong and Marion caution that this does not mean that standardization is good and flexibility is bad. Designing assessments to coherently link the nature of cognition to observation and to intended inferences for this small group of students does not lend itself to rigid standardization.

This complexity of design issues is not limited to alternate assessments. In her 2007 AERA presidential address, Eva Baker suggests, “Tests only dimly reflect in their design the results of research on learning, whether of skills, subject matter, or problem solving. These test-design properties matter to researchers but rarely are observable in the tests because the naked eye is drawn to test format, not educational soundness” (Baker, 2007, p. 310). The work of NHEAI and NAAC was meant to focus on educational soundness, not format, and the Gong and Marion 2006 report includes concepts and tools to help states do so as well.

Scoring Criteria and Procedures

As discussed above, there are many unanswered questions about what scoring criteria are appropriate for use with alternate assessments of students with significant cognitive disabilities. Basic questions remain:

- How can scoring protocols be designed and carried out with fidelity when tasks need to be adapted across such a broad range of student communication methods?

- How do we measure degree of independence in responses for students with limited response repertoires?
- How do we account for traditional understanding of baseline growth in a standards-based system?
- Who administers items or tasks and then scores responses when many of these students respond only to familiar test administrators?
- Who checks, and how do we verify that consistent administration and scoring is occurring?

Design of scoring rubrics and procedures, along with design of tasks, are among the greatest challenges that states face as they balance the need for flexibility versus standardization with the unusual and varied learning characteristics of the students.

Performance/Achievement Level Descriptors and Standard Setting

Scoring and task decisions ultimately need to be driven by how proficiency is defined for these students. Here again, basic questions still remain. What should these students know and be able to do? How well? Is the content clearly referenced? How good is good enough?

NAAC has developed a paper summarizing the issues of alternate assessment that provides a framework for states to use to answer these questions (Perie, 2007). The paper emphasizes the importance and challenges of writing detailed alternate achievement level descriptors that clearly link to the grade level content standards while also reflecting performance expectations, and that also address the context of any system supports that the students require, including level of prompting. States have struggled to accurately represent what the student performance actually means. The nature of the link to grade-level content that is appropriate for students with significant cognitive disabilities, and that is also appropriately challenging and consistent with what similar age peers are learning, has been both praised and ridiculed. States need to grapple with precise language that describes exactly what is and is not represented by various proficiency determinations, or the credibility of alternate assessments will be suspect. Understanding and describing clearly what success in academic content is for these students, and then matching those descriptions to test results is very, very difficult. The actual standard-setting procedures described in the Perie (2007) paper and those used in many states thus far are relatively straightforward by comparison. Because we understand so little about what students with significant cognitive disabilities know and can do in academic content when taught well and given the support to communicate effectively, we can anticipate dramatic changes in what proficiency means for these students. Initial descriptions and standards will need careful monitoring and adjusting over time.

Reporting and Accountability

Public reporting requirements of participation and performance of all students is defined in both NCLB and IDEA. NCEO has been compiling IDEA required reporting on state annual performance reports, in addition to reporting on assessment data that are publicly reported by states. It is clear from these reports that some states are struggling to provide clean and clear data on the participation and performance of students with disabilities in the assessment system in either type of report. Some of the struggle comes from limited capacity for data management or communication across divisions in some states, but we still do not have readily comparable data on the participation and performance of students with disabilities across all 50 states, including those students who participate in alternate assessments of all types.

The lack of clarity about participation and performance on alternate assessment carries across the entire alternate assessment effort. It is far more difficult to quickly peruse a state's alternate assessment description and materials and judge quality from the outside than it is for regular assessments. In NCEO's systematic analyses of state alternate assessments during the past decade, it is clear that alternate assessments sometimes are more or less than meets the eye on first glance. A primary reason for this lack of clarity is the number of unknowns that still remain in the field about what these students can know and do when they are taught well in the academic content. The technical issues of these new assessments are huge, but until we build a common understanding of the learning characteristics of these students, how they can be expected to learn in the academic domains, and what their performance looks like when they have been taught well, the technical efforts are simply an attempt to put order on rapidly shifting chaos.

Considerations for State Practice

State departments of education must move forward regardless of chaos or clarity. There are several strategies for states to consider as they continue efforts to, as was commonly expressed a decade ago about alternate assessments, "build the plane while we are flying." Because of the number of uncertainties still in play, we need:

1. **Transparency.** We do not know as yet what will work the best in teaching and in assessing students with significant cognitive disabilities in the academic content. We are seeing evidence of remarkable achievement, but this group is so varied in characteristics and the field of severe disabilities is still divided on what appropriate outcomes we can and should expect. It is appropriate that states vary so much in their assessment practices at this point, even appropriate that the content targets of alternate assessment are still taking so much time and struggle to refine. The key to resolving this lack of clarity is transparency of processes and outcomes. We need to

know what varying practices and targets yield for student outcomes, and the only way to build that knowledge base is to ensure that assessment development, implementation, and results are transparent and open to scrutiny. Although quantitative approaches to outcome measures are valued in general assessment, as are statistical approaches to documentation of technical quality, in order for the numbers to tell us something we have to know what the desired assessment processes and outcomes are. We do not know this, as yet, for students with significant cognitive disabilities.

2. **Integrity.** Building on the need for transparency is the need for integrity. The amount of flexibility needed to ensure that all students can demonstrate what they know and can do is higher in alternate assessments for this group of students than in more typical student populations. Flexibility can mask issues of teaching and learning unless it is carefully structured and controlled. Research on teachers' ability to assess and score their own students' work with fidelity and integrity is limited. Research from the 1980s suggests that teachers can predict which items of a norm-referenced test their typical students will get right (e.g., Colardarci, 1986; Hoge & Colardarci, 1989). In the 1986 Colardarci study, teachers were right in their item-level judgments more often than not, but accuracy was higher for some tasks than others, for example, computation versus problem solving (mathematics), literal versus figurative meaning (reading). Teachers were more accurate with higher-ability students than with lower-ability students. According to David Niemi, research on teacher scoring of performance assessments (at the National Center for Research on Evaluation, Standards, and Student Testing) suggests that teachers can be trained to reliably score work other than their own students' (e.g., writing assessments), but it is less likely that they will score their own students' work as reliably (personal communication, March 15, 2007). For students with significant cognitive disabilities, we have not built a shared understanding in the field of what acceptable performance is in the academic domains at each level, nor do we understand how varying prompting approaches affect the content being assessed, so teacher self-scoring remains a murky issue.

Similarly, standardization as a solution risks reducing the integrity of the assessment results when the methods do not match the population being assessed and how that population demonstrates competence in the academic domains. Given the uncertainties of what can be expected for these students, and the small numbers of students with highly varying learning characteristics in most states, many traditional tools of large-scale assessment development and documentation are of limited use. It is tempting to make use of tidy and traditional solutions for technical defense, but when the underlying assumptions of testing models and tools are not met, it is inappropriate to use them. Brennan (1998), in his NCME address commented:

In general, strong assumptions lead to strong results. . . . However, a claim that a model solves a thorny measurement problem is credible only to the extent that the assumptions engaged in addressing the problem can be shown to withstand serious challenge.

Too frequently, in my opinion, we act as if assumptions are met without question. Such unrestrained confidence can easily lead to excessive (or at least unsubstantiated) public claims about what our models can accomplish in real life educational testing contexts (pp. 5-6).

For example, one concern is the use of internal consistency reliability coefficient as a central piece of reliability evidence for alternate assessment scores. Some alternate assessments have few items/tasks, which are evaluated using a rubric designed to be rated holistically on different dimensions. The purpose and context of an assessment should determine the reliability value to apply and the degree of reliability required (Parkes, 2007). Cronbach's alpha may serve some value in examining the internal consistency of the alternate assessments items/tasks; however, designing reliability methodology that moves beyond sampling theories and dimensionality assumptions and focuses on conceptual-structural replications are needed to fully evaluate alternate assessment reliability issues.

3. **Validity studies.** Building on the issues of transparency and integrity, we have an obligation to monitor carefully the effects of alternate assessments over time, as well as to ensure the claims we are making for the use of the results are defensible. Several states are currently designing and carrying out validity studies as part of the General Supervision Enhancement Grants offered by the United States Department of Education's Office of Special Education Programs. These approaches can serve as models for all states as we work to understand whether claims based on alternate assessment results are warranted. We cannot afford to "hope" that our initial guesses of what will work to improve outcomes for these students will play out as we intend. We have less than two decades of experience in large-scale alternate assessment of these students and even less in understanding how they build competence in mathematics, reading, and science.

4. **Planned improvement over time.** In building a validity argument, we study whether the interpretations and uses of the test are defensible, and whether consequences that are hoped for and those that are to be avoided are in fact falling into their respective places. An important part of validity studies is the ongoing day-to-day oversight of the assessment development, implementation, and use of testing results, and high quality data collection and continuous improvement based on the data are absolutely necessary for these assessments. Several states have good examples of this kind of continuous improvement process in their state documentation. These states have built in data collection to routine assessment procedures to allow them to identify problems and address them year by year.

Conclusion

Why does it matter? The 1997 IDEA legislation was pivotal in changing expectations for students with disabilities. The preamble to the 1997 reauthorization stated, “Almost 20 years of research and experience has demonstrated that the education of children with disabilities can be made more effective...” Unfortunately, the preamble to the 2004 reauthorization includes different words simply by the addition of another decade of neglect: “Almost 30 years of research and experience has demonstrated that the education of children with disabilities can be made more effective by-- (A) having high expectations for such children and ensuring their access to the general education curriculum in the regular classroom, to the maximum extent possible, in order to-- (i) meet developmental goals and, to the maximum extent possible, the challenging expectations that have been established for all children; and (ii) be prepared to lead productive and independent adult lives, to the maximum extent possible...”

What is the “maximum extent possible”? We have learned that we have expected too little of students with significant cognitive disabilities in the past, but they still have much to teach us about what is possible. States can design their alternate assessments to reflect what we know and believe about these students and their learning, appropriately raising the bar for the students and their teachers. States can do so by building on what we have learned during the past decade, and ensuring that the process and outcomes of their approach to alternate assessment are transparent and subject to review, stand up to both technical and ethical scrutiny, push practices and outcomes in the expected and desired directions, and can be improved through data-based oversight over time.

References

- Arnold, N. (2003). *Washington Alternate Assessment System technical report on standard setting for the 2002 portfolio* (Synthesis Report 50). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.
- Baker, E. L. (2007). 2007 Presidential address: The end(s) of testing. *Educational Researcher*, 36 (6), 309-317.
- Bechard, S. (2001). *Models for reporting the results of alternate assessments within state accountability systems* (Synthesis Report 39). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes
- Brennan, R. L. (1998). Misconceptions at the intersection of measurement theory and practice. *Educational Measurement: Issues and Practice*, 17 (1), 5-9, 30.
- Browder, D. M., Ahlgrim-Delzell, L., Courtade, G., Gibbs, S. L., & Flowers, C. (in press). Evaluation of the effectiveness of an early literacy program for students with significant developmental disabilities using group randomized trail research. *Exceptional Children*.
- Browder, D., M., Gibbs, S., Ahlgrim-Delzell, L., Courtade, G., Mraz, M., & Flowers, C. (in press). Literacy for students with significant cognitive disabilities: What should we teach and what should we hope to achieve? *Remedial and Special Education*.
- Browder, D.M., & Spooner, F. (Eds.). (2006). *Teaching language arts, math, & science to students with significant cognitive disabilities*. Baltimore, MD: Brookes Publishing.
- Brown, L., Nietupski, J., & Hamre-Nietupski, S. (1976). The criterion of ultimate functioning and public school services for severely handicapped children. In M. Thomas (Ed.), *Hey, don't forget about me!* Reston, VA: Council for Exceptional Children.
- Colardarci, T. (1986). Accuracy of teacher judgments of student responses to standardized test items. *Journal of Educational Psychology*, 78 (2), 141-46.
- Cortiella, C. (2007). *Rewards and roadblocks: How special education students are faring under No Child Left Behind*. New York: National Center for Learning Disabilities.
- Cronbach, L. (1988). Five perspectives on validity argument. In H. Wainer (Ed.) *Test validity* (p. 3-17). Hillsdale, NJ: Erlbaum

Flowers, C., Wakeman, S., Browder, D., & Karvonen, M. (2007). *Links for academic learning: An alignment protocol for alternate assessments based on alternate achievement standards*. Charlotte, North Carolina: University of North Carolina at Charlotte. Retrieved from <http://www.nceo.info>.

Flowers, C., Wakeman, S., Browder, D., & Karvonen, M. (in press). An alignment protocol for alternate assessments based on alternate achievement standards. *Educational Measurements: Issues and Practice*.

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.

Hoge, R.D., & Colardarci, T. (1989). Teacher-based judgments of academic achievement: A review of literature. *Review of Educational Research*, 59 (3), 297-313.

Kane, M. (2002). Validating high-stakes testing programs. *Educational Measurement: Issues and Practices*, 21 (1), 31-41.

Kearns, J., & Towles-Reeves, E. (2007). Kearns, J., & Towles-Reeves, E. (2007). *Alternate assessments on alternate achievement standards student population*. Presentation at University of Maryland MARCES Conference, October 11, 2007. Retrieved June 5, 2008 from http://www.marces.org/conference/alt_assessment/07main.htm

Kearns, J., Towles-Reeves, E., Kleinert, H., & Kleinert, J. (2006). *Learning characteristics inventory report*. Lexington, Kentucky: University of Kentucky, National Alternate Assessment Center.

Kleinert, H., Browder, D., & Towles-Reeves, E. (in press). Models of cognition for students with significant cognitive disabilities: Implications for assessment. *Review of Educational Research*.

Kleinert, H., Haigh, J., Kearns, J., & Kennedy, S. (2000). Alternate assessments: Lessons learned and roads to be taken. *Exceptional Children*, 67 (1), 51-66.

Kleinert, H., & Kearns, J. (1999). A validation study of the performance indicators and learner outcomes of Kentucky's alternate assessment for students with significant disabilities. *Journal of the Association for Persons with Severe Handicaps*, 24 (2), 100-110.

Kleinert, H., Kennedy, S., & Kearns, J. (1999). Impact of alternate assessments: A statewide teacher survey. *Journal of Special Education*, 33 (2), 93-102.

Kohl, F., McLaughlin, M., & Nagle, K. (2006). Alternate achievement standards and assessments: A descriptive investigation of 16 states. *Exceptional Children*, 73, 107-123.

Lazarus, S. S., Thurlow, M. L., Christensen, L. L., & Cormier, D. (2007). *States' alternate assessments based on modified achievement standards (AA-MAS) in 2007* (Synthesis Report 67). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.

Marion, S., & Pellegrino, J. (2006). A validity framework for evaluating the technical quality of alternate assessments. *Educational Measurement: Issues and Practice*, 25 (4), 47-57.

Marion, S.F., Quenemoen, R.F., & Kearns, J.F. (2006). Introductory presentation to October 2006 Seminars on Inclusive Assessments. Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.

National Alternate Assessment Center. (2005). Part III: Theory of learning: What students with the most significant cognitive disabilities should know and be able to do. *Access and Alignment to Grade Level Content for Students with the Most Significant Cognitive Disabilities: A Training Module for Large-scale Use*. Retrieved October 30, 2007, from <http://www.naacpartners.org/Products/Pre/slide5.htm>

National Research Council (2001). *Knowing what students know: The science and design of educational assessment*. Washington, DC: National Academy of Sciences.

New Hampshire Enhanced Assessment Grant, National Alternate Assessment Center, & National Center for the Improvement of Educational Assessment. (2006a, October). *Documenting the technical quality of your state's alternate assessment system: An annotated workbook: Volume I: "Nuts and bolts"*. Lexington, KY: University of Kentucky, National Alternate Assessment Center.

New Hampshire Enhanced Assessment Grant, National Alternate Assessment Center, & National Center for the Improvement of Educational Assessment. (2006b, October). *An annotated workbook for documenting the technical quality of your state's alternate assessment system: Volume II: The validity evaluation*. Lexington, KY: University of Kentucky, National Alternate Assessment Center.

Olson, B., Mead, R., & Payne, D. (2002). *A report of a standard setting method for alternate assessments for students with significant disabilities* (Synthesis Report 47). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.

Parkes, J. (2007). Reliability as argument. *Educational Measurement: Issues and Practice*, 26 (4), 2-10.

Perie, M. (2007). *Setting alternate achievement standards*. Lexington, KY: University of Kentucky, National Alternate Assessment Center, Human Development Institute.

Quenemoen, R. F., Lehr, C. A., Thurlow, M. L., & Massanari, C. B. (2001). *Students with disabilities in standards-based assessment and accountability systems: Emerging issues, strategies, and recommendations* (Synthesis Report 37). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.

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.

Quenemoen, R., Thompson, S., & Thurlow, M. (2003). *Measuring academic achievement of students with significant cognitive disabilities: Building understanding of alternate assessment scoring criteria* (Synthesis Report 50). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.

Quenemoen, R., Thurlow, M., & Ryan, J. (2004). *I say potato, you say potahto: An AERA Conference discussion paper and side-by side glossary*. Minneapolis: National Center on Educational Outcomes.

Roeber, E. (2002). *Setting standards on alternate assessments* (Synthesis Report 42). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.

Shriner, J.G., & Thurlow, M.L. (1993). *State special education outcomes, 1993*. Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.

Thompson, S., & Thurlow, M. (1999). *1999 State special education outcomes: A report on state activities at the end of the century*. Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.

Thompson, S. J., & Thurlow, M. L. (2000). *State alternate assessments: Status as IDEA alternate assessment requirements take effect* (Synthesis Report No. 35). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.

Thompson, S., & Thurlow, M. (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.

Thompson, S. J., Johnstone, C. J., Thurlow, M. L., & Altman, J. R. (2005). *2005 state special education outcomes: Steps forward in a decade of change*. Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.

Towles-Reeves, E., Kearns, J., Kleinert, H., & Kleinert, J. (in press). Knowing what students know: Defining the student population taking alternate assessments based on alternate achievement standards. *Journal of Special Education*.

U.S. Department of Education, Office of Elementary and Secondary Education. (December, 2003). *Title I--Improving the academic achievement of the disadvantaged, final regulations*. Washington, DC: U.S. Department of Education.

Wiener, D. (2002). *Massachusetts: One state's approach to setting performance levels on the alternate assessment* (Synthesis Report 48). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.

Wiener, D. (2005). *One state's story: Access and alignment to the GRADE-LEVEL content for students with significant cognitive disabilities* (Synthesis Report 57). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.

Ysseldyke, J. E., & Olsen, K. R. (1997). *Putting alternate assessments into practice: What to measure and possible sources of data* (Synthesis Report No. 28). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.

Ysseldyke, J., Thurlow, M., Erickson, R., Gabrys, R., Haigh, J., Trimble, S., & Gong, B. (1996). *A comparison of state assessment systems in Maryland and Kentucky with a focus on the participation of students with disabilities* (Maryland-Kentucky Report 1). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.

Ysseldyke, J.E., Thurlow, M.L., McGrew, K.S., & Shriner, J.G. (1994). *Recommendations for making decisions about the participation of students with disabilities in statewide assessment programs: A report on a working conference to develop guidelines for statewide assessments and students with disabilities* (Synthesis Report 15). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.