Alternate Assessment Design–Mathematics Technical Report 1: Project Overview

Applying Evidence-Centered Design to Alternate Assessments in Mathematics for Students with Significant Cognitive Disabilities

R. Cameto, G. Haertel, A. DeBarger, and K. Morrison
SRI International

Students with significant cognitive disabilities challenge conventions with respect to the teaching, learning, and assessing of academic content. Assessment has been instrumental in changing the learning expectations of these students which in turn is beginning to influence classroom instructional practices. Assessment designers are challenged to develop assessments that adequately and reliably show what these students know and can do. The sheer variability in this target population, the assumptions about measuring their achievement, and the variability of design implementation procedures make traditional assessment design approaches inapplicable without some reformulation (Gong & Marion, 2006; Government Accountability Office, GAO, 2009; Ryan, Quenemoen, & Thurlow, 2004). The methods used to date in designing alternate assessments and selecting their content are varied but typically do not match the technical rigor used for designing general education assessments (Bechard, 2005). The Alternate Assessment Design–Mathematics (AAD-M) project is the first to address systematically the specification of grade-level academic content for alternate assessments of students with significant cognitive disabilities through the application of evidence-centered design (ECD) and the principles of universal design for learning (UDL).

ECD directly addresses these most pressing issues by using a replicable assessment design process that can be applied to all content areas and all types of evidence, from performance tasks and portfolio activities to technology-based simulations and animations to traditional multiple-choice item formats. The use of ECD can enhance the quality of assessments and improve the efficiency with which future assessments are developed while documenting the myriad design decisions required when developing a valid assessment of student learning (Mislevy, Steinberg, & Almond, 2003). The AAD-M project is innovative in two aspects: It is applying ECD for the first time to assessments for students with significant cognitive disabilities, and it is integrating ECD and UDL approaches in the design of tasks for alternate assessments based on alternate academic achievement standards (AA-AAS). This work extends current knowledge in the field and provides a prototype for future alternate assessment development.

Utah, Idaho, and Florida have formed a consortium with SRI International to improve their AA-AAS using ECD to design and develop assessment tasks that are linked to state extended content standards in mathematics. In this report, we describe

- Project goals and activities
- The development of assessments for accountability purposes for students with significant cognitive disabilities
- ECD and UDL frameworks and describe how they are applied through a co-design process
- Our plan to produce a series of technical reports, including procedural guidelines, design documents, and associated sample assessment tasks

- Our dissemination plan including the project website, www.alternateassessmentdesign.sri.com

History

A succession of federal laws, including the Individuals with Disabilities Education Act (IDEA) of 1997 and 2004 and the 2001 reauthorization of the Elementary and Secondary Education Act (ESEA), require that *all* students be assessed in reading/language arts, mathematics, and science and be included in state accountability systems. Most students with disabilities participate in general assessments even with accommodations, but some students, including those with significant cognitive disabilities, may need alternate ways to access assessments. To include these students in educational accountability systems, all states have developed alternate assessments based on alternate achievement standards (Kohl, McLaughlin, & Nagle, 2006; Thompson & Thurlow, 2003). However, states have faced a number of challenges, including (1) clearly documenting links between their general education content standards and their alternate assessments, (2) developing a clear rationale for their choice of particular content standards in their alternate assessments, and (3) providing strong evidence that the intended assessment content is actually being assessed, as called for by Flowers, Wakeman, Browder, and Karvonen (2007) in *Links for Academic Learning.* Although valuable work is under way in the area of technical adequacy of alternate assessments (for example, by the New Hampshire Enhanced Assessment Initiative and the National Alternate Assessment Center), the reliability and validity of alternate assessments remain problematic and complete confidence cannot be placed in results of such tests (GAO, 2009; Quenemoen, 2008; Quenemoen, Kearns, Quenemoen, Flowers, & Kleinert, 2010). A compelling need exists for well-designed, evidence-based AA-AAS to measure and document the performance of students with significant cognitive disabilities.

Federal education laws enacted during the past decade have produced a frenetic pace of change in alternate assessments and generated a marked shift to the full inclusion of students with significant cognitive disabilities in accountability systems across the states, accompanied by a shift in instructional emphasis from functional skills to academic content (Thompson, Johnstone, Thurlow, & Altman, 2005). A review conducted by Quenemoen (2008) indicated that states use several different approaches when gathering information on the performance and progress of these students. These approaches include rating scales, portfolios, performance tasks, multiple choice, or a blend of multiple formats (Cameto et al., 2009). These, in turn, are implemented with varying degrees of local decisionmaking, Individualized Education Plan team involvement, scoring, and criteria for inclusion in calculations for adequate yearly progress (Cameto et al., 2009). Design and implementation of alternate assessments are in considerable flux (for example, existing assessments are likely to be revised to align with the 2010 Common Core State Standards Initiative[1]).

[1] The Common Core State Standards Initiative (CCSSI) is a state-led effort coordinated by the National Governors Association Center for Best Practices (NGA Center) and the Council of Chief State School Officers (CCSSO). Governors and state commissioners of education from 48 states, 2 territories, and the District of Columbia developed a draft common core of state standards in English/language arts and mathematics for grades K–12. When the draft is formalized, the participating states will adopt the Common Core Standards. The CCSSI plans to develop a common core of standards in science.

The *Standards for Educational and Psychological Testing* (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 1999) continues to be the authoritative source of information on test validity. Several standards are particularly relevant to the design of AA-AAS. Among other dictates, the *Standards* require that procedures for specifying and generating test content be described, that the relation of the items to the dimensions of the domain be stated clearly, and that steps be taken to ensure that test score inferences accurately reflect the intended construct rather than any disabilities.

Nonregulatory guidance explained the December 9, 2003, regulation to ensure that students with the most significant cognitive disabilities were fully included in state accountability systems and that students had access to challenging instruction linked to state content standards. The guidance clarified that states are responsible for designing assessment systems that permit all students in the tested grades to be assessed against grade-level content and achievement standards, ensuring that assessments are based on state content standards. States are expected to field-test assessments by sampling the types of students expected to participate in the final assessment administration, define the assessment's measurement constructs precisely, and develop accessible test forms that allow for a wide range of accommodations in test administration. For AA-AAS in grades 3 through 8 and at the high school level, the assessment materials should show a clear link to the content standards for the grade the student is enrolled in, although the grade-level content may be reduced in complexity or modified to reflect prerequisite skills. The AAD-M holds this guidance as the target for performance task assessment and design and selected evidence-centered design and alternate assessment development methods.

The application of ECD to alternate assessment addresses validity issues as described by Shafer (2005) and Tindal et al. (2003) by applying a replicable process that makes explicit the content to be assessed, the evidence to be collected, and the features of tasks to be developed. Furthermore, this process is generalizable and can be applied to all content areas and types of evidence. The use of this approach in the AAD-M project will contribute much-needed information for improving AA-AAS and will further inform efforts to improve assessment practices generally across the ability spectrum and specifically for students with significant cognitive disabilities.

A review of relevant literature on alternate assessments and the results of the peer review process applied by the U.S. Department of Education to alternate assessments indicate the technical quality of alternate assessments continues to be a significant challenge (GAO, 2009; Quenemoen, 2008; Quenemoen et al., 2010). Meeting the standards adopted by the American Psychological Association and the American Educational Research Association requires access to large item pools, large samples of students to establish item and scale functioning and difficulty characteristics, and the use of standardization at every step in the assessment development process. Alternate assessment systems vary greatly in the design of the system, type of evidence collected, and the standardization that is applied in part because of the nature of the evidence collected in the alternate assessment systems. Portfolios are still the most common, and they frequently lack evidence supporting their reliability and validity (Cameto et al., 2009; Quenemoen, Thompson, & Thurlow, 2003; Thompson et al., 2005). Some states have been moving toward use of performance tasks to assess students with significant cognitive disabilities, which have the advantage of producing scores that can be evaluated through modern item response theory methods and can be administered to groups of students. In addition, formative assessments for this population are beginning to be explored.

Historically, large-scale assessments have not focused on how content, design, or task characteristics influence the ability of students to perform, especially those students in the tails of the achievement distribution. Alternate assessment designers in particular have often lacked systematic design processes that (1) define the focal knowledge, skills, and abilities (KSAs) required to demonstrate proficiency in academic content areas; (2) design assessment tasks with features that are well aligned with the focal KSAs; (3) design assessment tasks that minimize nonfocal KSAs and thereby mitigate construct-irrelevant variance; and (4) take into account the many ways that students perceive test content and express their responses. Those using the rigorous multistep design process that is central to ECD carefully consider how the content, task, and learner characteristics interact in the creation of assessment tasks.

Application of ECD to the Assessment Needs of a Challenging Population

Students with significant cognitive disabilities may come from any of the 13 regulatory categories included in the Individuals with Disabilities Improvement Act. In a survey of special education teachers of students with significant cognitive disabilities in several states, however, Cameto and colleagues (2010) found that, when asked to report on a randomly selected "target" SWSCD, teachers reported these students were primarily clustered into three disability categories: mental retardation, autism, and multiple disabilities. Although these data represent only three states, they are consistent with findings reported by Kearns (2007). An additional finding was that the majority these teachers reported that the target SWSCD had multiple disabling conditions (Cameto et al., 2010). The teachers surveyed by Cameto and colleagues (2010) also provided information on students' communication level and academic ability. Teachers were provided with descriptions of three communication levels developed by Browder, Flowers, and Wakeman (2008[2]) and asked to indicate which one best reflected the highest level at which their target student currently communicated. A majority of teachers (68%) reported that the target SWSCD communicated with symbols or words and had basic or emerging functional academic skills. A small percentage (12%) indicated that the target students had no reliable communicative response.

---

[2] Communication levels were described as follows:

- Level 1—Pre-symbolic. Has not yet acquired the skills to discriminate between pictures or other symbols (and does not use symbols to communicate). May or may not use objects to communicate. May or may not use idiosyncratic gestures, sounds/vocalizations, and movements/touch to communicate with others. A direct and immediate relationship between a routine activity and the student's response may or may not be apparent. The student may have the capacity to sort very different objects, may be trial and error. Mouthing and manipulation of objects reads to knowledge of how objects are used. May combine objects (e.g., place one block on another).
- Level 2—Early symbolic. May use some symbols to communicate (e.g., pictures, logos, objects). Beginning to acquire symbols as part of a communication system. May have limited emerging functional academic skills. Representations probably need to be related to the student's immediate environment and needs.
- Level 3—Symbolic. Communicates with symbols (e.g., pictures) or words (e.g., spoken words, assistive technology, ASL, home signs). May have emerging or basic functional academic skills. Emerging writing or graphic representation for the purpose of conveying meaning through writing, drawing, or computer keying.

In the past, students with significant cognitive disabilities typically lived in residential institutions where they were provided little in the way of education (McDonnell, Hardman, and McDonnell, 2003). Since the passage of Public Law 94-142, also known as Part B of the Education of the Handicapped Act (1975), renamed in 1990 the Individuals with Disabilities Education Act (IDEA), these students have become integrated into home and community life, living with their families, attending their neighborhood schools, learning to read, traveling independently in their communities, and engaging in productive employment as adults. Browder and Spooner (2003) reviewed the evolution of special education from the developmental, mental-age-based perspective of the 1970s through the functional, life-skills view of the 1980s; the social inclusion and self-determination view of the 1990s; and the academic standards-based demands since the turn of the century. Each time expectations have been raised, students have exceeded previous expectations, and now most are members of their communities, have friends, and enjoy social memberships like their nondisabled peers (Wagner, Balladeers, & Marder, 2003).

Although federal requirements hold students with the most significant cognitive disabilities to high academic expectations, the strongest argument for such high expectations for these students is their own performance over the last three decades (Marion & Pellegrino, 2006). Since the advent of IDEA, expectations for students with disabilities have been raised repeatedly, and students have consistently outperformed what had previously been perceived to be their limits. Initial research indicates that including students with disabilities in large-scale accountability testing results in higher expectations, improved instruction, and improved performance for those students (Cortiella, 2007; Kleinert, Kennedy, & Kearns, 1999; Quenemoen, Lehr, Thurlow, & Massanari, 2001; Towles-Reeves, Garrett, Burdette, & Burge, 2006; Ysseldyke, Dennison, & Nelson, 2003). The collaborating states – Utah, Idaho, and Florida – and the SRI team recognize that efforts to design alternate assessments must proceed within this context of the possibilities signaled by previous advances in special education.

High expectations—a hallmark of good education—now include academic performance for these students. But how can their academic performance be assessed? For general education students, most if not all statewide assessments have been developed following careful plans and blueprints linking content standards to assessment items with known psychometric properties, and processes and links have been well documented. Thus, their validity is well understood. For students with many types of disabilities, such assessments may be accommodated or modified. For students with significant cognitive disabilities, accommodations or modifications to the general education assessments are not sufficient. Although assessments for these students must by law be linked to general education content standards, they may use alternate academic achievement standards to measure KSAs.

Many students with significant cognitive disabilities also have coexisting physical or sensory disabilities that can interfere with their assessment performance. In recent years, augmentative and alternative communication devices and assistive technologies have reshaped the way such students are taught and learn, raising even further our expectations about what they may achieve. For students eligible to take AA-AAS, at least three important factors must be attended to: accommodations and technology, including universal design for learning and assessment; alternative and augmentative communication systems; and systematic prompting with feedback that has been used extensively in research with students with severe disabilities

(Browder & Cooper-Duffy, 2003). Each of these considerations will be integrated into the design of assessment tasks based on the ECD process in this project.

The AAD-M project emulates and extends the ECD approach to the design of alternate assessment tasks in mathematics. ECD is a practical theory-based approach to developing quality assessments that combines developments in cognitive psychology and advances in measurement theory and technology. ECD is a well-understood process that can be used in all stages of assessment design and development, from domain analysis to the specification of student, evidence and task models to the creation of items and tasks and finally to the design of an assessment delivery system. Although each of the collaborating states has unique needs, ECD provides a robust and suitable approach that can be customized to each state's needs. The tasks designed in the AAD-M project can be implemented in portfolio or performance task assessment systems or in formative benchmark applications. The implementation of tasks can be guided according to the assessment specifications of each participating state—portfolio systems in Idaho, on-demand summative assessments in Utah, and diagnostic formative assessments in Florida.

A synergistic application of ECD and UDL facilitates the development of assessment tasks aligned with academic content standards, increases the accessibility of these tasks, and raises expectations for the performance of students with significant cognitive disabilities.

Project Description and Goals

The AAD-M project combines current knowledge from multiple disciplines to advance the design of alternate assessment performance tasks for students with significant cognitive disabilities. The approach integrates recent work in (1) the pedagogy of special education for students with significant cognitive disabilities (Browder & Spooner, 2003), (2) alternate assessment design (Bechard, 2005), and (3) universal design for learning (CAST, 2008) with (4) evidence-centered assessment design (Mislevy & Haertel, 2006). This work is guided by federal guidelines for alternate assessment design that specify that "all students, including students with disabilities, be held to grade-level achievement standards when taking assessments" (U.S. Department of Education, 2005).

The goals of the project are to

1. Extend the conceptual framework of evidence-centered design to alternate assessment based on alternate achievement standards using the Principled Assessment Design for Inquiry (PADI) assessment design system
2. Integrate the principles of universal design for learning with ECD to guide the development of tasks that are accessible to all learners
3. Use the National Council of Teachers of Mathematics (NCTM) Principles and Standards for School Mathematics to identify common expectations that represent critical areas of learning common across all three states in number and operations, algebra, geometry, measurement, and data analysis and probability
4. Develop AA-AAS assessment design patterns, task templates, assessment task specifications, and exemplar tasks that address priority state academic standards in mathematics for students with significant cognitive disabilities

5. Evaluate the exemplar assessment tasks produced using ECD through pilot-testing in all three states
6. Enhance the human capital of state departments of education staff in learning how to use ECD to design and develop assessment tasks for use with students with significant cognitive disabilities, including the provision of procedural guidelines
7. Support state department of education staff and teachers in the development of additional performance tasks in mathematics to expand the task bank for each state
8. Provide participating states with a library of design patterns, task templates, and task specifications that are reusable and extendable to the authoring of additional performance tasks in mathematics.

Evidence-Centered Design

Evidence-centered design is a recommended approach for the development of educational assessments and can be applied to a range of content standards and assessment types. The rigorous multilayer design process central to ECD enables designers to consider systematically the content, task, and learner characteristics that influence student performance. ECD provides a foundation for assessments that states can use to address the validity of their assessment systems.

A strength of ECD is the support it provides for the development of items and tasks for all students that focus on construct-relevant content, minimize the impact of construct-irrelevant skills, and take into account appropriate accessibility options. For example, in a mathematics examination, math content would be targeted and the need for non-construct-relevant skills such as reading would be minimized; designers would consider supports such as use of a large font or alternate response options during item design instead of modifying items and tasks after they have been written.

The ECD process involves five layers of activity. The layers focus in turn on the identification of the content to be assessed; the creation of a model of the assessment; the design of assessment elements such as potential observations, work products, rubrics, and psychometric models; the creation of these elements including the assessment tasks; and the design of the assessment delivery, scoring, and reporting. Each layer is described below.

1. Domain analysis involves determining the specific content to be included in the assessment. Use of the common core standards and existing state standards exemplify starting points for domain analysis.
2. Domain modeling entails creation and documentation of a high-level description of the assessment. Design patterns are one example of this type of activity.
3. Conceptual assessment framework specifies in detail the knowledge, skills, and abilities to be assessed, the evidence that needs to be collected, and the features of the tasks that will elicit the evidence. Also identified are nontargeted KSAs, which, although required for successful performance on an item, are not the intended target of the assessment. By identifying nontargeted KSAs, designers can minimize construct-irrelevant variance and maximize accessibility. Finally, the psychometric model and evaluative decision rules for task scoring are considered and assessment task features are detailed and carefully aligned with the targeted and nontargeted KSAs.

4. Implementation is the creation of the assessment items or tasks, along with appropriate accessible alternate representations of item or task content.
5. Delivery involves specification of the processes for the assessment administration, scoring, and reporting, including accessibility features that are allowed without violating the targeted KSAs.

Universal Design for Learning

Universal design emphasizes the importance of addressing accessibility for the broadest range of potential users during the initial stages of designing a product and throughout the development and implementation of the product. The use of universal design principles creates flexible solutions because designers consider from the start the diverse ways in which individuals will interact with a product and the environment.

The tenets of universal design have been extended to the education arena; this extension is referred to as UDL. When sources of construct-irrelevant variance in an assessment are identified by ECD, the application of UDL principles can be used to minimize construct-irrelevant variance by incorporating appropriate options for how students interact within the assessment environment. In this way, ECD works synergistically with UDL. By considering multiple means of perception, expression, cognition, language and symbol use, executive functioning, and engagement, the application of UDL in the ECD process accounts for individual differences in how students recognize, strategize, and engage in learning and testing situations. This synergistic process minimizes the unintended negative influence that access needs may have on student performance and maximizes the opportunities for students to show what they know and can do.

UDL principles are incorporated into this ECD process during assessment design and item authoring by considering multiple means of perception, expression, cognition, language and symbol use, executive functioning, and engagement. This can include consideration of augmentative and alternative communication systems.

Crosswalk of States' Extended Standards in Mathematics and the National Council of Teachers of Mathematics (NCTM) Expectations

In an analysis of test design and development methods, Bechard (2005) reported that the best approach to designing alternate assessments aligns the assessment content with a state's academic content standards, thus both promoting access to the general curriculum and increasing instructional opportunities for students with significant cognitive disabilities. States have developed alternate assessment items, tasks, or types of evidence of student performance in two ways. Most states based the design of their items/tasks for their AA-AAS on extensions of the grade-level content standards referred to as extended standards and adopted by the state boards of education (Cameto et al., 2009). Other states based the design of the items/tasks for their AA-AAS on the grade-level content standards adopted by their boards of education for all students, often referred to as general education grade-level standards. The three AAD-M collaborating states had adopted extended content standards in mathematics for their AA-AAS. The state extended standards were the foundation for the domain analysis reported, the first layer of the ECD process.

This study team analyzed the three states' content standard extensions and aligned them with the mathematics expectations put forth in the *Principles and Standards for School Mathematics* (National Council of Teachers of Mathematics, NCTM, 2005). The NCTM expectations are intended to help in focusing curricula within a grade band and in developing curricula progressively across grade bands. They are used to design instructional programs and curricular frameworks for K–12 mathematics. The expectations represent a consensus among educators about the content and processes that are essential to understanding the knowledge, skills, and abilities required for successful performance in mathematics. NCTM identifies expectations in five domains: numbers and operations, algebra, geometry, measurement, and data analysis and probability. To focus the states' efforts on a common set of expectations, the project team conducted a crosswalk between the NCTM expectations and the three states' extended mathematics standards. This crosswalk identified areas of overlap and uniqueness in the NCTM expectations that each state reflected through its extended standards. The crosswalk resulted in identification of a total of 30 NCTM expectations that were common across all three states in grades 3 through 8 and high school. Ten of these expectations were in the domain of numbers and operations and between four to six expectations were in each of the remaining four domains. These expectations were the foundation on which the design patterns and associated tasks were developed.

The ECD Co-Design Process

Co-design is a process of bringing together the expertise of assessment specialists, special educators, and content area specialists to create design patterns, task templates, task specifications, and exemplar tasks. In this project, the co-design team members were specialists in large-scale and formative assessment, special educators of SWSCD, and mathematics educators with experience in instruction and assessment. This constellation of expertise was contributed by members of each state department of education, by SRI International, and by nationally recognized experts in special education, mathematics, and assessment.

To facilitate the design process, project staff used products that are associated with the PADI online assessment design system. This technology systematically supports the design of evidence-based assessment items and tasks using design patterns and task templates

Step 1: Co-design team develops design patterns. Design patterns are guiding structures that are part of the domain modeling layer of ECD. Design patterns comprise attributes that are necessary for constructing an evidentiary-based assessment. These attributes are based on the work of Messick (1994) and Mislevy and his colleagues (Mislevy et al., 2003; Mislevy, Steinberg, and Almond, 2003). Each design pattern articulates an assessment argument by identifying the focal KSAs that are to be measured, the kinds of observations that can provide evidence of this knowledge or skill, and the features of task situations that allow the students to provide this evidence. Also specified in the design patterns are any nonfocal KSAs that may be required for students to respond correctly to the assessment tasks but are not the target of the assessment task (for example, reading comprehension and decoding skills needed to respond to a mathematics word problem). Design patterns also capture the ways assessment tasks can be varied to increase or decrease demands for knowledge and specify the work products and rubrics that the assessment designer may want to use. In the AAD-M project, 30 design patterns in mathematics were created and apply to the mathematics content covered in each state.[3]

[3] For more information about design patterns, see PADI Technical Report 1, *Design Patterns for Assessing Science Inquiry* (Mislevy, R., Hamel, L., Fried, R. G., Gaffney, T., Haertel, G., Hafter, A., … Wenk, A., 2003, Menlo Park, CA: SRI International); PADI Technical Report 5, *The Case for an Integrated Design Framework for Assessing Science Inquiry* (Baxter, G., & Mislevy, R., 2005, Menlo Park, CA: SRI International); and Technical Report 8, *An*

Step 2. Co-design team develops summary task templates. A summary task template was completed for each state. This template provides an overview of the assessment system used by the state including an overview of its student model, which consists of the constructs to be assessed (for example, overall mathematics proficiency; subdomain proficiency as appropriate, such as numbers and operation, geometry, etc.); scoring and evaluation rubrics; measurement models; and descriptions of the kinds of stimulus materials and presentation used in each state's tasks and items.

Step 3. The co-design team develops task specifications and authors tasks. Task specifications provide guidelines for the design of individual assessment tasks. Designers specify the particular stimuli and response options that will be presented to students. For example, in a task specification, the designer indicates that four data points (rather than three or five) will be presented to students who are asked to create a line graph. In completing the task specification, designers also specify how students' responses will be scored, give administration guidelines, and identify the variable features that can be used to increase or decrease the difficulty of the tasks. Prompts, graphics, diagrams, and supporting materials are described in detail for each task.

Thirty task specifications are linked to each of the 30 design patterns; four tasks are associated with each task specification. For each task specification, the first task is designed to be the most cognitively complex and to assess one of the following depth-of-knowledge (DOK) levels: Application, Comprehension, or Performance.[4] The second task is designed to be less complex and targets a lower DOK level (either Performance or Recall). The third task is even less complex and targets the Recall DOK level. If students are unable to respond to the third task at the Recall level, he or she is asked to respond to a task at the Attention DOK level (the fourth task). The first and second tasks are designed to align with a single focal KSA which was selected to be the target of the assessment tasks at the beginning of the co-design process. The third task is designed to align with an additional KSA, which is also selected at the beginning of the co-design process. The additional KSA could be described as a foundational skill in that it is typically a prerequisite for successful performance on the first and second tasks.

Step 4. State department of education staff and teachers create additional performance tasks. Using the design patterns, summary task templates, task specifications, and the written description of exemplar tasks, state department of education (SDE) staff members and selected teachers will replicate the process of generating performance tasks for AA-AAS in mathematics with guidance from the SRI team.

Design for Pilot Task Tryouts

The collaborating states will pilot-test the newly developed assessment tasks with teachers administering them to students eligible to take state AA-AAS. Grant funds will be used to reimburse teachers for their time in administering the pilot tests. The details of the pilot-testing, such as sampling criteria and size, timing and scheduling, recruitment, administration, and data collection activities, are summarized below and will be fully described in a future technical

Example-Based Exploration of Design Patterns in Measurement (DeBarger, A. H., & Riconscente, M., 2005, Menlo Park, CA: SRI International). Technical reports are available at padi.sri.com.

[4] Flowers, C., Wakeman, S. Y., Browder, D. M., & Karvonen, M. (2007). *Links for Academic Learning: An Alignment Protocol for Alternate Assessments Based on Alternate Achievement Standards.* Charlotte, NC: National Alternate Assessment Center, University Of North Carolina at Charlotte.

report (Technical Report 7: Pilot Test Design). To pilot the 120 newly designed tasks, each state will gather information by administering tasks using common instructions. The focus of the pilot will be to collect information about task variability and the appropriateness of the tasks to measure a range of student performance levels.

Task viability. Teachers will administer pilot task tryouts to students to judge the viability of the tasks. Can the four tasks associated with a design pattern be administered as designed? Are the task instructions and materials clear to the teacher? Are they clear to the student? Data will be collected through a teacher questionnaire and observations of task administration. Data will inform improvement of the tasks.

Appropriateness of tasks to measure a range of student performance levels. Tasks will be administered to students with significant cognitive disabilities whose teachers characterize them as demonstrating low, medium, or high symbolic functioning (Browder et al., 2008). Which students successfully perform the first task (most complex task) and at what level of symbolic functioning? Which students successfully perform the second, third, and fourth tasks (decreasingly less complex tasks) and at what level of symbolic functioning? Data will inform modification of tasks so all or most students can gain access to at least one task associated with each design pattern.

Thirty-six tasks associated with nine task specifications that represent the five conceptual areas of the NCTM and span two or more grade bands are considered core tasks. Collecting data from each state on these 36 core tasks spanning two or more grade bands permits a review of the progression of complexity and DOK from one grade span to the next. These core tasks will be piloted with 216 students in each of three states ($N = 648$). Twenty-eight additional tasks associated with seven task specifications will be piloted by each state so that all 120 tasks associated with the 30 task specifications will be piloted. (See Appendix A for a list of NCTM expectations linked to the design patterns.)

Dissemination Plan

Holding high expectations for the academic achievement of students with significant cognitive disabilities is now a widely accepted practice, in part because of its being included in legislation: IDEA in 1997 and 2004; ESEA in 2001. Evidence-centered design holds promise as a way of furthering this practice by providing a method of ensuring that challenging academic content will be the focus of AA-AAS test design. By applying ECD principles to the development of alternate assessments for students with significant cognitive disabilities, the AAD-M project is breaking new ground. It is essential that the outcomes and lessons learned from this research be shared with other states so that the quality of assessment for this population can be improved nationally. To achieve this goal, several dissemination activities are planned, as described below.

Website. SRI has developed a project website to post study reports and selected products that will be available to the public. The site includes links to relevant websites, such as those for the Council of Chief State School Officers (CCSSO), NAAC (National Alternate Assessment Center), NCEO (National Center on Education Outcomes), and the NCSA (National Conference on Student Assessment), and has links to the websites for each of the collaborating states and related contact information. The URL for the site will be distributed to interested parties at national meetings and presentations and via a listserv provided by SRI. In the final months of the project, the website will be transferred to the U.S. Department of Education for continued maintenance.

Webinar. SRI will host a free 2-hour web-based interactive seminar presentation to disseminate project procedures and outcomes and to answer questions. Information about and invitations to attend the webinar will be on the website and distributed via e-mail using the listserv described above.

Technical Report Series including Procedural Guidelines. As each report in the Technical Report series is completed, it will be posted to the project website. Two of the technical reports are user-friendly documents describing the process and procedures used in the AAD-M project so that states outside the consortium can learn about the ECD principles used. The planned technical reports are as follows.

Technical Report 1: *Project Overview: Applying Evidence-Centered Design to Alternate Assessments in Mathematics for Students with Significant Cognitive Disabilities.*

Technical Report 2: *Current State of Mathematics Assessment in Alternate Assessment* A description of (1) the state of the art in alternate assessment in mathematics and (2) the current state of practice in alternate assessment design in mathematics.

Technical Report 3: *Crosswalk — Domain Analysis Aligning NCTM Expectations with State Extended Mathematics Standards.* The results of an analysis of the three collaborating states' extended content standards to identify common NCTM expectations for which evidence-centered design patterns, task specifications, and assessment tasks were completed.

Technical Report 4: *Design Patterns — Developing Design Patterns for Students with Significant Cognitive Disabilities in Mathematics*. A description of the theoretical foundations of evidence-centered design that underlie design patterns, the processes used to create design patterns, and a description of the current library of patterns available in this project.

Technical Report 5: *Synergistic Use of Evidence-Centered Design and Universal Design for Learning for Improved Assessment Design*. A brief description of the integration of the ECD and UDL approaches used to develop the assessment design tools in this project.

Technical Report 6: *Design and Development of Assessment Tasks.* A description of the application of ECD and UDL that underlies the newly developed assessment tasks, the co-design process used to develop the tasks, and the library of tasks available from this project.

Technical Report 7: *Pilot Task Tryouts Design*. The design of the tryouts, including a description of the sample of students to be tested, the logistics of the data collection, and the qualitative and quantitative analyses completed.

Technical Report 8: *Finding, Conclusions, and Recommendations for AAD-M Tasks*. The qualitative and quantitative findings of the assessment task tryouts, conclusions about the design and development process, and recommendations for further research and development.

Technical Report 9: *Procedural Guidelines for Design Patterns*. The steps to follow in applying the co-design process to the creation of design patterns using ECD and UDL.

Technical Report 10. *Procedural Guidelines for Assessment Tasks.* The steps to follow in applying the co-design process to create assessment tasks using ECD and UDL

Technical Report 11: *Project Evaluation Results.* The final report prepared by the external project evaluator, including findings and recommendations.

References

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 1999. *Standards for Educational and Psychological Testing NEED FULL REF*

Bechard, S. (2005). *Developing alternate assessments using expanded benchmarks from a nine-state consensus framework in reading, writing, mathematics, and science.* Paper presented at the 2005 National Council on Measurement in Education (NCME) Annual Conference, Montreal, Canada.

Browder, D. M., & Cooper-Duffy, K. (2003). Evidence-based practices for students with severe disabilities and the requirement for accountability in "No Child Left Behind." *The Journal of Special Education*, *37*(3), 157–163.

Browder, D. M., & Spooner, F. (2003). Understanding the purpose and process of alternate assessment. In D. Ryndak & S. Alper (Eds), *Curriculum and instruction for students with significant disabilities in inclusive settings* (2nd ed., pp. 51-72). Boston: Allyn & Bacon.

Browder, D. M., Flowers, C., & Wakeman, S. Y. (2008). Facilitating participation in assessments and the general curriculum: Level of symbolic communication classification for students with significant cognitive disabilities. *Assessment in Education: Principles, Policy, and Practice*, *15*(2), 137–151.

Browder, D., Spooner, F., Ahlgrim-Delzell, L., Flowers, C., Algozzine, B., & Karvonen, M. (2004). A content analysis of the curricular philosophies reflected in states' alternate assessment performance indicators. *Research and Practice for Persons with Severe Disabilities*, 28, 165–181.

Cameto, R., Knokey, A.-M., Nagle, K., Sanford, C., Blackorby, J., Sinclair, B., & Riley, D. (2009). *National study on alternate assessments: National profile on alternate assessments based on alternate achievement standards.* (NCSER 2009-3014). Menlo Park, CA: SRI International.

Cameto, R., Bergland, F., Knokey, A.-M., Nagle, K., Sanford, C., Kalb, S. C., Blackorby, J., Sinclair, B., Riley, D., & Ortega, M. (2010, April). *Teacher perspectives of school level implementation of alternate assessments for students with significant cognitive disabilities*. A report from the National Study on Alternate Assessments (NCSER 2010-3007). Menlo Park, CA: SRI International

CAST (2008). *Universal design for learning guidelines version 1.0.* Wakefield, MA: Author.

Cortiella, C. (2007). *Rewards and roadblocks: How special education students are faring under No Child Left Behind*. Retrieved April 29, 2010, from http://www.ncld.org/images/ stories/oncapitolhill/policyrelatedpublications/rewardsandroadblocks/rewardsandroadblocks.pdf

Flowers, C., Wakeman, S. Y., Browder, D. M., & Karvonen, M. (2007). *Links for academic learning: An alignment protocol for alternate assessments based on alternate achievement standards.* Charlotte, NC: National Alternate Assessment Center, University of North Carolina at Charlotte.

Gong, B., & Marion, S. (2006). *Dealing with flexibility in assessments for students with significant cognitive disabilities.* Minneapolis: University of Minnesota, National Center On Educational Outcomes. Available at http://cehd.umn.edu/NCEO/onlinepubs/ synthesis60. html

Government Accountability Office. (2009). *No Child Left Behind: Enhancements in the Department Of Education's review process could improve state academic assessments* [GAO 09-911]. Retrieved May 4, 2010, from www. gao. gov/new. items/D09911. pdf

Kearns, J. (2007). *Defining the population to be served.* Paper presented at the 8th annual Maryland Conference on Alternate Assessments. University of Maryland, College Park.

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

Kohl, F. L., Mclaughlin, M. J., & Nagle, K. M. (2006). Alternate achievement standards and assessments: A descriptive investigation of 16 states. *Exceptional Children, 73*(1), 107–123.

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

McDonnell, J. J., Hardman, M. L., & McDonnell, A. P. (2003). A*n introduction to persons with moderate and severe disabilities: Educational and social issues* (2nd Ed.). Boston, MA: Allyn and Bacon.

Messick, S. (1994). The interplay of evidence and consequences in the validation of performance assessments. *Educational Researcher, 23*(2), 13–23.

Mislevy, R., & Haertel, G. (2006). Implications of evidence-centered design for educational testing. *Educational Measurement: Issues and Practice, 24*(4), 6–20.

Mislevy, R., Hamel, L., Fried, R., Gaffney, T., Haertel, G., Hafter, A., … Wenk, A. (2003). *Design patterns for assessing science inquiry* (PADI Technical Report 1). Menlo Park, CA: SRI International.

Mislevy, R. J., Steinberg, L. S., & Almond, R. G. (2003). On the structure of educational assessments. *Measurement: Interdisciplinary Research and Perspectives, 1*, 3–67.

National Council of Teachers of Mathematics (2005). *Principles and standards for school mathematics.* Reston, VA: Author.

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

Quenemoen, R., Kearns, J., Quenemoen, M., Flowers, C., & Kleinert, H. (2010). Common misperceptions and research-based recommendations for alternate assessment based on alternate achievement standards (Synthesis Report 73). Minneapolis: University Of Minnesota, National Center on Educational Outcomes. Retrieved April 22, 2010, from http:// www. cehd. umn. edu/NCEO/onlinepubs/synthesis73/synthesis73. pdf

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: 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: University of Minnesota, National Center on Educational Outcomes.

Ryan, J. M., Quenemoen, R. F., & Thurlow, M. L. (2004). *I say potato, you say potato: The assessment-speak gap between general and alternate assessment experts.* Minneapolis: University of Minnesota, National Center on Educational Outcomes.

Schafer, W. D. (2005). Technical documentation for alternate assessment. *Practical Assessment, Research and Evaluation, 10*(10), 1–9.

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: University of Minnesota, National Center on Educational Outcomes.

Thompson, S., & Thurlow, M. (2003). *2003 State special education outcomes: Marching on*. Minneapolis: University of Minnesota, National Center on Educational Outcomes. Available at http://www.education.umn.edu/NCEO/onlinepubs/2003statereport.html

Tindal, G., McDonald, M., Tedesco, M., Glasgow, A., Almond, P., Crawford, L., & Hollenbeck, K. (2003). Alternate assessments in reading and math: Development and validation for students with significant disabilities. *Exceptional Children, 69*(4), 481–494.

Towles-Reeves, E., Garrett, B., Burdette, P., & Burdge, M. (2006). Validation of large-scale alternate assessment systems and their influence on instruction—What are the consequences? *Assessment for Effective Intervention, 31*(3), 45–57.

U. S. Department of Education. (2005). *Alternate achievement standards for students with the most significant cognitive disabilities. Non-regulatory guidance*. Washington, DC: Author. Available at http://www.ed.gov/policy/elsec/guid/altguidance.doc

Wagner, M., Balladeers, T., & Marder, C. (with Cameto, R., Cardoso, D., Garza, N., Levine, P., & Newman, L.). (2003). *Life Outside the Classroom for Youth with Disabilities. A Report from the National Longitudinal Transition Study-2 (NLTS2).* Menlo Park, CA: SRI International. Available at www.nlts2.org/reports/2003_04-2/nlts2_report_2003_04-2_complete.pdf.

Ysseldyke, J., Dennison, A., & Nelson, R. (2003). *Large-scale assessment and accountability systems: Positive consequences for students with disabilities* (Synthesis Report 51). Minneapolis: University of Minnesota, National Center on Educational Outcomes.

Appendix: NCTM Expectations Associated with Core and Additional Tasks for Pilot Testing

CORE TASKS

Number and Operations

1. A3 grades 3–5: Develop understanding of fractions as parts of unit wholes, as parts of a collection, as locations on number lines, and as divisions of whole numbers
2. A1 grades 6–8: Work flexibly with fractions, decimals, and percents to solve problems

Algebra

3. B2 grades 3–5: Represent the idea of a variable as an unknown quantity using a letter or a symbol
4. B1 grades 6–8: Develop an initial conceptual understanding of different uses of variables
5. B3 grades 9–12: Use symbolic algebra to represent and explain mathematical relationships

Geometry

6. A1 grades 3–5: Identify, compare, and analyze attributes of two- and three-dimensional shapes and develop vocabulary to describe the attributes
7. A1 grades 6–8: Precisely describe, classify, and understand relationships among types of two- and three-dimensional objects using their defining properties
8. A1 grades 9–12: Analyze properties and determine attributes of two- and three-dimensional objects

Measurement

9. B2 grades 3–5: Select and apply appropriate standard units and tools to measure length, area, volume, weight, time, temperature, and the size of angles

ADDITIONAL TASKS

Utah:

Number and Operations

1. A1 grades 3–5: Understand the place-value structure of the base-ten number system and be able to represent and compare whole numbers and decimals
2. A2 grades 3–5: Recognize equivalent representations for the same number and generate them by decomposing and composing numbers
3. A7 grades 6–8: Develop meaning for integers and represent and compare quantities with them

Algebra

4. C1 grades 3–5: Model problem situations with objects and use representations such as graphs, tables, and equations to draw conclusions

Data Analysis and Probability

5. A2 grades 6–8: Select, create, and use appropriate graphical representations of data, including histograms, box plots, and scatter plots

Measurement

6. B3 grades 3–5: Select and use benchmarks to estimate measurements

7. A1 grades 9–12: Make decisions about units and scales that are appropriate for problem situations involving measurement

Florida:

Number and Operations

1. A4 grades 3–5: Use models, benchmarks, and equivalent forms to judge the size of fractions
2. B1 grades 3–5: Understand various meanings of multiplication and division
3. C4 grades 6–8: Develop, analyze, and explain methods for solving problems involving proportions, such as scaling and finding equivalent ratios
4. A1 grades 9–12: Develop a deeper understanding of very large and very small numbers and of various representations of them

Data Analysis and Probability

5. A3 grades 3–5: Represent data using tables and graphs such as line plots, bar graphs, and line graphs

Geometry

6. A4 grades 3–5: Explore congruence and similarity

Measurement

7. A2 grades 6–8: Understand relationships among units and convert from one unit to another within the same system

Idaho:

Number and Operations

1. B3 grades 3–5: Identify and use relationships between operations, such as division as the inverse of multiplication, to solve problems
2. C2 grades 3–5: Develop fluency in adding, subtracting, multiplying, and dividing whole numbers
3. C1 grades 9–12: Develop fluency in operations with real numbers, vectors, and matrices, using mental computation or paper-and-pencil calculations for simple cases and technology for more-complicated cases

Algebra

4. A1 grades 3–5: Describe, extend, and make generalizations about geometric and numeric pattern

Data Analysis and Probability

5. B1 grades 3–5: Describe the shape and important features of a set of data and compare related data sets, with an emphasis on how the data are distributed
6. B1 grades 9–12: For univariate measurement data, be able to display the distribution, describe its shape, and select and calculate summary statistics

Measurement

7. B2 grades 6–8: Select and apply techniques and tools to accurately find length, area, volume, and angle measures to appropriate levels of precision