EDUCATIONAL POLICY REFORM RESEARCH INSTITUTE

Policy Symposium Proceedings
February 4-6, 2004

Ensuring Accountability for All Children in an Era of Standards-Based Reform: Alternate Achievement Standards

Arlington, VA

Table of Contents

Introduction

The Educational Policy Reform Research Institute (EPRRI) and the US Department of Education's Office of Special Education Programs (OSEP) held a policy symposium entitled "Ensuring Accountability for All Children in an Era of Standards-Based Reform: Alternate Achievement Standards" February 4-6, 2004 at the Crystal City Marriott in Arlington, Virginia. The purpose of the symposium was to open a dialogue on alternate achievement standards, content standards, and assessments with regard to the December 9th regulations of *The No Child Left Behind Act of 2001*. Participants included personnel from the U.S. Department of Education, representatives from state and local education agencies, and researchers in the field of special education.

A number of lectures and presentations concentrating on alternate achievement standards and alternate assessments took place during the symposium. Dr. David Malouf gave a history of alternate assessments. Stephanie Lee and Christine Wolfe provided an overview of *No Child Left Behind* as it relates to students with disabilities. Dr. Martha Thurlow proposed universal design as a promising approach for expanding students' access the general education curriculum and assessments. Dr. Jacqueline Kearns reported on current research studying the benefits of digitally accessible learning and assessment for students with disabilities. Dr. David Rose demonstrated and discussed general education computer software programs that have built-in accommodations to help students with disabilities access the general education curriculum. Dr. Margaret McLaughlin presented an EPRRI white paper examining the development of meaningful and valid accountability systems for students with disabilities. A panel of state and school district representatives answered questions concerning the process and purposes of their states' established alternate achievement standards. Summaries of the lectures and presentations are provided in the following symposium proceedings.

A large portion of the three day symposium was spent in structured discussion groups. One discussion group session addressed questions about alternate achievement standards and assessments. The groups responded to questions of how to align alternate achievement standards with content standards, how to make rigorous alternate achievement standards, and how to ensure the technical adequacy of alternate assessments. Another discussion group session focused on determining what technical assistance states and school districts need to establish and implement appropriate alternate achievement standards and assessments within the overall accountability system. In addition, each group identified specific implications and tasks for OSEP. A third discussion group session was held to compile suggestions for areas of research and development related to alternate achievement standards and assessments. The discussions from these sessions are summarized in these proceedings.

Ensuring Accountability for All Children in an Era of Standards-Based Reform: Alternate Achievement Standards
February 4-6, 2004

Purpose

The purpose of this symposium was to provide a forum in which to share knowledge, perspectives, and ideas, and to develop guidance for states and districts on how to establish alternate achievement standards and alternate assessments within the regulations of *No Child Left Behind.* A parallel purpose was to discuss and identify elements and processes still needed to reach our "vision" for students with disabilities under the December 9th regulations of *No Child Left Behind*.

Participants

A wide variety of interests and expertise were represented at this symposium:

Researchers:

- Diane Browder – Professor, University of North Carolina, Special Education Program
- Jacqueline Farmer Kearns – ILSSA, University of Kentucky
- David Rose – Co-Executive Director, CAST, Massachusetts
- James Shriner – Associate Professor, University of Illinois, Department of Special Education
- Sandy Thompson – Research Associate, National Center for Education Outcomes, Minnesota
- James Ysseldyke – Professor and Associate Dean for Research, University of Minnesota
- Carol Allman – Consultant, American Printing House for the Blind, Tallahassee, Florida
- Martha Thurlow – Director, National Center for Education Outcomes, Minnesota
- Margaret McLaughlin – Co-Principal Investigator, Educational Policy and Reform Research Institute, Maryland
- David Riley – Executive Director, Urban Special Education Leadership Collaborative, Massachusetts

State Departments of Education and School Representatives:

- Sharon Hall – Section Chief, Maryland State Department of Education
- Leslie Lightbourne – Education Program Consultant, Louisiana Department of Education
- Alexa Pochowski – Assistant Commissioner, Kansas State Department of Education

- Terri Rogers Connolly – Supervisor, Exceptional Student Services Unit, Colorado Department of Education
- Daniel Wiener – Assessment Coordinator, Students with Disabilities, Massachusetts Department of Education
- Kai Graves – MA DOE MCAS/ALT Teacher Training Specialist, Springfield Public Schools, Massachusetts
- Patricia Longo – Regional Director, Instructional Support Services, Douglas County, Colorado

US Department of Education Representatives:

- Cynthia Bryant – Education Program Specialist, Office of Special Education Programs
- Teresa Cahalan – Office of Deputy Secretary, Budget Services
- Louis Danielson – Director, Research to Practice Division, Office of Special Education Programs
- Patty Guard – Deputy Director, Office of Special Education Programs
- Kelly Henderson – EPRRI Project Officer, Office of Special Education Programs
- Jackie Jackson – Office of Elementary and Secondary Education
- Troy Justesen – Acting Deputy Asst Secretary, Special Education and Rehabilitative Services, Office of Special Education Programs
- Kristen Lauer – Office of Special Education Programs
- David Malouf – Research Analyst, Office of Special Education Programs
- Darla Margurger – Office of Elementary and Secondary Education
- Mike Monfore – Office of Special Education Programs
- Kay Rigling – Office of General Council
- Sue Rigney – Office of Elementary and Secondary Education
- Ruth Ryder – Director, Division Monitoring and State Improvement Planning, Office of Special Education Programs
- Stephanie Smith Lee – Director, Office of Special Education Programs
- Wendy Tada – Office of Special Education Programs
- Laurie Wise – Office of Special Education Programs
- Christine Wolfe – Director of Policy

Wednesday, February 4, 2004: Reception and Introduction

An opening reception was held to greet the symposium participants. Dr. David Malouf, Research Analyst from the Office of Special Education Programs, provided a historical perspective of alternate assessments. State and district representatives gave brief overviews of how they established alternate achievement standards.

Thursday, February 5, 2004:

Opening Remarks

Troy Justesen stated how important EPRRI's work is to the issues related to students with disabilities surrounding *No Child Left Behind* and the reauthorization of *IDEA.*

Stephanie Smith Lee proposed a close examination of how both general and alternate standards and assessments can fit into a universal design, which she promoted as a way of providing access for all students. Although the Department of Education does not tell the states what to do as far as setting standards, it would like to provide some guidance to the states and identify best practices. She emphasized that alternate standards must promote students' progress through the curriculum towards grade-level standards. The ultimate goal is to have truly universally designed standards, curriculum, instruction, and assessments that are aligned with each other so that all students will benefit.

Overview and key requirements of the U.S. Department of Education's final regulations concerning assessments based on Alternate Achievement Standards

Presentation by Stephanie Lee and Christine Wolfe, Director of Policy, Office of the Under Secretary at the U.S. Department of Education

No Child Left Behind (NCLB) aims to close the achievement gap and to ensure that all students, including those with disabilities, those from minority backgrounds, and linguistically diverse students, achieve academic proficiency. It requires that all children be assessed on challenging, state-developed content and achievement standards in grades 3-8 and once in grades 10-12. Alternate achievement standards must be aligned to the content standards. Adequate Yearly Progress (AYP), defined individually by states to establish state, district, and school accountability, includes benchmarks for percent of students, both in the aggregate and by subgroup, expected to demonstrate proficiency.

The safe harbor provision within NCLB takes into account progress towards meeting AYP criteria. Specifically, the safe harbor measure protects states, districts, and schools that do not make AYP but show, within a subgroup, a 10% decrease of students not making AYP, and an improvement on other academic indicators. The safe harbor indicator is determined by the states.

NCLB allows for a range of tools and methods to assess students with disabilities. Regulations require assessments to be accessible and valid for the widest range of

students, thus promoting application of universal design. Children with disabilities may take regular assessments, with or without accommodations, or take alternate assessments, as determined by the IEP team. Students with the most significant cognitive disabilities can be administered alternate assessments based on alternate achievement standards, provided that these standards are clearly being measured by the assessment items. States can use more than one alternate assessment, but each must be reliable, valid, technically adequate, and aligned with content standards. They must also be designed so that scores for math and reading are reported separately and can be included in AYP calculations.

On March 20, 2003, the U.S. Department of Education stated that students' proficient scores on assessments based on alternate achievement standards will be counted in AYP calculations up to a 1% cap. This provision was established to benefit those students with the most significant disabilities. Each state defines who "students with the most significant cognitive disabilities" are; however, they must have one of the 13 categorical disabilities and be deemed unable to achieve grade-level standards by their IEP team. The goal of alternate achievement standards is to promote access to the general education curriculum, with the same knowledge and skills as the basis for both standards. Extending the content standards to make them more accessible for all students is one possible technique for developing alternate achievement standards. States are allowed to have more than one set of alternate achievement standards. The students being assessed on the alternate achievement standards are included in the 1% cap. While the 1% cap does not affect the number of students that may take alternate assessments, it does limit to 1% those scores that can be calculated in AYP as "proficient" or "advanced." Alternate assessments based on grade-level achievement standards are not affected by the 1% cap. This cap is applied at the state and district levels, but states can apply for a slightly higher cap if they can prove that it is warranted. Out-of-level testing is discouraged and included in the 1% cap, since it is considered to be based on alternate achievement standards. Given the concern regarding possible consequences of alternate achievement standards and assessments, these consequences must be identified and explained to children's parents. Furthermore, states need good guidelines on alternate achievement standards and alternate assessments. These guidelines should be structured from the examples of other states that have effectively established alternate achievement standards.

This summer, the Office of Elementary and Secondary Education will begin a Peer Review Process. The Peer Review panel will closely examine the process states underwent to set alternate and general standards and assessments. The Peer Review process will hold states accountable for providing evidence that they have met the criterion.

Whole group discussion

During the question/ answer section following the presentation, the symposium participants elicited further information on the peer review process. Every state will eventually go through the peer review process set to begin this summer. The Peer Review panel will look at how states set and established the regular and alternate

standards and assessments. States will need to provide evidence of these standards and assessments and their technical adequacy. The peer review process is not intended to reprimand states but to help them in their process. The results of the peer review process will likely vary from state to state, since one-size-fits-all concepts cannot be applied to this situation. Given the fluid nature of standards and assessments, changes will likely occur due to variables such as shifts in the governing body or experts' predictions. Unfortunately, the peer review process must freeze a state in time, which is hard to do but necessary for the peer review to be completed.

Other participants sought explanations of terms and processes discussed. For example, out-of-level assessments were distinguished from certain alternate assessments that test grade-level content in a different format. Out-of-level testing is not measuring a student's performance against grade-level standards and is thus based on alternative achievement standards. The process of having IEP teams set alternate achievement standards for a particular child was brought into question; Dr Wolfferesponded by stating that Massachusetts has multiple levels of alternate standards, which is allowed by the regulations.

Increasing Accessible Testing

Martha Thurlow – Universal Design

Universal Design is a way to make testing more accessible to a larger group of students. The goal is to optimize student assessment conditions for today's diverse population of students. Test items that seem obvious and clear to test makers are often not clear to students taking the tests. This creates barriers for students with disabilities.

The 7 principles of Universal Design are:

1. Consider all types of students
2. Include all students (including students with disabilities and students with limited English proficiency) in early field testing of assessments
3. Measure what the test intends to measure, which should reflect content standards
4. Minimize skills required beyond those skills being measured
5. Provide a clear format for text
6. Make text concise and readable: use commonly used words, use vocabulary that is on grade level, minimize use of words
7. Make assessments amenable to accommodations

There is a great deal of research being conducted on universal design, including the Johnstone research. Johnstone conducted a study with low-performing 6th grade students in a rural area. He transformed released state-wide assessments into universally designed assessments with built-in accommodations. Each student participating in the study took one regular released state assessment and one new, universally designed assessment. Johnstone found statistically significant differences between the performances on the two types of assessments in favor of the universally designed assessment.

The design of the test can have a significant impact on individual student performance. This is especially evident when the student knows the information but does not answer the test items correctly because of the design of the assessment. Universally designed assessments do not provide extra help for students who do not know the information; they enable a larger range of students to display their knowledge of the material being assessed.

Based on a participant comment, it was acknowledged that using a universal design may take away from the authenticity of the items. This is a concern that will need to be investigated. There may be a tradeoff between authentic items and universally designed items.

Jacqueline Kearns, ILSSA at University of Kentucky – Digitally Accessible Learning and Assessment

Jacqueline Kearns presented research conducted by Preston Lewis and Scott Trimble on technology/ web/ internet-based assessments. The principle concept of the research was that technology-based assessment would expand the accessibility to a wider range of students.

Currently, the concentration in Kentucky is for a universal design of learning which includes access to general education content, materials, and curriculum. The research was conducted in Kentucky, where currently the technology assessments are used for students with disabilities. Kentucky also has an E-text school program and currently has 3 textbooks in an electronic format. Technology is neither sufficient nor necessary for universal design of learning, but is an enabling factor. A well-developed lesson that has many entry and exit points to accommodate a wide range of students can make content just as accessible to students with disabilities.

The research study dealt with a web-based, individual CATS assessment. The piloted study conducted last year included 16 districts, 31 schools, and 204 students. Ultimately the goal is for all students to take the online assessments, which are currently only administered to students with disabilities. The students participating in the pilot web-based assessment study were required to have an IEP, 504, or LEP plan, access to digital technology in class, previous experience on the practice site, and the accommodation of needing a reader during an assessment. The assessment items were put online but were not universally designed.

On the testing date, the students took the CATS through a website on the computer. The website was designed to make the assessment items available only during specified windows of time. The students would log into the system with a password that would only work for that administration of the test, so the test could not be accessed again The students were able to have the assessment material read to them by the computer as many times as they wanted to meet their reader assessment accommodation. There were

provisions in place to prevent a student from going back to previous pages. The scoring process was completely electronic. The answers were submitted and scored.

The results of the study showed the students to have very positive attitudes towards web-based assessments. They did not mind taking the test as much as they had with the pencil/paper assessments. It has been found that students who take pencil/paper assessment with human readers rarely ask for a test item to be reread. In the study, the students working on the web-based assessments had test items reread by the computer up to three times.

More work needs to be done on the skills students need to be able to use a web-based assessment. When there have been comparisons between pencil/paper assessment and online assessments, there is little difference shown in scores. There is no data to conclude that online assessments are any easier than pencil/paper assessments.

David Rose, Co-Executive Director of CAST – The Future is in the Margins

It is the people outside of the mainstream of culture that make changes happen. This is the population of people that the current methods aren't working for. Children with disabilities are outside the mainstream of culture and the issues surrounding their education will shape the future of education. In order to begin this process we need to change our view of curriculum.

Curriculum needs to be universally designed so all students can access it. There are many technologies that are marketed for regular education but are universally designed for the use of all students, even students with disabilities.

Wiggle Works is universally designed software that was created from research at CAST. It allows printed books to be electronic. The program focuses primarily on K-3 early literacy programs. It is a regular education program that provides tremendous supports for students with disabilities. *Read 180* is research-based middle school remedial reading software. *Yearly Progress Pro* is technology derived from Fuchs work. *Thinking Reader* is a new technology designed to help students in middle school read books in the general education curriculum.

Thinking Reader is an example of how the curriculum should one day look. It is general education technology that gets to the core standards of the general education curriculum, but is designed in a way that makes the curriculum accessible for students with disabilities. The program has digital versions of a number of commonly used middle school books. The digital version has a number of entry levels and provides as much or as little support as the student needs. The reading accommodations that students can use are changing the font of the text; having the text read to the student, highlighting sections of the text; accessing glossary of words used in the text; accessing images of some words in the text; and having words translated and read to student in other languages.

There is also a reading comprehension part to the *Thinking Reader* program. It is modeled after reciprocal teaching principles. Within the reading of the text there are places where the program prompts students to think about the story. When a student is using a "think about the story" section, the program will provide scaffolding. The scaffolding can be as strong as asking a student to choose from a multi-choice bank a good summary of the text to having the student independently type or speak a summary. The scaffolding may also provide models and mentors that will coach them through using a reading comprehension strategy. The students will receive immediate feedback when they respond to a task or question. A work log keeps track of the dates, times, questions asked, sections read, and quizzes completed for the student.

Research results have shown that students with disabilities and students that are struggling with reading understand the text better when they are able to use these types of technologies. The students' state wide assessment reading comprehension scores showed these students to be better strategic readers. The research also found that teachers began to teach better when this type of technology was used in classrooms. The programs seemed to be professional development mentors to the classroom teachers.

These programs are examples of universally designed general education curriculum that work well for students with disabilities. A new way of thinking about curriculum must take place. People should no longer be asking what disabilities students have that prevents them from participating in the general education curriculum, but should be asking what disabilities the curricula have. *Thinking Reader* is an example of a curriculum that has fewer disabilities than many other curricula used.

There is a push for publishers to get on board with making universally designed curricula. There is a movement to make it natural for a publisher to publish not only a text version of a book but also a digital version of that same book. The National Instructional Materials Accessibility Standard, which requires states to work with publishers to obtain digital versions for all books in the schools, is a part of the working draft of the reauthorization of IDEA. If digital texts become part of a student's curriculum, then there will also be a need for digital tests, since the assessment should match the curriculum.

State and district panel on implementing alternate achievement standards

A panel of representatives from Kansas, Louisiana, Maryland, Colorado, Kentucky, and Massachusetts state departments of education answered questions relating to their state's work on alternative achievement standards. Diane Browder facilitated the panel discussion.

1. What prompted the use of alternate achievement standards in your state?
Many states discussed federal or state legislation that prompted states to use alternate achievement standards. Alexa Pochowski from Kansas State Department of Education said the IDEA 1997 requirement for an alternate assessment to be in place caused Kansas to refocus its standards to make sure they applied to students with disabilities. Sharon

Hall from the Maryland State Department of Education credited NCLB for bringing about the Maryland's alternate achievement standards. Terri Rogers Connolly from the Colorado Department of Education pointed to Colorado legislation from 1993 that adapted content standards to include extended benchmarks in order to enable every student access to the curriculum. Jacqueline Farmer Kearns from the University of Kentucky said Kentucky required that portfolio performance-based assessments be in place by 1992 so that students with the most significant disabilities could participate in the new educational reform policies. Daniel Wiener from the Massachusetts Department of Education explained that when the standards were originally written in Massachusetts, they encompassed the "big ideas" that all students should be working on but set the bar where they felt all students should be. When many students missed reaching that bar, they realized they needed to develop some way to help the students with disabilities that were falling short. After IDEA 1997, Massachusetts decided that all special education teachers needed to know what was being taught in general education classrooms and how their students could fit into the general education classrooms and curriculum. They also created a framework guide that allows for different entry points into the general curriculum. Leslie Lightbourne from the Louisiana Department of Education explained that Louisiana took the existing standards and modified them into what they call target indicators. Legislators along with regular and special education leaders attended a meeting to look over and revise the target indicators to make sure they were aligned with the general education curriculum.

2. *Give an example of a state standard and how to show performance on alternate assessments.*

In North Carolina, the writing portion of an alternate assessment would ask the students to place pictures in a sequence to tell a story. The student response would then be assessed for mastery of certain criteria. In Colorado high schools, they use the notion of applied information and work on community skills. In Kentucky, students with disabilities are included in the teaching of linear equations by being instructed on simple math patterns or number identification, which are lower skills of the same standard.

3. Who helped in development when determining the format for the alternate standards?

In Massachusetts, a panel of special educators, regular educators, administrators, content specialists, community members, and parents collaborated to adapt content standards to the needs of students with disabilities, as outlined by the special educators. Other stakeholders were asked for input on the alternate standards. In Maryland, content experts designed a set of content standards that cover Pre-K through grade 8, with the indicators changing at each grade-level. Since the general scope and sequence of skills forms multiple natural entry points, the alternate standards are embedded in the general content standards. Colorado has used teacher input to make changes in the number of standards. University faculty, special educators, and parents helped Louisiana develop their content standards and grade-level expectations.

4. What impact did teacher, parent, and student responses have on alternate standards?

In Louisiana, families and teachers had mixed feelings about alternate standards. The teachers liked the idea of collaborating with general educators, but agreed that it was a lot

of work. In Massachusetts, the alternate standards encountered a great deal of resistance. The teachers wanted to know why they needed alternate standards and were met with the response that it is the law. The first step of the process was to look at the IEP process and develop a continuum of standards with many different entry points. There was little opposition after the continuum was in place. In Kentucky, the move to alternate standards was originally challenged, as were many other special education movements, but people like it now.

5. *Have you had to make changes with alternate achievement standards to comply with NCLB?*

In Colorado, there are performance-based assessments in all grades 3-8 which likely do not need to be changed because the standards and assessments all focus on reading and math. The reliability, validity, and technical quality of the assessment must still be evaluated. In Maryland, there was a need to shift from a functional skills/ community involvement curriculum to a more reading- and math-oriented curriculum and to assess students on this curriculum in every grade. This is beneficial to the students with disabilities because it pulls special educators into the scope and sequence of the regular curriculum. Kansas did not have to make many changes to alternate standards. It has grading rubrics with clear requirements for performance, but is working to help teachers and parents understand what needs to be taught and the meaning of "proficient". Kentucky has a state advisory panel that looks at trends in scores and determines how to make changes without effects to the trend line. Massachusetts' alternate standards are sound but the technical adequacy of its alternate assessments is yet to be evaluated.

6. *Are parents involved ?*

In Massachusetts, the parents are very involved in the IEP process and the IEP guides the alternate assessments. Also, parents have to sign their students' final portfolios before they are sent to the state. Maryland is looking for more parental involvement in the process so they are adding it to the grading rubric.

7. *What advice would you give to someone developing alternate standards?*

Jacqueline (Kentucky) suggested establishing guiding principles or objectives you hope to see completed or achieved. She also recommended broad thinking about access to the general curriculum. Patricia (Colorado) suggested that special and general educators establish and continue collaboration in order to foster changes in instruction, and added that professional development is a very important aspect of developing alternate standards. Sharon (Maryland) advised against making any assumptions and promoted the use of explicit terminology and instructional techniques. Alexa (Kansas) recommended engagement of the whole community in planning and development. Leslie (Louisiana) stated that alternate assessments need to be seen as part of the whole assessment program, calling for collaboration so that a wide range of students with disabilities are represented. She added that the process is a constant cycle of review and feedback.

Other Questions:

Why do we look at generalization on alternate assessments and not regular education assessments?
The panel response was that the students with disabilities may be asked to provide different evidence of learning, but any type of assessment is really a form of generalization.

How do you standardize an assessment for a student reading below grade level?
Sharon (Maryland) suggested determining where the student is functioning in the general education content standards and then setting a goal for progress. Alexa (Kansas) suggested "middle" assessments that are scored against grade level standards but are less complex. Patricia (Colorado) suggested analyzing how students reach the standard, identifying the supports the student needs, and then dealing with the assessment.

Who decides that what is measured in the alternate standards is outlined by the content standards?
Daniel (Massachusetts) advised getting the same people to write both the alternate and the content standard frameworks. Alternately, those who wrote the general education standards can help to guide those writing the alternate standards.

The panel discussion ended with lingering questions and ideas surrounding alignment of alternate achievement standards and content standards. The issue of whether we should base alternate assessment off regular assessments given that those regular assessments may not really be aligned with content standards was raised but not addressed.

Breakout session: Addressing Questions about Alternate Achievement Standards and Assessments

Facilitated by Stephanie Smith Lee and Louis Danielson, Director of Division of Research to Practice at OSEP

***Group 1: How do we align alternate achievement standards with content standards and ensure access to the general education curriculum*?**

Group 1 discussed several different possible approaches for aligning alternate achievement standards with content standards, concluding that no one model or approach discussed is perfect and that each has both benefits and drawbacks. The group also pointed out the need for clear definitions of terms and a common understanding when discussing alternate achievement standards.

The group identified a number of terms and concepts in need of clarification. A common understanding of the purposes of alternate achievement standards needs to be reached to make discussion of alternate achievement standards more effective. The term "alternate achievement standards" can be discussed in relation to accountability, assessment, or curriculum and instruction. When a person is talking about alternate achievement

standards it is important for their audience to understand in what context they are talking. The distinction between content standards and performance standards also needs to be made when discussing the issue. The group found it very difficult to define the term "proficient" when discussing alternate achievement standards. Should it be the number of skills mastered? What would be a reasonable number of skills mastered for a student to be considered "proficient?" Another issue arose when including the term "mastered" in the definition of the term "proficient." What is the definition of "mastered" and does mastering a skill mean that the student should be able to generalize that skill? These terms and the context in which a person is discussing alternate achievement standards must be defined and explained for proper understanding.

One idea proposed to align alternate achievement standards with content standards is using a bottom-up approach. Currently we are using a top-down approach, moving from state assessments to task analyses of state content standards to benchmarks. If we moved in the other direction, we could take all the goals stated in IEPs and aggregate them together to form standards. Then the aggregated data can be monitored for student progress. In the bottom-up approach we can take what a student is working on in class and create an assessment for it. Then we would have the alternate assessment. This illustrates the idea of back-mapping, in which the assessment is derived from what the special educator is teaching. One concern is that special education teachers may not be following a specific curriculum.

The North Carolina example of setting alternate achievement standards is similar to the bottom-up approach because it places emphasis on the individual student's needs. The teachers select skills from the content standards for their children to work on. The data is collected over time to show progress on the skill throughout the year until it is mastered. This approach, however, leads to difficulties in deciding which skills are actually reading- and math- related and determining standards for "proficiency" on a skill. It was suggested that curriculum content people could act as gate keepers, ruling if the skills being taught are really reading or math. Content area specialists could review portfolio items to see if they are aligned with the general education curriculum. A criteria list or a rubric could be posted on the web to check for alignment. This would be an easy way for teachers to know if what they are teaching counts as reading or math content.

The North Carolina example of teachers choosing from a group of skills aligned to the content standard gave away to another "strict but flexible" approach. Because some flexibility in setting alternate achievement standards for individual students along with a strict outline and process to that flexibility is required, one could analyze the content standards benchmarks for a sub-set cluster of core benchmarks that are essential. The student's IEP team would then decide on a number of benchmarks to work on and those specific benchmarks would be monitored. This design includes curriculum, assessment, and instructional aspects of alternate achievement standards but does not address the issue of accountability. This raises the question of how many benchmarks need to be mastered in order to be considered "proficient" or "basic." The benchmarks could be replaced by grade level indicators that create support and access. This approach based off the grade level indicators would achieve grade level alignment, as well.

Alternate achievement standards can also be looked at through an extended or expanded curriculum. The expanded curriculum would provide different entry points for students to access the general curriculum based on their individual abilities. These expanded content standards also would take the guess work out of the teacher's job. The teachers, in other words, would not have to be masters of alignment. The technical adequacy of this type of approach would have to be considered. The "Massachusetts model" is an example of using grade level standards and extending them to include many different entry points into the curriculum. The group struggled with the possibility of students only learning "splinter skills." If a student were to be expected to learn the first three skills under a standard for that grade level and didn't master them before they moved on to the new grade level and new standards/indicators, would they really be learning anything? The issue of limited instructional time brings to the surface questions of which skills are most important to teach.

Another approach is to maintain the same standard but expect something different from students held to alternate achievement standards. This may include partial participation or using a sub-set strain of benchmarks. It may also include different ways of responding or different tasks, so the student would not be assessed with a paper/pencil assessment.

Louisiana uses two sets of standards consisting of either state-specified skills or teacher-specified skills. Currently, the state-specified skills are the ones being assessed for the accountability system. The teacher-specified skills may be the type to look at when making alternate achievement standards.

When aligning alternate achievement standards to content standards, it is necessary to pinpoint the essence of what students need to know. Stakeholders at the state level should identify benchmarks, extensions, support skills, and/or building block skills to represent the essential, core skills of each content standard. The curriculum should be built around the essential skills and then the assessments should evaluate a students' ability on these skills. It is important to keep in mind that alternate achievement standards must be related to content standards, not regular assessments. When developing the standards and benchmarks/indicators, there must be a way for the students to move from alternate achievement standards to general content standards. The standards should be reviewed by outsiders to elicit honest feedback. All these concerns in each approach need to be evaluated and discussed.

Group 2: How do we establish rigorous alternate achievement standards?

Group 2 also found the need to define certain terms before discussing how to establish rigorous alternate achievement standards. The term "alternate achievement standard" was defined by the regulations as an expectation of performance that differs in complexity from a grade-level achievement standard. The group suggested that the word "standard" be replaced with the work "benchmark." They also discussed the difference between content standards and performance standards, with content standards being what is included in the test with a certain percentage of items equaling proficient, while

performance standards represent a range of demonstration or proficiency levels based on grade-level standards. The standards should be the skills and knowledge we want the students to know. The term rigorous was defined as reflecting the highest achievement standards possible based on what we want the students to learn. But who is to evaluate how rigorous the alternate achievement standards should be? If most of the students being assessed against the alternate achievement standards do really well, does this mean that the standards are not rigorous enough? The group decided the rigor of the alternate achievement standards could be defined externally and/or internally. An example of internally determining the rigor of a portfolio system would be to have a system of checks and balances to evaluate it. The portfolio would be signed off on by a team of professionals including administrators. After a student has mastered and been evaluated on one goal, the student could not be assessed again on that same skills in the following year. The externally defined rigor could be decided by means of outside auditors checking data for oddities. The field must be concerned with other levels of rigor as well. There needs to be rigor in instruction, standard setting, and scoring.

A challenge of alternate achievement standards is to match the content standards and alternate achievement levels. Kai Graves reviewed again how things are done in Massachusetts. The group reviewed the break down of the standards in Massachusetts and felt that this is a good approach to developing rigorous alternative achievement standards. Since it is difficult to divorce achievement standards and assessments, the alternate achievement standards must match the alternate assessments.

The concept of "building a ramp" to ensure that students assessed on alternate achievement standards could progress on to content standards was evaluated. The group agreed that this was promising for students with disabilities, but noted that stakeholders need to acknowledge the fact that the students bridging over from one standard system to another would receive the lowest scores while in the transition process.

We must continue to be concerned about what impact alternate achievement standards have on individual students, especially given NCLB's focus on collective groups. We need to evaluate why we are setting up this accountability system. If the alternate achievement standards are made too difficult, it would lead to tracking individual students. The tracking of individual students could lead to lower expectations and less of a possibility for students to progress up to the regular content standards and general assessments. Assessments designed take into account all students and to display the skills of the highest to lowest achieving students may answer this concern.

Group 3: How do we ensure that alternate assessments are technically adequate?

Concerns surrounding reliability and validity make it difficult for alternate assessments to be technically adequate. There are small numbers of students taking these assessments and much variability between states, methods of assessment, and students' individual needs and abilities. It is hard to outline a set of guidelines for technical adequacy that will apply in every situation. One very important principle that should be remembered when discussing technical adequacy of alternate assessments is if a test is not reliable

then it's not valid; but if a test is not valid and based on standards, then it doesn't matter if it is reliable. Scorer/inter-rater reliability as well as process, content, construct, predictive, and empirical validity were all discussed. Concerns arose over an acceptable level of scorer or inter-rater reliability on portfolio assessments. The small size of the populations taking the alternate assessment makes empirical validity difficult. Content validity involves whether the test is constructed to evaluate the students' understanding of meaningful, academic life skills. Predictive validity has to do with whether the results of a test can predict the students' related performance and is therefore an issue of generalizability. The validity of the alternate assessment is influenced by classroom instruction and alternate achievement standards.

The reliability of alternate assessment is a concern when discussing technical adequacy. Multiple-choice tests can increase reliability but diminish the authenticity of the test items. However, there are a number of ways to make authentic alternate assessments more reliable, such as high degree of professional development, subjective and diverse evidence using a general rubric, and being scored by people and not computers. There is little research done on inter-rater reliability of portfolios. The standard for portfolios is two raters and a 3rd "resolver" rater. This lack of research highlights an urgency to get empirical advancements and a literature base for technical adequacy.

Since most states use experts in content and assessment to write assessment items, the validity of alternate assessments is a concern. The validity of an assessment deals with whether it is measuring what it should be. This illustrates the difficulty in accomplishing both alignment and technical adequacy. Should students be assessed on the same content but have different performance cuts, or should they be asked for different outcomes? Students must learn skills that are specific to them as individuals. Once a special educator decides on these skills, a regular educator should be consulted to ensure criterion validity by evaluating if a specified skill is really math or reading.

The goal of alternate assessment should be to improve instruction that is valuable to the individual student. New curricula may need to be developed with content embedded in a functional skills curriculum, so that students can learn academics as well as functional skills that will be beneficial to them later in life. There needs to be communication with parents about how the content skills and knowledge of the alternate assessments will help their children towards an independent, post-secondary life. No one assessment is the silver bullet. The large scale assessments could not be used for instructional or predictive purposes. Curriculum based assessments and portfolios were suggestions of types of alternate assessments. Assessments should be created to work with accommodations and accommodations used on assessments should be the same that are used during instruction.

Language descriptions and purposes of alternate assessments need to be consistent. The conversation about alternate assessments must include the following questions. What inferences are we making from alternate assessments? What does "proficiency" mean? Are the students accountable? Are the schools accountable? Are the teachers accountable?

Assessments currently exist that can be used as alternate assessments but the teachers need supports in place to help use them. Once data is collected from the alternate assessments, the states need to understand thoroughly how to read and interpret it. Alternate assessments should provide useful information on the child's progress, increase instruction, and provide information for parents. A parallel set of assessments aligned with general education content standards will help to guide teaching practices. States need a plan to show how assessments and standards are interrelated. States could have multiple levels of alternate achievement standards. Ultimately, assessments need to serve a purpose for the states, and we need to give states guidelines for student growth.

Breakout Session Report Out

The breakout session ended with a summary of each group's discussions.

Group 1 explained the need for identification and differentiation between terms such as "alignment," "proficient," "advanced," "alternate achievement standards," "performance standards," "content standards," and for recognition that discussions about alternate achievement standards can come from various contexts. The differences between top-down and bottom-up models of aligning alternate achievement standards with content standards were discussed. Five possible approaches were explained as ways to align the two sets of standards. The primary concepts that emerged regarding alignment were extending content standards to lower the complexity of the skills and reducing the standards to the core essence of knowledge while building supports and structures for students with disabilities. Various successful systems of alignment should be identified so that they can serve as models for others. One system will not work for all states and schools.

Group 2 also offered working definitions of the terms "alternate achievement standard" and "rigorous." Rigor should be defined by internal or external bodies. Rigor can be defined by internal bodies such as parents and administrators signing off on a student portfolio. External bodies can audit for oddities in the data. There needs to be rigor in content, scoring procedures, and standards- setting. The group also discussed one concern surrounding building a ramp for students with disabilities to go from alternate achievement standards to content standards.

Group 3 explained how technical adequacy of alternate achievement standards/ alternate assessments is very difficult to determine. There are concerns with the reliability of scoring/rating for portfolio assessments, and with the validity of the content students are working on based alternate achievement standards. Are these alternate achievement standards aligned to the content standards? Are the alternate achievement standards necessary skills that these children need? The validity in terms of alignment to content standards and necessity of the skills to the students' repertoires is in question.

Friday, February 6, 2004

EPRRI White Paper: "Developing Accountability Systems that are Meaningful and Valid for Students with a Range of Abilities"

Margaret McLaughlin presented an EPRRI paper, currently in draft form, on alternate achievement standards and assessments, titled "Developing Accountability Systems that are Meaningful and Valid for Students with a Range of Abilities." Information included in the paper was taken from the most recent literature on general education standards and assessments. It was then applied to the development and implementation of alternate achievement standards and assessments. In reviewing the literature EPRRI staff found that although there is convergence on the literature, not all researchers agree across all aspects. Many researchers demonstrated a great deal of consensus of *what* accountability systems are, especially with regard to closing the achievement gap. However, the literature reviewed for this paper did not demonstrate a "shared understanding" on the definitions of commonly used terminology. Five assumptions were taken from Fuhrman and Elmore (2003) and were used as a framework for the paper:

1. Student achievement is the key goal of school and accountability focuses the attention on this goal.
2. Student performance can be accurately and authentically assessed
3. Consequences, inducements, and sanctions motivate school personnel and students.
4. Improved instruction and higher levels of student achievement will result from the increased accountability.
5. Unintended negative consequences will be minimized.

Dr. McLaughlin clarified that while not all participants may agree on the assumptions, they are the ones that exist in the literature on general education standards and assessments. EPRRI staff then derived topics and questions from the existing literature and categorized them under the five assumptions, in the form of questions aimed at alternate achievement standards and assessments.

After a review of the information contained in the draft version of the paper, Dr. McLaughlin opened the discussion to symposium participants for their input, questions, and critiques. Both state and federal representatives contributed to the conversation, and highlights are as follows:

- A representative from the US Department of Education noted that "most schools' mission statements seek to have outcomes for students that surpass merely achievement" and also encompass citizenship, life-long learning, etc. He further explained that "the standards-based reform movement is leading more regular education teachers to think of achievement as only academic areas," while special education teachers often include social and behavioral goals under the achievement heading.
- Other participants strongly advocated academic instruction for *all* children. Without instruction in academic areas, how is a child ever going to have anything to use in social situations with their peers?

- Many participants were in agreement that allowing reading and math goals to have a broader scope introduces the opportunity for more functional skills to be recycled and called reading and math all over again.
- A state-level representative expressed concern over developing entirely different standards for students with the most significant cognitive disabilities. By developing separate goals for them, are we lessening the value of these children?
- Some participants expressed concern for the preparation of new and veteran special educators with regard to content-area training.
- When implementing alternate achievement standards, curriculum and placement shouldn't be confused. Accessing the general education curriculum doesn't necessarily mean placement in a general education classroom.
- A representative from the Office of Special Education Programs recommended that instead of developing alternate achievement standards, it would benefit students more if we develop a direct link to the grade level content standards. Rather than aligning one set of standards to the other, we should extend grade level standards to encompass all learners. "If we can figure out the essence of fourth grade, it will also help the children who were previously struggling to make proficiency at grade level."

The discussion concluded with an announcement from David Riley that an online forum posted at the EPRRI website (www.eprri.org) would give participants further opportunity to discuss the paper. EPRRI staff will utilize these comments, as well as the discussion from the symposium, when adding to and revising the paper itself.

Small Group Discussions

The protocol for small group discussions was as follows:

Participants were assigned to one of four smaller groups for two discussions. During each breakout session, participants were given a protocol to follow and a timekeeper and facilitator were chosen. The first session called for recommendations on what technical assistance and development is most needed by districts and states. The second session focused on eliciting research and development needs. Both sessions followed the same procedures, including the following steps:

1. Participants spent the first five minutes individually brainstorming and writing recommendations.
2. Each group member then shared his or her recommendations, after which the group discussed similarities and/or differences within the group.
3. The group refined and merged these recommendations into one or two group recommendations that were summarized on chart paper.
4. Each small group then shared their recommendations with the larger group, with discussion following as needed.

Discussion on Technical Assistance and Development Needs

Small breakout groups were charged with discussing what technical assistance and development states and schools districts need over the next one or two years. Groups were also asked to discuss what implications their recommendations have for OSEP. The summaries from each group which follow reflect the top two to three Technical Assistance needs identified by each group.

Group 1 Summary

States and districts need to better understand the meaning of terms. Although no one felt than OSEP should or could standardize how states are required to implement alternate achievement standards and assessments, the group did discuss the need to operationally define terms. If there is more consistency in the terms that are used and the ways in which accountability systems are designed, states will be able to clearly understand the intents of alternate achievement standards, and how these fit into an overall accountability system.

States and districts need technical assistance in making data driven decisions. The group felt that professional development may be needed to demonstrate best practices in collecting and using the types of information which will be generated from alternate assessments. Given the broad range of abilities for the target population that will be held to alternate achievement standards, we need to ensure that appropriate data are being collected at all levels.

States and districts need technical assistance in content-area training. While most special educators, both pre-professional and veteran, have expertise in accommodating and designing methods of instruction for learners with exceptional needs, few have exposure and training in specific academic areas such as math and science. Special education teachers need assistance to interface special education skills with content area subjects.

These recommendations lead to several implications for OSEP. Information needs to reach all of the stakeholder groups. Professional development is needed for teachers on teaching subject matter content and on writing IEP goals to reflect a convergence of content and functional goals. Teachers need instruction in utilizing and making data-driven decisions, especially given the wide range of abilities of students who fall into the small population for whom alternate standards are appropriate.

Group 2 Summary

Group 2 believed states and districts have a need for core principles or constructs to facilitate common understanding and a common language. The definitions of “content” and “performance/achievement” standards need to be consistent. The differences and similarities between “functional” and “access” skills and a description of the terms’

relation to standards and assessments need to be clearly stated. Assessments should shift the focus from functional skills to access skills, non-academic skills that help students with significant cognitive disabilities access the curriculum. These must be directly taught to students with disabilities and high-risk students. The alignment of alternate achievement standards and content standards needs to be completely understood by states and districts in order for them to carry out the process.

States and districts also need technical assistance in the area of effective practices, strategies, and models for developing alternate achievement standards and alternate assessment. Assistance should offer best practices, including the creation of a broad, representative stakeholder group (e.g., department of assessment, department of school accountability, department of curriculum and instruction, department of equity). Assistance in how to improve and optimize psychometrics is also needed. Many representatives expressed concern over interpreting both aggregated and disaggregated alternate assessment scores, given the broad ability range and limited number of students represented in the alternate assessment.

States and districts need OSEP to develop guiding principles that will help them to establish appropriate goals, strategies, and accountability systems. OSEP should also provide assistance and guidance to states in ensuring reliability and validity for alternate assessments.

Group 3 Summary

States and districts need models, approaches, and exemplars in many areas. Content and special education alignment models are needed. One of the most important needs is clarity of the term "alignment." The group members felt states have varied interpretations when determining whether alternate achievement standards are considered in "alignment" with the regular grade-level content standards and curriculum. A model is needed of what "extended standards" look like for grades 3-8 and one high school grade level. The states and districts need models of partial participation. When designing alternate achievement standards, it is very important to keep in mind the social validity of that standard.

States need to understand how technical adequacy applies to alternate achievement standards and alternate assessments. "Rigor in standards proves/avoids rigor-mortis for technical adequacy." Do broader constructs of technical adequacy apply to alternate assessments based on alternate achievement standards? There is a need to draw in reformers from general education (e.g., Bobhinn). There is also a need to apply technical adequacy standards to existing alternate assessments.

OSEP needs to assist in developing examples for what extended standards should look like and how accountability and school improvement can be implemented in the in grade-level classrooms. There is also a need for OSEP to examine decision-making regarding individual students and how they are "assigned" to alternate achievement standards.

Group 4 Summary

States and districts need technical assistance in the area of models and exemplars. A fluid and seamless accountability system needs to have:

1) alternate ways to represent regular content standards
2) alternate ways for students, including those with the most complex disabilities to demonstrate proficiency on regular and alternate achievement standards
3) alternate ways to engage students in assessments.

Additionally, representatives felt that states need assistance in articulating the clear and defensible purposes/goals of their alternate achievement standards. The goals should be measurable (e.g., students are gaining literacy, etc.). Furthermore, states need clear and shared understanding of terms and constructs (e.g. alternate achievement standards, alignment, the meaning of "proficient," extended standards, etc.)

The states and districts need OSEP to develop a working definition of what "proficiency" means for students who are held to the alternate achievement standards. OSEP needs to develop different constructs of validity. For example, at the end of four years, what are we seeing? How are student performing on measures other than the alternate achievement standards? What inferences can we make from students' performance on the alternate achievement standards? Do we see more inclusion? "We should have more evidence beyond just their performance on the alternate achievement standards."

Group members also suggested that OSEP specify the purpose for alternate assessments. Is the purpose to create lifelong learners? Is it to measure progress on IEP goals? Is it to access and master the general curriculum? What decisions should be made with the resulting data? Will IEP decisions be made based on this data? Will future placement and/or high school graduation or exit decisions be made based on this data?

Discussion on Research and Development

The same small groups discussed and generated a list of priority areas of research and development. The groups were to focus on longer term development and to pose questions. Groups were also asked to consider the investments each research question or development would require, as well as the principal partnerships involved. Below are the research and development priorities or questions chosen by each group.

Group 1 Summary

The first area of Research/Development should deal with comparing/contrasting different types of assessments (large-scale "system" assessment vs. on-going curriculum based assessment, etc.). Following are specific research questions:

1. What are their relative effects on instruction, as measured by another assessment or other evidence? In other words, what is the relationship between type of assessment (alternate assessment, curriculum based) and instruction delivery?

2. What are the psychometric properties of each type of assessment, which of these properties are essential for developing an alternate assessment, and how can we ensure rigor and technical adequacy when developing an alternate assessment?
3. What differences exist between assessing various groups of children (minority populations, ELL, etc.)?
4. What constructs are actually measured by each type of assessment?
5. How does each type of assessment fit within the accountability system overall, and what is the effect of the 1% of students held to alternative achievement standards (and students with disabilities in general) on the system as a whole?

This development work would involve many partners, including both special education assessment experts and general education assessment experts. As one participant noted, "We need to have all of the people in the same room at the same time)." The monetary investment would be large, but this focus requires immediate attention.

The second area of Research/Development should investigate the "nature of alignment." For example, how do general content standards align with alternate achievement standards, and how is access to general curriculum provided? How do you measure alignment? This also would also involve many partners and a large investment, and would be a long-term research and development activity.

Group 2 Summary

Group 2 concluded that the focus of research and development should be on answering the question, "so what?" The first area should address the circumstances and the extent to which inclusion in the accountability system for students with disabilities has an impact on outcomes for students with disabilities. Specifically, to what extent does the focus on "academic skill curricula" impact specific life goals and broader educational goals of students with disabilities? Although research on outcomes is difficult, it is important, especially given predictions that entry level positions will soon require higher level reading skills than those of managers and administrators.

The second area of Research/Development should examine the extent to which general content and alternate achievement standards are actually aligned and represented in instructional experiences and assessments.

Group 3 Summary

The first area of Research/Development should explore different accountability models in search of seamless systems. A possible scope would be 4 centers dealing with multiple states and state/university partnerships. Existing and new models (e.g. CBM model) should be considered. In order to assess academic gains on state standards, longitudinal data should be collected on intervention models to show if and how students are mastering state standards. Different models of standards setting should be tested, as well.

The second area of Research/Development should consider dynamic alternate assessments. How can we come up with a model that informs instruction, and compares or guides different instructional strategies? The scope would be multiple research grants that consider:

- technical quality
- how assessments inform instructional strategies
- how schools respond when students score low on alternate achievement standards
- the impact and next steps in these situations
- standard setting models
- students' longitudinal academic progress on other measures.

Group 4 Summary

The first focus of Research/Development should be to comparatively study different accountability models. Focus should be on both alternate assessments and alternate achievement standards and the overall competence enhancement for students and systems. In other words, what is the impact on students, and does it make a difference which model is used? A cost-benefit analysis should be completed. The implementation integrity should be investigated, such as its relationship with the latitude allowed teachers and the impact of professional development and changing perspectives over time. The stability over time and autonomy for teachers should also be looked at within the comparative study.

The second area of Research/Development should address issues with technical adequacy. Specifically, what are alternative ways of looking at technical adequacy "in this new realm of achievement standards and the assessments"? The reliability, validity, sensitivity, and usefulness of the alternate assessments should be looked at. The role of prompts, cues, supports should be investigated in terms of their potential impact on technical adequacy. The stakeholders' involvement should be analyzed as well.

Additional Comments

Other comments were made during whole group discussions of these Research/ Development proposals. One participant discussed the possibility of using the Bob Stakes evaluation model in which you examine how things were before implementation, state what you intend to do and the intended outcomes, look at what you actually did, study observed outcomes. He noted that there is right now a "natural experiment in place", since different states are implementing different models (portfolios, performance assessments, downward extensions of other assessments). A contractor could do an evaluation of these models relatively easily, with data on relative merits and limitations of different models available in two or three years. This could provide an answer to whether it matters which model is used. Another participant responded that a multi-state consortium (including Colorado) is already in the process of comparing different models of alternate assessments, such as portfolios and performance based assessments. Preliminary findings should be available within the next year or so.

Other questions were raised as well. For example, what happens when the assessment system is more sophisticated than the NCLB requirements? NCLB is not intended to deal with multiple measures; rather, it sets the bottom line, but if you exceed bottom line, there is no incentive. Related to this, to what degree can the assessment scale be stretched when using universally designed learning standards? At some point the scale might break. Possible areas to examine were multi-trait, multi-level (matrix) models, which could potentially be stretched via universal design.