How Design Patterns Integrate Universal Design for Learning (UDL) Into Assessments for Students With Disabilities

To be presented at the 2011 annual meeting of the American Educational Research Association, New Orleans, LA.

Study One Authors:
Elizabeth Murray, Center for Applied Technologies, Inc.
Britte Haugan Cheng and Geneva Haertel, SRI International

Study Two Authors:
Renee Cameto, Geneva Haertel, Angela DeBarger, and Kathryn Morrison
SRI International

Table Of Contents

Designing Large-Scale Science Assessment Tasks for Students with High-Incidence Disabilities: The Use of Evidence-Centered Design, Design Patterns, and Universal Design for Learning

Elizabeth Murray
CAST, Inc.

Britte Haugan Cheng and Geneva Haertel
SRI International

Purpose

The No Child Left Behind Act requires that students with disabilities be included in state assessments and accountability. However, the use of accommodations, modifications, and alternate assessments to permit the inclusion of students with disabilities has given rise to a number of issues related to fairness and test validity. Recently, researchers have begun to explore whether tests can be designed from the outset to be more accessible and valid for a wider range of students; this approach is termed "universal design." The researchers on this project are studying the use of universal design for learning (UDL) paired with an approach termed "evidence-centered design" (ECD) to redesign or develop assessment items that can more accurately evaluate the knowledge and skills of all students on statewide tests. The academic content focus of this study is middle school science, but if successful the approach can be applied to other topics and age ranges. In this study, the researchers' specific goals are (1) to evaluate the validity of inferences that can be drawn from existing state science assessments for students with and without high incidence disabilities (learning disabilities and mild mental retardation), (2) to redesign assessment items to increase the validity for students both with and without disabilities, (3) to conduct empirical studies of the validity of inferences drawn from the scores on the redesigned items, and (4) to develop research-based guidelines that can be used in large-scale assessment design and development to increase the validity of inferences from science assessment scores for all students.

This paper describes a design methodology for improving the validity of inferences about the performance of students with disabilities on large-scale science assessments. We present work to date from a study that combines the use of "universal design for learning" (UDL) with "evidence-centered design" (ECD) to redesign statewide science items to more accurately evaluate the knowledge and skills of all students, including those with high incidence disabilities. The presentation will: (1) describe the state of science assessment for students with disabilities, (2) overview the ECD and UDL frameworks and describe how these frameworks were integrated within a working Web-based assessment design system; (3) describe how the Web-based system helps guide designers through the complex decisions prerequisite to the development of assessments for students with disabilities; and (4) present examples of redesigned science assessment items and design documentation.

State of Science Assessment for Students with Disabilities

Our Focus on Middle School Science

The decision to focus project research on science assessments for students with disabilities was motivated by the extension of NCLB to science in 2007 and an understanding that success in science coursework serves as a pipeline to scientific careers, as well as greater postsecondary education and labor market opportunities for students. Our focus on middle school level students was motivated by the formal introduction of scientific reasoning and problem solving in grades 6 through 8 and by the interdependence of reading, math, and science knowledge, skills and abilities. Science instruction and assessment are noted for abstract content, challenging vocabulary, text (books) written at difficult readability levels, and complex lab activities. Inability to successfully engage with these curricula and the more complex science content can lead to high school students' decisions to opt out of science classes and scientific career trajectories. Although some special education researchers have developed interventions and outlined best practices for instructing students with disabilities in science (Mastropieri & Scruggs, 1992; Mastropieri & Scruggs, 1995; McClery & Tindal, 1999; Norman, Caseau, & Stefanisch, 1998), science education for students with disabilities has historically been a lower priority in research programs than reading/language arts and mathematics. Moreover, whereas science assessment tasks that entail declarative and procedural knowledge (Li & Shavelson, 2001) require students to recognize and recall information, tasks that entail schematic or strategic knowledge further challenge students to execute or evaluate problem solutions as well as to judge the appropriateness of knowledge applied — precisely the areas affected by many cognition-based disabilities.

NLTS2 Background

The National Longitudinal Transition Study-2 (NLTS2), funded by the U.S. Department of Education, Institute of Education Sciences (IES), collected information from parents, youth, and schools from 2001 to 2010. The study provides a national picture of the educational programs, accommodations, and in-and out-of-school outcomes of young people with disabilities as they transition from secondary school to early adulthood roles. The NLTS-2 sample is comprised of 11,275 students in all disability categories stratified by geographic region and Local Education Agency (LEA) size and LEA wealth. NLTS-2 data summaries generalize to the national population of youth with disabilities, as well as to each disability category individually. NLTS-2 collected longitudinal data via telephone surveys of parents and youth, paper-based surveys of teachers, and face-to-face assessments of academic performance. To be included in the assessment, students were required to be able to speak and understand English or ASL and be able to complete all measures required for the study using the same accommodations provided to them in the course of day to day instruction and assessment. Data presented here are student scores on the version Woodcock Johnson III (WJ3) assessment (Woodcock, McGrew, & Mather, 2001). Data from four subtests were obtained: science concepts, applied problems, passage comprehension, and calculation. This study examines the science concepts subtest scores.

Evidence from the WJ3 (Woodcock, McGrew, & Mather, 2001) shows that students with disabilities have difficulties in science in addition to reading and mathematics. On average, students with disabilities nationally score at the 24th percentile on the science concepts subtest, and one in three students has scores below the 5th percentile. Figure 1 illustrates that there is considerable variation in performance both within and across disability categories. This implies that the careful analysis of task requirements, both relevant and irrelevant to the constructs being measured, is appropriate in this project, as task requirements may differentially reflect performance of students with different disabilities. For example, in the boxplot below (see Figure 1), students with mental retardation, autism, multiple disability diagnoses, and traumatic brain injury have both significantly lower and less variable performance relative to students in other disabilities (for example, learning disability, visual impairment) on the science concepts subtest. This pattern is evident in all subject areas tested as part of NLTS-2, including the mathematics applied problems subtest that presents students with problems consistent with the types of scientific reasoning commonly introduced in middle school science. It is important to note that the student performance on the WJ3 passage comprehension and calculation subtests are even more variable than those shown below.

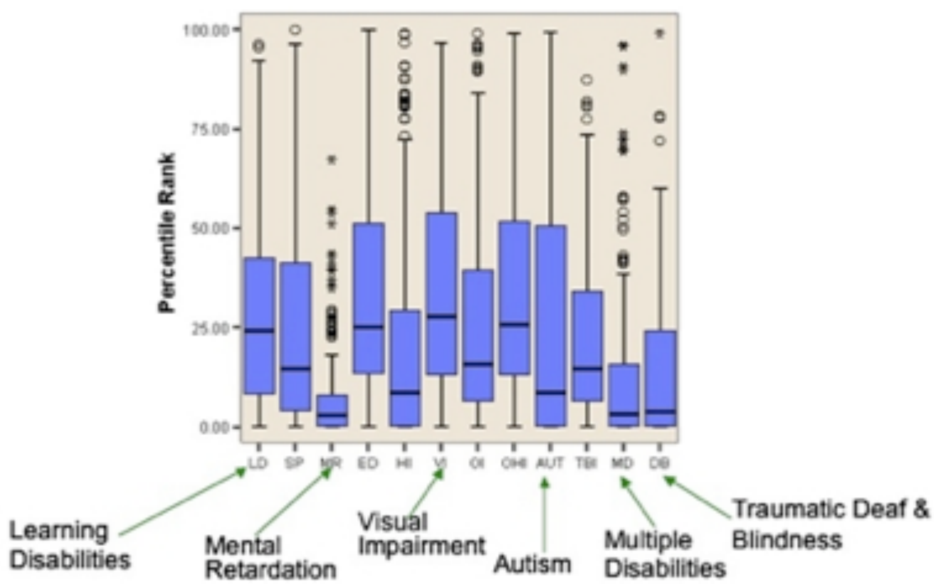

Figure 1. Scores on the WJ3 Science Concepts Subtest

Figure 2 presents an alternative view of these data that illustrates the need to better understand the ways that disabilities differentially limit student performance on assessments. Given that achievement data are often reported in terms of proportions of students above or below a particular threshold, we use this method with a 40% percentile rank threshold to illustrate the variation found in scores on the WJ3 science concepts subtest. This is the way that accountability systems organize achievement data and also provides a rough estimate of the level of improvement required for students to meet proficiency targets. This figure illustrates that students with disabilities may not be well represented by traditional reporting of achievement data, due to the highly skewed distribution of achievement scores, although the degree of skewness is highly dependent on students' disability category. Variation across disability categories indicates that, to more accurately gauge student performance (and possibly progress over time), we need to understand the variation in student achievement data and, specifically, assessment designers need to document the ways we anticipate task requirements, students' abilities, and their particular disabilities will interact in testing situations. In addition, from a policy perspective, it is important to explore if some of the achievement gap in test scores can be closed by improvements in test design and administration; for example, by applying Universal Design for Learning principles implemented through the Principled Assessment Designs for Inquiry (PADI) online assessment design system-- the goal of our project.

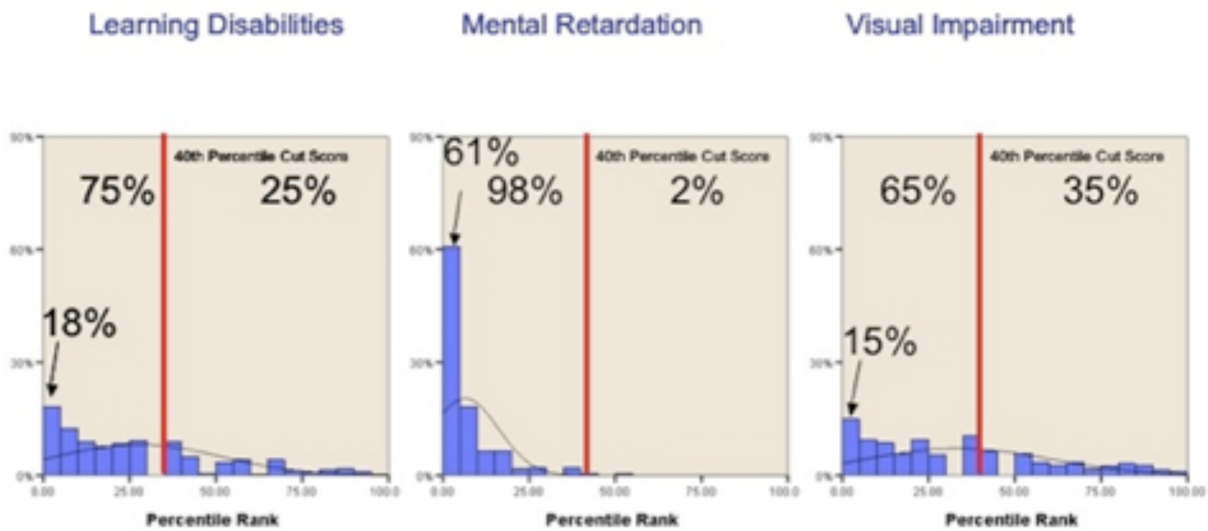

Figure 2. Scores on the WJ3 Science Concepts Subtest by Disability Category

Theoretical Frameworks

The redesign of statewide science items in this project was based on principles of ECD and UDL. In the following section, we describe the principles that underlie each of these frameworks.

Universal Design for Learning

Universal Design for Learning (UDL) helps to meet the challenge of diversity by suggesting flexible assessment materials, techniques, and strategies (Dolan, Rose, Burling, Harris & Way, 2007). The flexibility of UDL empowers assessors to meet the varied needs of students and to accurately measure student progress. The UDL framework includes three guiding principles that address three critical aspects of any learning activity, including its assessment. The first principle, multiple means of representation, addresses the ways in which information is presented. The second principle is multiple means of action and expression. This principle focuses on the ways in which students can interact with content and express what they are learning. Multiple means of engagement is the third principle, addressing the ways in which students are engaged in learning (Meyer & Rose, 2006; Rose & Meyer, 2002; Rose, Meyer, & Hitchcock, 2005). These principles provide structure for the infusion of UDL into assessment.

Principle I. Provide Multiple Means of Representation (the "what" of learning). Students differ in the ways that they perceive and comprehend information that is presented to them. For example, those with sensory disabilities (e.g., blindness or deafness), learning disabilities (e.g., dyslexia), language or cultural differences, and so forth, may all require different ways of approaching content. Others may simply grasp information better through visual or auditory means rather than printed text. In reality, there is no one means of representation that will be optimal for all students; providing options for representation is essential.

Principle II: Provide Multiple Means of Action and Expression (the "how" of learning). Students differ in the ways that they can interact with materials and express what they know. For example, individuals with significant motor disabilities (e.g. cerebral palsy), those who struggle with strategic and organizational abilities (executive function disorders, ADHD), those who have language barriers, and so forth, approach learning tasks very differently and will demonstrate their mastery very differently. Some may be able to express themselves well in writing text but not oral speech, and vice versa. In reality, there is no one means of expression that will be optimal for all students; providing options for expression is essential

Principle III: Provide Multiple Means of Engagement (the "why" of learning). Affect represents a crucial component to learning. Students differ markedly in the ways in which they can be engaged or motivated to learn. Some students enjoy spontaneity and novelty, while others do not, preferring strict routine. Some will persist with highly challenging tasks while others will give up quickly. In reality, there is no one means of engagement that will be optimal for all students; providing multiple options for engagement is essential.

Evidence-Centered Design

Evidence-centered assessment design (ECD) was formulated by Robert Mislevy, Linda Steinberg, and Russell Almond (2003) at Educational Testing Service. ECD builds on developments in fields such as expert systems (Breese, Goldman, & Wellman, 1994), software design (Gamma, Helm, Johnson, & Vlissides, 1994), and legal argumentation (Tillers & Schum, 1991) to make explicit, and to provide tools for, building assessment arguments that help in both designing new assessments and understanding familiar ones (Mislevy & Riconscente, 2005). Two complementary ideas organize the effort. The first is an overarching conception of assessment as an argument from imperfect evidence. Specifically, it involves making explicit the claims (the inferences that one intends to make based on scores) and the nature of the evidence that supports those claims (Hansen & Mislevy, 2008). The second idea is distinguishing layers at which activities and structures appear in the assessment enterprise, all to the end of instantiating an assessment argument in operational processes. By making the underlying evidentiary argument more explicit, the framework makes operational elements more amenable to examination, sharing and refinement. Making the argument more explicit also helps designers meet diverse assessment needs caused by changing technological, social and legal environments (Hansen & Mislevy, 2008).

Domain Analysis. In ECD, assessment is expressed in layers that provide structure for different kinds of work and information at different stages of the process. In the Domain Analysis layer, research and experience about the domains and skills of interest are gathered—information about the knowledge, skills, and abilities (KSAs) of interest, the ways people acquire KSAs and use them, the situations under which the KSAs are employed, and the indicators of successful application of the KSAs.

Domain Modeling. In the Domain Modeling layer, information from Domain Analysis is organized to form the assessment argument. Domain modeling structures the outcomes of domain analysis in a form that reflects the narrative structure of an assessment argument, in order to ground the more technical models in the next layer. The PADI Online Assessment Design System uses objects called design patterns to assist task designers with domain modeling. Design patterns play a key role in the present project, as we consider the impact of universal design for learning principles and accommodations on task design and evaluation.

Conceptual Assessment Framework (CAF). The CAF layer concerns technical specifications for operational elements including measurement models, scoring methods, test assembly specifications, and requirements and protocols for assessment delivery. An assessment argument laid out in narrative form at the Domain Modeling layer is here expressed in terms of coordinated pieces of machinery: specifications for tasks, measurement models, scoring methods, and delivery requirements within templates. The central models within the CAF are the Student Model, Evidence Model, and Task Model. In addition, the Assembly Model determines how tasks are assembled into tests, the Presentation Model indicates the requirements for interaction with a student (e.g., simulator requirements), and the Delivery Model specifies requirements for the operational setting. Details about task features, measurement-model parameters, stimulus material specifications, and the like are expressed in the CAF model templates in terms of knowledge representations and data structures, which guide their implementation and ensure their coordination. These templates are essentially blueprints that specify, at a meta-level, the necessary elements for tasks. The present project includes some work at the CAF layer, as we developed example templates that demonstrate how tasks can be developed in accordance with UDL principles and modified in accordance with student needs.

Assessment Implementation. The work in this layer includes activities in preparation for testing examinees such as authoring tasks, calibrating items, finalizing rubrics, producing materials, producing presentation environments, and training interviewers and scorers, all in accordance with the assessment arguments and test specifications created in previous layers of ECD. The ECD approach links the rationales for each layer back to the assessment argument and provides structures that support opportunities for reuse and interoperability.

Assessment Delivery. The work in this layer includes activities, such as presenting tasks to examinees, evaluating performances to assign scores, and reporting the results to provide feedback to students themselves, teachers, decision-makers, or other stakeholders.

The ECD framework described in this report applies principles of evidentiary reasoning to handle the complexities of the validity argument (Cronbach & Meehl, 1955; Messick, 1989, 1994; Kane, 1992) associated with accessibility features. The key idea is to lay out the evidentiary structures, what may be termed the validity argument (or "validation argument" [National Research Council, 2004, p. 104]). An assessment argument can be summarized as comprising: (a) a claim about a person possessing at a given level a certain targeted proficiency, (b) the data (e.g., scores) that would likely result if the person possessed, at a certain level, the targeted proficiency, (c) the warrant (or rationale, based on theory and experience) that tells why the person's level of the targeted proficiency would yield the expected score, and (d) "alternative explanations" for the person's high or low scores (i.e., explanations other than the person's level of the targeted proficiency). The existence of *alternative explanations* that are both significant and credible might indicate that *validity* is threatened or being *compromised* (Messick, 1989).

Much of the analysis that is the focus of this project has to do with these alternative explanations, factors that can hinder an assessment from yielding valid inferences. When such alternative explanations are recognized at the earliest stages of test design, then later rework and retrofitting can be avoided. The ECD accessibility effort has focused on building argument structures that might help anticipate and address key details of these alternative explanations particularly as they relate to test takers with disabilities (Hansen & Mislevy, 2008).

Integration of UDL and ECD in PADI Online Assessment Design System

Principled Assessment Designs for Inquiry (PADI) was a project supported by the National Science Foundation to improve the assessment of science inquiry (through the Interagency Educational Research Initiative under grant REC-0129331). The PADI project has developed a design framework for assessment tasks based on the evidence-centered design (ECD) framework. PADI was developed as a system for designing blueprints for assessment tasks, with a particular eye toward science inquiry tasks—tasks that stress scientific concepts, problem solving, building models, using models, and cycles of investigation. The PADI framework guides an assessment developer's work through design structures that embody assessment arguments and take advantage of the commonalities across the assessments for sharing and reusing conceptual and operational elements (Mislevy & Haertel, 2006). PADI provides a conceptual framework, data structures, and software supporting tools for this work. The PADI online assessment design system is fully operational.

ECD seeks to integrate the processes of assessment design, authoring, delivery, scoring, and reporting. Work within PADI, however, is focused on design layers that lie above the level of specific environments for task authoring and assessment delivery. The key PADI design objects used in the present project are *design patterns* and *templates*.

PADI assessment *design patterns* (analogous to those in architecture and software engineering) capture design rationale in a reusable and generative form in the domain modeling layer of assessment. They help designers think through substantive aspects of an assessment argument in a structure that spans specific domains, forms, grades, and purposes (Mislevy et al., 2003). Assessment designers working with the PADI design system use the web-based design interface illustrated for design patterns. (See Design Pattern in Figure 3 below.)

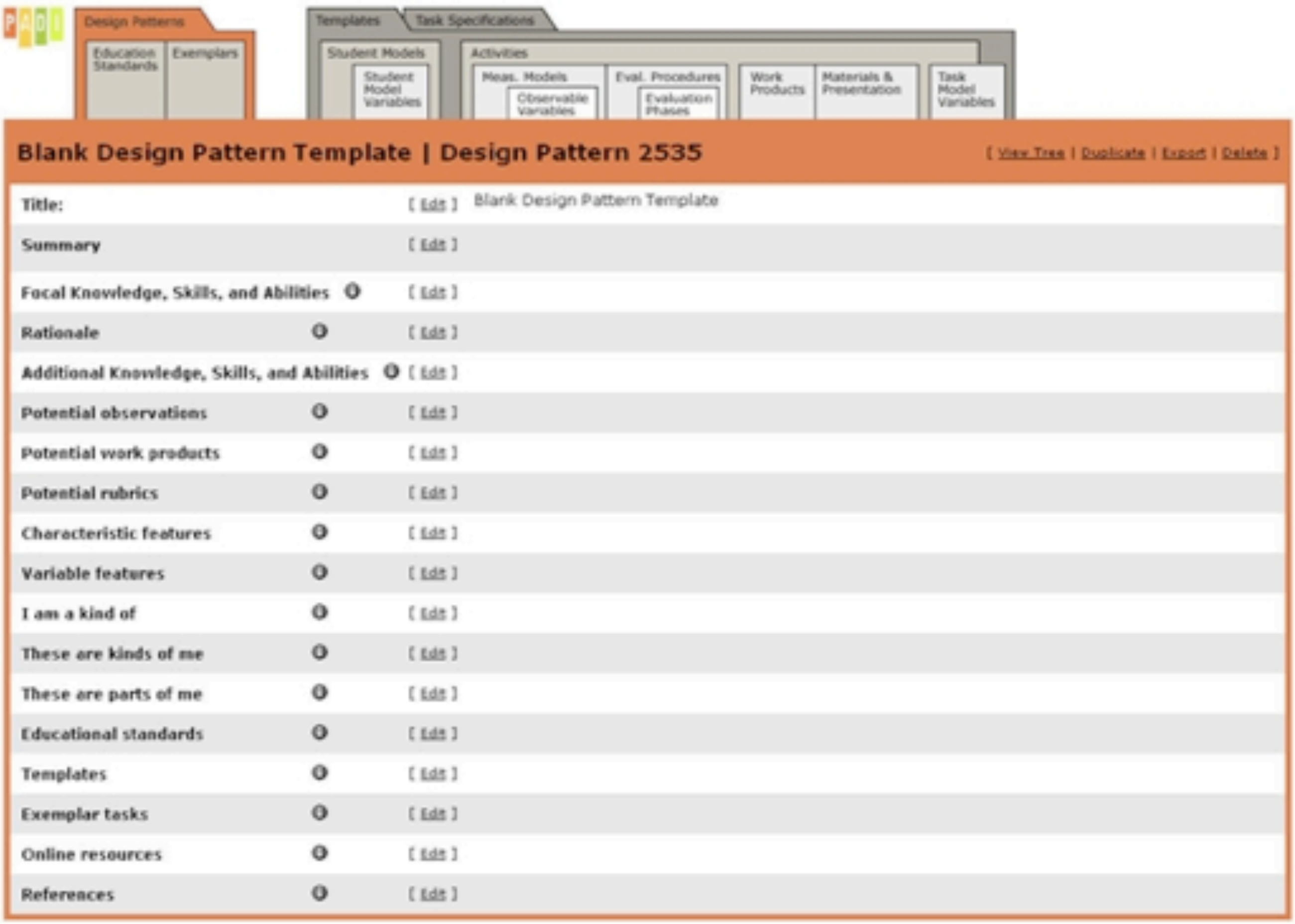

Figure 3. Design Pattern Template

Additional KSAs. These are the other knowledge/skill/abilities that may be required in a task (Mislevy et al., 2003). For tests of academic subjects, the abilities to “see” and “hear” are typically Additional KSAs. On the other hand, for assessment of sight and hearing, respectively, sight and hearing are likely to be defined as Focal KSAs. Notice that there are many disabilities that involve impairments of sight, hearing, or both (e.g., blind, low vision, color-blind, deaf, hard to hear, deaf-blind). Cognitive issues such as dyslexia, attention deficit, and executive processing limitations can also be addressed. Deficits in such Additional KSAs can cause unduly low scores among test takers with disabilities.

Characteristic Features. Characteristic Features of the assessment are the features that must be present in a situation in order to evoke the desired evidence about the Focal KSAs (Mislevy et al., 2003).

Variable Features. Variable Features are described as features that can be varied to shift the difficulty or focus of tasks (Mislevy et al., 2003). Variable Features have a particularly significant role with respect to test takers with disabilities and other sub-populations (e.g., speakers of minority language). Much of our attention will be on manipulating Variable

Features to reduce or eliminate demands for Additional KSAs in which there may be a deficit while making sure (to the extent possible) that demands for Focal KSAs have not been changed.

Method

In this study, ECD and UDL were applied to a subset of 20 statewide science items that will be delivered online. As part of the redesign process, design patterns were completed that represent the assessment argument underlying the item. Specifically, the design pattern helps identify whether task requirements elicit proficiency on intended test constructs (Focal KSAs) or inadvertently contribute variance to student scores but are not relevant to the construct being measured (construct-irrelevant Additional KSAs). Based on this analysis, revisions were made to item designs to reduce the influence of construct-irrelevant Additional KSAs. In the following section, we describe how we implemented revisions to the Kansas Statewide Science practice assessment items.

Construct validity is the sine qua non of assessment properties; to what degree do the evidence and rationale for the data gathered in an assessment support the inferences or decisions that a user wants to make? In the literature on accommodated assessment, the question typically centers on whether a given alteration of a task "changes the construct" (*Standards for Educational and Psychological Testing*, AERA, APA, NCME, 1985. p. 78). Specifically, if an alteration changes the construct, then construct validity has been violated, and if the alternation does not change the construct, then construct validity has not been violated.

Yet for assessment designers and developers as well as some other audiences, there is often a need to reason more deeply about the relationships between construct validity and task design. We would argue that it is important to specify more carefully what knowledge and skills need to be assessed, and at what levels; the assessment designers need to determine the essence of the intended construct that is to be assessed and what knowledge and skills influence test performance but are not the intended construct. This cannot be determined simply by examining the tasks on a test, because all of the knowledge, skills, and abilities needed to do well on a test are jointly required. In a given testing application, some of these KSAs will be relevant for the inference at hand and others will not (Phillips, 1994); the target examinee population may vary on some of them and not on others. It can even be the case that a given alteration on a test will introduce extraneous score variation in one application, and thus reduce validity, but reduce extraneous variation in a different application of the same test, and increase validity there. It is only by knowing the purpose of a test and the intended examinee population that one can answer how a given change will impact the evidentiary value of data for the construct meant to be assessed. A series of decisions needs to be made in the course of developing a specific test for a specific purpose and testing population to reason through the question of whether a given alteration "changes the construct."

Infusing UDL into PADI Design Patterns

The project team reviewed relevant background information on ECD and UDL to determine the intersection between UDL principles and PADI Design Patterns. Based on this analysis, six of the original nine UDL categories derived from UDL Principles I, II, and III are now used to categorize types of construct-irrelevant Additional Knowledge, Skills, and Abilities (Additional KSAs) that are likely to influence student performance. Definitions of UDL categories are provided in Table 1.

The three original Guidelines for Principle I — Perceptual, Language and Symbols, and Comprehension — have been included in the PADI design system. Comprehension was changed to Cognitive to reflect the types of supports that would be appropriate for assessments. The three guidelines for Principle II were modified for the PADI design system. Physical Action and Expressive Skills and Fluency were combined into one category — Skill and Fluency. The three guidelines for engagement have been condensed into one category — Affect— for the PADI design system. We used these categories to define potential construct-irrelevant barriers to assessment and Variable Features to consider when developing an assessment. These features can be embedded into assessments when appropriate in order to reduce barriers and gain a more accurate understanding of student learning. When using these categories it is essential to keep in mind the knowledge, skills, and abilities being assessed and the impact of any variable feature on the construct relevance of an item.

Variable Feature Categories Derived from UDL Principles
Appendix A summarizes the Variable Features associated with each UDL category and provides examples.

Perceptual. To be accurate for a diverse student population, assessments must present information in ways that are perceptible to all students. Perceptual barriers to assessment can be reduced by (a) providing the same information through different sensory modalities, and (b) providing information in a format that can be adjusted (e.g. text that can be enlarged, sounds that can be amplified). Multiple representations such as these can ensure that information is not only accessible to students with particular sensory and perceptual disabilities but also easier to access for many others. When the same information, for example, is presented in both speech and text, comprehension is enhanced for most students. Examples of variable features in this category include alternatives for visual information (e.g., providing text-to-speech) and options for representational format (e.g., enlarged text and graphics).

Language and Symbols. Students vary in their facility with language and symbols. Vocabulary that may clarify a test item for one student may be foreign to another. A graph that illustrates the relationship between two variables may be informative to one student but puzzling to another. An important assessment strategy is to ensure that alternative representations are provided not only for accessibility but to ensure information is clear and understandable to all students. Examples of variable features in this category include supports for vocabulary (e.g., definitions of construct-irrelevant terms) and supports for decoding graphs or charts (e.g., providing explanation of categories in a chart).

Cognitive. Decades of cognitive science research have shown that the ability to transform information into useable knowledge is an active, not a passive process. Constructing useable knowledge depends on active "information processing skills" such as selective attending, integrating new information with prior knowledge, and strategic categorizing. Individuals differ greatly in their skills in information processing and in their access to prior knowledge through which they can assimilate new information. Proper design and presentation of information can provide the cognitive ramps that are necessary to ensure that assessments accurately measure student knowledge. Examples of variable features in this category include supports for background knowledge (e.g., links to relevant background information) and supports for information processing (e.g., chunking information into smaller elements).

Skill and Fluency. Print format provides limited means of navigation or physical interaction (e.g., by turning pages with fingers, handwriting in spaces provided). Many interactive pieces of educational software similarly provide only limited means of navigation or interaction (e.g., via dexterously manipulating a joystick or keyboard). Navigating and interacting with these materials will raise barriers for some students. It is important to provide assessment materials that students can use easily. Furthermore, no medium of expression is equally suited for all students or for all kinds of communication. Alternative ways to respond to assessment items should be provided in order to ensure that the mode of response does not interfere with students' ability to demonstrate their true understanding. Additionally, students vary widely in their familiarity and fluency with different media. Media used for assessment, therefore, should include supports to scaffold and guide students who are using a less familiar medium so that they can express themselves competently. Examples of variable features in this category include supports for composition (e.g., sentence starters) and alternatives to writing (e.g., audiotaping responses).

Executive. Executive functions are at the highest level of the human capacity to act skillfully. We use these functions to overcome impulsive, short-term reactions to our environment and instead to set long-term goals, plan effective strategies for reaching those goals, monitor our progress, and modify strategies as needed. Executive functions have very limited capacity and are especially vulnerable to disability. In assessment situations, students' weaknesses in executive functions can hinder their ability to accurately demonstrate what they know. For these reasons, scaffolding for executive functions is important to consider in assessments. Examples of variable features in this category include supports for maintaining a goal (e.g., sentence starters) and supports for planning (e.g., graphic organizers).

Affect. Students differ significantly in what attracts their attention and engages their interest. Even the same student will differ over time and in different circumstances. Additionally, many tasks require not just initial engagement but sustained attention and effort. When motivated to do so, many students can regulate their attention and affect in order to sustain the effort and concentration that such learning will require. However, \students differ considerably in their ability to self-regulate in this way. Examples of variable features in this category include supports for intrinsic motivation (e.g., choice of item context) and supports for sustaining effort (e.g., explicit display of goal, such as number of items completed).

By incorporating Universal Design for Learning into assessment, all students' needs are taken into account. Providing options and supports will reduce potential construct0irrelevant barriers and lead to more accurate measurement for the range of learners who participate in a given assessment. We believe that the careful infusion of the six UDL categories designed for the PADI system will ultimately bring about a greater understanding of student learning.

The six categories within the Additional KSAs, along with the accompanying UDL Variable Task Features, guide designers to consider the diverse needs of all students. A similar extension of Potential Work Products that would support a range of ways of responding to tasks is being developed and linked with appropriate UDL-motivated KSAs. By infusing UDL into the PADI Design System, assessment designers are able to create flexible design patterns that will provide a more accurate measure of student learning.

Background on Design Patterns, Construct Validity, and Specific Assessment Contexts

A design pattern helps by laying out choices to be made as appropriate to specific testing applications. It is the specific test and context to which the property of construct validity applies, and the determination of which potential sources of variance among examinees' test scores would be construct relevant or construct irrelevant.

We have discussed above how important it is for tasks that are intended to assess a focal KSA to exhibit in some form the characteristic features denoted in the design pattern, and that by manipulating variable task features a test developer can increase, decrease, circumvent, or support particular Additional KSAs. Which Additional KSAs will be construct-relevant to require in a task in a given context depends on the test purpose and target population; that is, a test-for-purpose-with-population decision. The creator of the design pattern does not know what this decision will be, because it can be validly and appropriately different for different applications.

Table 1 below distinguishes what can be known at the time of creating a design pattern for any number of tests that in some way address the focal KSAs, and what must be determined at the time of specifying the application to a particular test. Note that the terms "Focal KSA" and "Additional KSA" describe KSAs in the *design space* while "construct relevant" and "construct irrelevant" describe KSAs in the *application space*.

Table 1. Design Space and Application Space Descriptors

	Application Space Descriptors (for thinking about KSAs for a particular test and its purpose and the intended examinee population)	
Design Space Descriptors (for thinking about KSAs in the design pattern stage)	Construct Relevant	Construct Irrelevant
Focal KSA. A design pattern is meant to support designing tasks and assessments that assess the Focal KSAs.	(1) Focal KSAs from the design pattern, at the right level and focus for the application.	(2) Focal KSAs, but too hard, too easy, or off focus for the intended application.
Additional KSA. Additional KSAs may be required at the designer's discretion	(3) The designer deems certain additional KSAs are appropriately part of the intended construct to assess.	(4) Additional KSAs that are required to apprehend, interact with, or respond to an implemented task, yet are not part of the intended construct to assess.

Both Cell 1 and Cell 2 specify what a particular test application needs to require for KSAs that are listed in the Focal KSA attribute of a design pattern. A test needs to have some requirement for focal KSAs (and perhaps some Additional KSAs as well) in order to be valid. Creating tasks that elicit these KSAs will contribute construct relevant variance in examinees' scores as long as it is done correctly.

Cell 1 addresses an implemented test application's requirement for the Focal KSAs listed in the design pattern, in a task meant to assess the capabilities the design pattern is meant to support, at a level that suits the application's intended use and examinee population. This is the essence of construct relevant variance in a test: Having the intended capability makes it more likely an examinee will perform well, and lacking it makes it more likely that he or she will not perform as well.

Cell 2 concerns requirements for the Focal KSAs described in the design pattern, but, in flawed test construction, the demand for the KSAs is not at the right level. For example, the word list for an in-class "spelling bee" for a second grade class might contain words that are much too hard for the students. The KSA of spelling English words is appropriate, but it has not been implemented appropriately for the intended use.

Cells 3 and 4 concern a particular test application's requirement for KSAs that are listed in the Additional KSA attribute of a design pattern. These demands may or may not contribute to construct relevant variance in that application, depending on the purpose and examinee population.

In Cell 3, the designer deems certain additional KSAs are appropriately part of the intended construct to assess. For example, it may be decided that working memory capability needed to spell words without writing them along the way is appropriate for an in-class "spelling bee", because it is intended to give the students feedback on how well they would do in the upcoming "spelling bee" competition that does not allow writing while spelling. More generally, prerequisite knowledge is often considered "fair game" in assessing school skills.

On a test of standards, at a given grade, including requirements for knowledge from standards at earlier grades are often considered appropriate and construct-relevant reasons for poor performance, and are therefore not scaffolded (i.e., the variable task feature "scaffolding" has been set to none for these Additional KSAs).

Cell 4 concerns Additional KSAs that are required to apprehend, interact with, or respond to an implemented task, but are not part of the intended construct to assess. For example, the standard 'spelling bee" requires a spoken response, and the KSA of speaking is almost certainly not of the essence of the capability at issue. It is a potential cause of poor performance. Allowing for typed, written, or pointed-to spelling of words as a task feature is a UDL approach to mitigating this problem. In general, requirements in a task for physical and cognitive KSAs that are not construct relevant can lead to poor performance, and mask the KSAs that are the intent of assessment (Focal KSAs, plus Additional KSAs that are construct relevant in the application at hand). They are thus potentially sources of construct irrelevant variation.

Note that the Additional KSA of being able to say letters aloud—i.e., to produce a work product in the form of a spoken sequence of letters—is not universally construct relevant or construct irrelevant in and of itself, but only in light of the purpose of a given test application. The design pattern cannot provide the answer, but it can alert the test developer to the question, and offer suggests for UDL and accommodation strategies when the Additional KSA is deemed construct irrelevant for the application and there are examinees in the test population who may not have the Additional KSA at the required levels.

The phrase "potentially construct irrelevant sources of variation" highlights the role of the intended examinee population in determining whether a requirement for a construct-irrelevant Additional KSA contributes to invalid inferences in a given application. Being able to speak letters in a "spelling bee" is a construct irrelevant requirement, but if it is known a priori that everyone in the class is able to spell works aloud, this will not be a source of poor performance for this population. But it might be for a different class that has a student who experiences difficulty responding in this manner. An alternative way of responding in that class, perhaps used only by that student, would be necessary in order to remove a construct irrelevant source of variance in the second class.

Examples of Item Variants

During the first three years of the "Principled Science Assessment Designs for Students with Disabilities" project, we have developed a great deal of machinery for infusing the principles of evidence-centered design and UDL into assessments. This includes developing a conceptual framework for the integration of UDL and ECD, extending the PADI design pattern tool to build in CAST's UDL guidelines, co-designing thirteen enhanced design patterns with four states, sharing the results at the 2009 annual meeting of the American Educational Research Association (AERA), and setting up arrangements to gather student data on alternate forms of items in Kansas and Nevada. The empirical phase of the work will create a set of twenty items with multiple variations, with each variation crafted to illustrate UDL principles to reduce or remove requirements for construct irrelevant Additional KSAs for targeted subpopulations of students.

We will be able to administer these variant forms using Kansas's web-based authoring and delivery system. This delivery system is currently used to administer the Kansas accountability tests to more than 80% of the schools in Kansas. It provides a flexible authoring system that will allow us to implement variant forms of items with little or no additional programming so that we can manipulate presentation material, response modes, supports, and UDL features. This system in fact brings a set of UDL features of its own that can be turned on or off to further create task variants.

Results

As part of this project, sixteen design patterns were created in collaboration with four partner state departments of education, including Kentucky, Kansas, Nevada and South Carolina. We present one of these design patterns, see Appendix B, as well as an original and revised item based on that design patterns. We discuss the rationale for the changes made to item directives, graphics, and distracters as well as the role of the exemplar design pattern in supporting the revision process.

The tasks that have initially provided the source material for item variants are released items from the Kansas assessment and formative assessment tasks, all of which are already implemented in the Kansas statewide delivery system. What follows is a discussion of an item that has been revised to take into account the ECD and UDL framework. We provide explanations of how the differences in features between the original and revised item reflect considerations of UDL and ECD assessment arguments, using the conceptual machinery we have developed in the project. Figures 4 and 5 present the original and revised items.

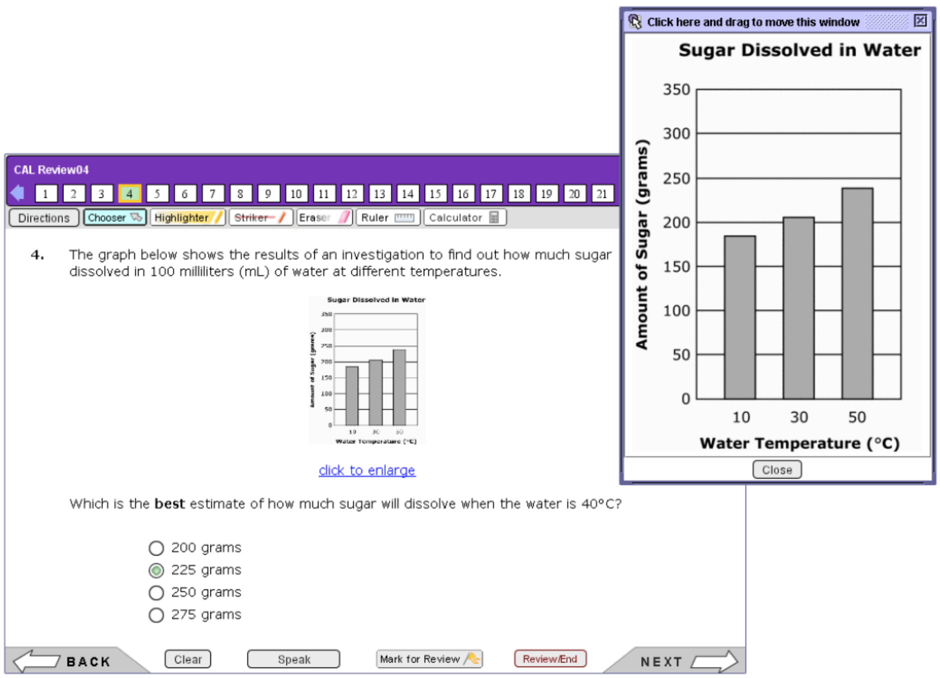

Figure 4. Original Version of Example Item

In its original form, this task (See Figure 5) presents a bar-chart of the number of grams of sugar that dissolved in 100ml of water at 10, 30, and 50 degrees Celsius, the result of an investigation. The chart indicates that at 10oC, approximately 175 grams of sugar dissolved, at 30oC approximately 205 grams of sugar dissolved, and at 50oC approximately 245 grams of sugar dissolved. The prompt asks, based on the graph, "which is the best estimate of how much sugar will dissolve when the water is 40oC." The multiple-choice options are "200 grams," "225 grams," "250 grams," and "275 grams."

Our analysis of the knowledge, skills, and abilities demanded by this task included the following:

- seeing and reading the text
- knowing the meaning of "mL"
- knowing the meaning of "oC "
- seeing the chart (a visual ability)

- "reading" the chart (knowledge and skill requirements to understand the
- relationship between temperature and amount of sugar dissolved, match up the units in the responses with segments on the X and Y-axes, and relate the verbal descriptions in the options with the bars on the chart)
- carrying through the steps relating the prompt to the chart (executive processing)
- identifying the bars that bound the range of values on which the interpolation will be based
- performing the appropriate mathematics estimation process within the range defined by the prompt
- identifying the response option that is closest to the mathematical estimate and/or identifying which responses do not fall within the range defined by the prompt and data points charted

Once the KSAs are identified, identifying which KSAs are both focal and construct relevant will determine which revisions will minimize demands of the task that are not those targeted and which revisions alter the targeted demands of the tasks.

Determining what is a construct relevant KSA of an existing item or task is not always clear cut. For example, in this item, reading the chart is a first step in the interpretation of data, however, the original designer may not have intended the item to test that particular KSA even though the presentation format makes it impossible to answer the questions without the application of that particular skill. If the designer did not intend this KSA to be construct relevant, any of multiple types of representations of the data could be presented to students from which they could perform the same interpolation. In this case, representation format would be considered an Additional KSA in the design pattern and would alert task developers to determine the extent of demand for this KSA to require in the task, if any. If the determination was that the KSA was construct relevant for the assessment application at hand, then one should maintain this level across revised versions of the tasks that might be administered to different students so that the same mix of construct-relevant KSAs was maintained. In the design pattern, the chart would then be a Characteristic Task Feature. In the case of this item, we maintained the chart representation based on our conversations with our state partners that confirmed their expectation that students be able to interpret data in bar charts in particular, as well as other data representation formats.

However, as the image provided with the item presents essential information, providing a way for students to enlarge it so that they can clearly see this information again may decrease the likelihood that some students would score incorrectly due to a factor that is not construct-relevant. Since this is a science task, we are probably safe in assuming that "seeing" the graph is not construct relevant, and it is appropriate to reduce or eliminate demands for this ability. One variable task feature we varied in the revised version of this task, therefore, was providing a larger version of the graph. This revision reduces demands on visual acuity, and makes the task more accessible to students with limited vision. Even the revised version, we may note, has some demand for visual acuity, which would be an alternative explanation for poor performance by a student with little or no visual capabilities. We provided machine-spoken descriptions (via a text-to-speech feature) to increase the accessibility to a still wider range of students. Those students who would were identified as requiring text-to-speech or other read-aloud accommodations during their typical school testing were provided this feature in our study.

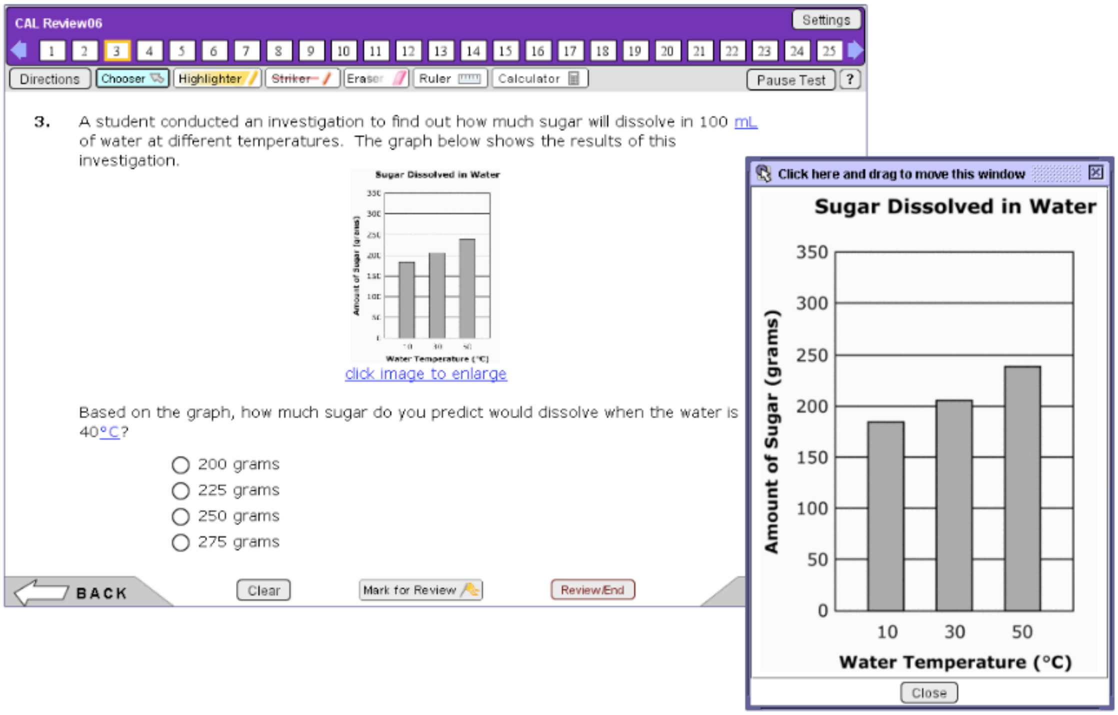

Figure 5. Revised Version of Example Item

Another consideration for test items is vocabulary. As illustrated in this example, terms and abbreviations such as mL or ^{o}C may not be construct relevant, and, if so, their definitions could be provided. If vocabulary terms are determined to be construct relevant KSAs, then, of course, this support would not be provided. In this case, interpreting the chart does not rely on any specific unit of measure and was therefore considered construct irrelevant. Embedded definitions were provided for students using a roll-over feature. If students hover their mouse over an underlined term, a definition appears. The UDL framework provided a range of possible variable features that would minimize the impact of the vocabulary KSA on student performance, including a glossary. In the testing situation we intended to provide to students, the additional demands of using a static glossary were deemed to be onerous for students with reading difficulties. As a rule, revisions prompted by the UDL principles that reduce construct irrelevant sources of cognitive load are usually worth removing if possible. The roll-over feature, although not a part of the original testing environment, was considered a common technological feature that students would be comfortable using.

The last four KSAs identified in analysis of this task describe complex cognitive processes required to complete this task. Many of these KSAs require executive processing as well as continued cognitive engagement with the task information, two hallmark challenges for students with learning disabilities and other high-incidence cognitive disabilities. Relying again on the UDL framework, revisions that supported students' executive functioning in the form of managing information and resources were identified and implemented to reduce the cognitive demands of the task that are implicit in the application of these KSAs. For example, modifying the physical layout of an item to clearly demarcate item context from the prompt (the questions students are supposed to answer) and using this layout consistently across the entire test can support students' perception and interpretation of the task presentation and language. In revising this task, we monitored that the layout matched the layout being used across the whole assessment.

Modifications to the language or wording of the task context and prompt can further support students' engagement and executive processing. The UDL framework identifies motivational elements as supports for student engagement. Providing a task context that situates the problem being posed that is familiar to students can increase intrinsic motivation. To provide a context that students might recall from their own experience, we situated this task in a typical classroom investigation and asked students to predict how much sugar would dissolve in water at 40 degrees. The interpolation task, then, has potentially more meaning for students than just a mathematical estimation. We reinforced this context by using active voice in the wording of the task to draw students' attention to the use and origination of the data.

In each of the above, the task revision results in reducing a requirement for a construct irrelevant KSA such the student's lack of those abilities should not hinder him or her from demonstrating ability in the targeted proficiency (which, in this case has one KSA). Essentially, revisions to the task ensures that the student can satisfy requirements for all non-construct relevant KSAs, thereby enabling their ability in the construct relevant KSAs to be expressed. Of course, for any number of reasons, the student may have a lack in the construct relevant KSA and therefore may perform poorly. However, the assessment will still be valid, i.e., validly low. And with assessments that are more valid for students with disabilities, we will more likely be able to address the range of causes for low academic achievement in many students with disabilities.

Significance

This application of UDL principles to the revision of statewide assessment tasks systematically documents the integration of UDL and ECD frameworks to enhance construct validity. Our findings speak in particular to those interested in building assessment design arguments that address the issue of student diversity. In the final year of the project, design strategies will be identified that proved especially helpful in the improvement of items for all students and, in particular, for students with high incidence disabilities. These design strategies will be articulated in a set of guidelines that state departments of education can apply in developing items for their statewide science assessments. While these strategies were developed in the context of middle school science assessments, we believe they will be applicable to most subject areas and across students of different ages. Likewise, even though this project focused on the refinement of items in "general education" science assessments, the variable feature categories based on UDL principles include features that are appropriate for students with low incidence disabilities.

References

Breese, J. S., Goldman, R. P., & Wellman, M. P. (1994). Introduction to the special section on knowledge-based construction of probabilistic and decision models. *IEEE Transactions on Systems, Man, and Cybernetics, 24*, 1577-1579.

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, *52*, 281-302.

Dolan, R. P., Rose, D. H., Burling, K., Harms, M., & Way, D. (April, 2007). The Universal Design for Computer-Based Testing Framework: A Structure for Developing Guidelines for Constructing Innovative Computer-Administered Tests. Paper presented at the National Council on Measurement in Education Annual Meeting, Chicago, IL.

Gamma, E., Helm, R., Johnson, R., & Vlissides, J. (1994). *Design patterns*. Reading, MA: Addison-Wesley.

Hansen, E. G., & Mislevy, R. J. (2008). *Design patterns for improving accessibility for test takers with disabilities (RR-08-49).* Princeton, New Jersey: ETS Research Report Series.

Kane, M. (1992). An argument-based approach to validation. *Psychological Bulletin*, *112*, 527-535.

Li, M., & Shavelson, R. J. (2001). (April 12, 2001). *Examining the links between science achievement and assessment.* Paper presented at the annual meeting of the American Educational Research Association, Seattle.

Mastropieri, M.A., & Scruggs, T.E. (1992). Science for students with disabilities. Review of Educational Research, 62, 377-411.

Mastropieri, MA. & Scruggs, T.E. 1995. Teaching Science to Students with Disabilities in General Education Settings. Teaching Exceptional Children, 27(4). pp. 10-13.

McCleery, J.A., and G.A. Tindal. 1999. Teaching the scientific method to at-risk students and LD students through concept anchoring and explicit instruction. *Remedial & Special Education* 20(1): 7-18.

Messick, S. (1989). Validity. In R.L. Linn (Ed.), *Educational Measurement* (3rd ed., pp. 13-103). New York: Macmillan.

Messick, S. (1994). The interplay of evidence and consequences in the validation of performance assessments. *Educational Researcher*, *32*(2), 13-23.

Meyer, A., & Rose, D. (2006). Preface. In D. Rose, & A. Meyer (Eds.), *A practical reader in universal design for learning.* (pp. vii-xi). Harvard Education Press, Cambridge, MA

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

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

Mislevy, R., & Riconscente, M. (2005). *Evidence-centered assessment design: Layers, structures, and terminology (PADI Technical Report 9).* Menlo Park, CA: SRI International.

Mislevy, R., Hamel, L., Fried, R., G., Gaffney, T., Haertel, G., Hafter, A., Murphy, R., Quellmalz, E., Rosenquist, A., Schank, P., Draney, K., Kennedy, C., Long, K., Wilson, M., Chudowsky, N., Morrison, A., Pena, P., Songer, N., Wenk, A. (2003). *Design patterns for assessing science inquiry* (PADI Technical Report 1). Menlo Park, CA: SRI International.

National Research Council (2004) Keeping score for all: The effects of inclusion and accommodation policies on large-scale educational assessment In J. . A. Koenig & L. F. Bachman (Eds). Committee on Participation of English Language Learners and Students with Disabilities in NAEP and Other Large-Scale Assessment. Washington, DC: National Academy of Sciences.

Norman, K.I., D. Caseau, and G.P. Stefanich. 1998. Teaching students with disabilities in inclusive classrooms: Survey results. Science Education 82: 127-146.

Phillips, S. E. (1994). High-stakes testing accommodations: Validitiy Versus Disabled Rights. Applied Measurement in Education, 7(2), 93-120.

Rose, D. H., & Meyer, A. (2002). Teaching every student in the Digital Age: Universal Design for Learning. Alexandria, VA: ASCD.

Rose, D.H., Meyer, A., & Hitchcock, C. (2005). The universally designed classroom: Accessible curriculum and digital technologies. Cambridge, MA: Harvard Education Press.

Standards for Educational and Psychological Testing, AERA, APA, NCME, 1985. p. 78.

Tillers, P. and Schum, D. A. (1991) A Theory of Preliminary Fact Investigation. University of California Davis Law Review 24, pp.931 - 1012.

Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). Woodcock-Johnson III. Itasca, IL: Riverside Publishing.

Appendix A: Variable Features by UDL Category with Examples

Perceptual Features

1. Representational Format
    - Flexible size of text and images
    - Flexible amplitude of speech or sound
    - Adjustable contrast
    - Flexible colors
    - Flexible layout
2. Auditory Information
    - Text equivalents (e.g. captions, automated speech to text)
    - Visual graphics or outlines
    - Virtual manipulatives, video animation
    - Verbal descriptions
    - Tactile graphics, objects
3. Visual Information
    - Spoken equivalents for text and images
    - Automatic text to speech
    - Tactile graphics
    - Braille

Language and Symbols

1. Supports for Vocabulary and Symbols
    - Pre-taught vocabulary and symbols
    - Embedded support for key terms (e.g. technical glossary, hyperlinks/ footnotes to definitions, illustrations, background knowledge)
    - Embedded support for non-technical terms (e.g. non-technical glossary, hyperlinks/ footnotes to definitions, illustrations, background knowledge)
    - Embedded alternatives for unfamiliar references (e.g. domain specific notation, jargon, figurative language, etc.)
2. Supports for Syntactic Skills and Underlying Structure
    - Alternate syntactic levels (simplified text)
    - Grammar aids
    - Highlighted syntactical elements (e.g. subjects, predicates, noun-verb agreement, adjectives, phrase structure, etc.)
    - Highlight structural relations or make them more explicit
3. Supports for English Language
    - All key information in the dominant language (e.g. English) is also available in prevalent first languages (e.g. Spanish) for second language learners and in ASL for students who are deaf
    - Key vocabulary words have links to both dominant and non-dominant definitions

and pronunciations

- Domain-specific vocabulary (e.g. "matter" in science) is translated for both special and common meanings
- Electronic translation tools, multi-lingual glossaries

4. Supports for Decoding and Fluency
    - Digital text with automatic text to speech
    - Digital Braille with automatic Braille to speech

Cognitive Features

1. Supports for Background knowledge
    - Advanced organizers, pre-teaching, relevant analogies and examples
    - Links to prior knowledge (e.g. hyperlinks to multimedia, concrete objects in students' environments)
    - Provision of an example
2. Supports for Critical features, Big Ideas, and Relationships
    - Concept maps, graphic organizers, outlines
    - Highlight features in text, diagrams, graphics, and illustrations
    - Reducing the field of competing information or distractions, masking
    - Using multiple examples and non-examples to emphasize critical concepts
3. Options that Guide Information Processing
    - Explicit prompts for each step in a sequential process
    - Interactive models that guide exploration and inspection
    - Graduated scaffolds that support information processing strategies
    - Multiple entry points and optional pathways through content
    - Chunking information into smaller elements, progressive release of information, sequential highlighting
    - Discrete question (s) or scenario-based text presentation
    - Complexity of the scientific investigation presented in the scenario
    - Cognitive complexity (Webb's Depth of Knowledge Levels)
    - If selected response, distracters based on misconceptions/typical errors vs. non-misconceptions
4. Supports for Memory and Transfer
    - Checklists, organizers, sticky notes, electronic reminders
    - Prompts for using mnemonic strategies and devices
    - Templates, graphic organizers, concept maps to support note-taking
    - Scaffolding that connects new information to prior knowledge
    - Embedding new ideas in familiar ideas and contexts, use of analogy, metaphor, example

Skill and Fluency

1. Supports for Manipulations
    - Virtual manipulatives, Snap-to constraints
    - Nonskid mats, Larger objects
2. Supports for Navigation

- Alternatives for physically interacting with materials: by hand, by voice, by single switch, by keyboard, by joystick, by adapted keyboard

3. Alternatives to Writing
    - Voice recognition, Audio taping, Dictation, Video, Illustration
4. Supports for Composition
    - Keyboarding and alternative keyboards, Onscreen keyboard,
    - Wider lines, Larger paper, Pencil grips
    - Drawing tools - with shapes, lines, etc.
    - Blank tables, charts, graph paper
    - Spellcheckers, calculators, sentence starters, word prediction, dictation (voice recognition or scribe), symbol-to-text, sentence strips

Executive Features

1. Support for Goal and Expectation Setting
    - Prompts and scaffolds to estimate effort, resources, and difficulty
    - Animated agents that model the process and product of goal-setting
    - Guides and checklists for scaffolding goal-setting
2. Supports for Goal Maintenance and Adjustment
    - Maintain salience of objectives and goals (e.g. reminders, progress charts)
    - Adjust levels of challenge and support (e.g. adjustable leveling and embedded support, alternative levels of difficulty, alternative points of entry)
3. Supports for Planning and Sequencing
    - Embedded prompts to "stop and think" before acting
    - Checklists and project planning templates for setting up prioritization, schedules, and steps
    - Guides for breaking long-term objectives into reachable short-term objectives
4. Supports for Managing Information
    - Graphic organizers and templates for organizing information
    - Embedded prompts for categorizing and systematizing
    - Checklists and guides for note-taking
5. Supports for Working Memory
    - Note-taking, mnemonic aids
    - Locate items near relevant text
6. Supports for Monitoring Progress
    - Guided questions for self-monitoring

- Representations of progress (e.g. before and after photos, graphs and charts)
- Templates that guide self-reflection on quality and completeness
- Differentiated models of self-assessment strategies

Affect Features

1. Supports for Intrinsic Motivation (Challenge and/or Threat)
    - Offer individual choice
    - Enhance relevance, value, authenticity (e.g. contextualize to students' lives, provision of an example)
    - Options to vary level of novelty and risk (e.g. options in peer and adult support, alternatives to competition, alternatives to public display or performance, alternative consequences)
    - Options to vary sensory stimulation (e.g. shortened work periods, frequent breaks, noise buffers, optional headphones, alternative settings, presentation of fewer items at a time)
2. Supports for Sustaining Effort and Persistence
    - Maintain salience of goals (e.g. explicit display of goals, periodic reminders, replacement of long-term goals with short-term objectives, prompts for visualization)
    - Adjustable levels of challenge and support
    - Encourage collaboration and support
    - Communicate on-going, mastery-oriented feedback
3. Support for Self-regulation
    - Guide motivational goal-setting
    - Scaffold self-regulatory skills and strategies
    - Develop emotional self-assessment and reflection

Appendix B: Example Design Pattern

[NV] Interpreting Data in Tables, Charts, and Graphs - AERA 2011 | Design Pattern 2130

[| Permit | Delete | View: View (vertical)]

Title	[Edit]	[NV] Interpreting Data in Tables, Charts, and Graphs - AERA 2011
Overview	[Edit]	This Design Pattern describes key components of tasks that might be designed to measure students' ability to understand relationships among data as represented in canonical science and mathematical forms (i.e., tables, charts and graphs). Webb's Depth of Knowledge (DOK) framework is used throughout to scaffold design of items that tap this ability at each level of Webb's framework.
Focal Knowledge, Skills, and Abilities	[Edit]	FK1. Ability to identify data points in one or more representational forms. FK2. Ability to compare and /or contrast multiple representations and the data represented therein. FK3. Ability to extrapolate or interpolate data points from given data. FK4. Ability to describe simple mathematical relationships or trends among data. FK5. Ability to draw conclusions or make predictions based on data.
Rationale	[Edit]	R1. A key activity of science inquiry is working with data in many forms or representations. Students' ability to analyze relationships among data is integral to their participation in science inquiry and to represent and think critically about relationships among experimental variables, observed phenomena, etc.
Additional Knowledge, Skills, and Abilities	[Edit]	AK1. == The following Additional KSAs are prerequisite knowledge that can be required for tasks that address the Focal KSA. Whether they are to be supported or not (e.g., glossary, background facts, equation list) is a decision to be made -- either by the assessment design team, at the level of the testing program, or at the level of the individual task if that is appropriate in the testing program. == AK2. Awareness of different representational forms AK3. Knowledge of what data are AK4. Ability to identify dependent and independent variables AK5. Knowledge of mathematics AK6. Scientific content knowledge AK7. == The following Additional KSAs are generally construct-irrelevant knowledge, skills, or other attributes that may be involved in tasks generated under this design pattern. The task author can consider offering supports, presenting material, or getting work products that reduce or avoid requirements for these Additional KSAs, either through accommodated forms of a task or UDL principles. Many of these Additional KSAs are linked to Variable Task Features or Potential Work Products for suggestions on how to do this. == AK8. Perceptual . vision . hearing . touch AK9. Language and symbols . vocabulary and symbols . syntax and underlying structure . English-language proficiency . decoding text or math notation . decoding charts, graphs, or images AK10. Cognitive . background knowledge . concepts and categories . information processing strategies . memory and transfer AK11. Skill and fluency

		. dexterity, strength, and mobility . navigation and object manipulation . automaticity (e.g., calculations, writing) . familiarity with media . facility with tools AK12. Executive (problem solving) . goal and expectation setting . goal maintenance and adjustment . planning and sequencing steps in a process . managing information and resources . working memory . monitoring progress AK13. Affective . intrinsic, task-specific motivation (challenge and/or threat, interest) . sustaining effort and persistence . coping skills and frustration management
Potential observations	[Edit]	Po1. Correct identification of the location of a data point in chart or graph OR the accurate identification of a value to complete a data table. Po2. Identification of representational forms of data that communicate the same mathematical relationships among data (or trends in data). Po3. Accuracy of conclusions drawn from data that are intended to inform predictions. Po4. Appropriateness of inferences drawn from data tables, charts, and graphs.
Potential work products	[Edit]	Pw1. Selection of inference or prediction (selected response) Pw2. Written interpretation of data from one or more representational forms Pw3. Written prediction based on interpretation of data
Potential rubrics	[Edit]	Pr1. Key for selected response items Pr2. Partial credit rubric for scoring of written responses
Characteristic features	[Edit]	Cf1. The presentation contains numeric data Cf2. The presentation includes at least one representational form Cf3. The presented data are in a scientific context
Variable features	[Edit]	Vf1. Number of representations Vf2. Type(s) of representations Vf3. Amount of data Vf4. Complexity of representational form(s) Vf5. Number of variables represented in the table, graph, or chart Vf6. Provision of an example Vf7. Amount of content knowledge required Vf8. Presence of color(s) in table, graph, or chart Vf9. Data source (student collected vs. provided) Vf10. Perceptual Features (1): Representational Format - Flexible size of text and images - Flexible amplitude of speech or sound - Adjustable contrast - Flexible colors - Flexible layout Vf11. Perceptual Features (2): Auditory Information - Text equivalents (e.g. captions, automated speech to text) - Visual graphics or outlines - Virtual manipulatives, video animation - Verbal descriptions - Tactile graphics, objects Vf12. Perceptual Features (3): Visual Information

- Spoken equivalents for text and images
- Automatic text to speech
- Tactile graphics
- Braille

Vf13. Language and Symbols (1): Supports for Vocabulary and Symbols
- Pre-taught vocabulary and symbols
- Embedded support for key terms (e.g. technical glossary, hyperlinks/ footnotes to definitions, illustrations, background knowledge)
- Embedded support for non-technical terms (e.g. non-technical glossary, hyperlinks/ footnotes to definitions, illustrations, background knowledge)
- Embedded alternatives for unfamiliar references (e.g. domain specific notation, jargon, figurative language, etc.)

Vf14. Language and Symbols (2): Supports for Syntactic Skills and Underlying Structure
- Alternate syntactic levels (simplified text)
- Grammar aids
- Highlighted syntactical elements (e.g. subjects, predicates, noun-verb agreement, adjectives, phrase structure, etc.)
- Highlight structural relations or make them more explicit

Vf15. Language and Symbols (3): Supports for English Language
- All key information in the dominant language (e.g. English) is also available in prevalent first languages (e.g. Spanish) for second language learners and in ASL for students who are deaf
- Key vocabulary words have links to both dominant and non-dominant definitions and pronunciations
- Domain-specific vocabulary (e.g. "matter" in science) is translated for both special and common meanings
- Electronic translation tools, multi-lingual glossaries

Vf16. Language and Symbols (4): Supports for Decoding and Fluency
- Digital text with automatic text to speech
- Digital Braille with automatic Braille to speech

Vf17. Cognitive Features (1): Supports for Background knowledge
- Advanced organizers, pre-teaching, relevant analogies and examples
- Links to prior knowledge (e.g. hyperlinks to multimedia, concrete objects in students' environments)
- Provision of an example

Vf18. Cognitive Features (2): Supports for Critical features, Big Ideas, and Relationships
- Concept maps, graphic organizers, outlines
- Highlight features in text, diagrams, graphics, and illustrations
- Reducing the field of competing information or distractions, masking
- Using multiple examples and non-examples to emphasize critical concepts

Vf19. Cognitive Features (3): Options that Guide Information Processing
- Explicit prompts for each step in a sequential process
- Interactive models that guide exploration and inspection
- Graduated scaffolds that support information processing strategies
- Multiple entry points and optional pathways through content
- Chunking information into smaller elements, progressive release of information, sequential highlighting
- Discrete question(s) or scenario-based text presentation
- Complexity of the scientific investigation presented in the scenario
- Cognitive complexity (Webb's Depth of Knowledge Levels)
- If selected response, distractors based on misconceptions/typical errors vs. non-misconceptions

Vf20. Cognitive Features (4): Supports for Memory and Transfer
- Checklists, organizers, sticky notes, electronic reminders
- Prompts for using mnemonic strategies and devices
- Templates, graphic organizers, concept maps to support note-taking
- Scaffolding that connects new information to prior knowledge
- Embedding new ideas in familiar ideas and contexts, use of analogy, metaphor, example

Vf21. Skill and Fluency (1): Supports for Manipulations
- Virtual manipulatives, Snap-to constraints
- Nonstick mats, Larger objects

Vf22. Skill and Fluency (2): Supports for Navigation

- Alternatives for physically interacting with materials: by hand, by voice, by single switch, by keyboard, by joystick, by adapted keyboard

Vf23. Skill and Fluency (3): Alternatives to Writing
- Voice recognition, Audio taping, Dictation, Video, Illustration

Vf24. Skill and Fluency (4): Supports for Composition
- Keyboarding and alternative keyboards, Onscreen keyboard,
- Wider lines, Larger paper, Pencil grips
- Drawing tools - with shapes, lines, etc.
- Blank tables, charts, graph paper
- Spellcheckers, calculators, sentence starters, word prediction, dictation (voice recognition or scribe), symbol-to-text, sentence strips

Vf25. Executive Features (1): Support for Goal and Expectation Setting
- Prompts and scaffolds to estimate effort, resources, and difficulty
- Animated agents that model the process and product of goal-setting
- Guides and checklists for scaffolding goal-setting

Vf26. Executive Features (2): Supports for Goal Maintenance and Adjustment
- Maintain salience of objectives and goals (e.g. reminders, progress charts)
- Adjust levels of challenge and support (e.g. adjustable leveling and embedded support, alternative levels of difficulty, alternative points of entry)

Vf27. Executive Features (3): Supports for Planning and Sequencing
- Embedded prompts to "stop and think" before acting
- Checklists and project planning templates for setting up prioritization, schedules, and steps
- Guides for breaking long-term objectives into reachable short-term objectives

Vf28. Executive Features (4): Supports for Managing Information
- Graphic organizers and templates for organizing information
- Embedded prompts for categorizing and systematizing
- Checklists and guides for note-taking

Vf29. Executive Features (5): Supports for Working Memory
- Note-taking, Mnemonic aids
- Locate items near relevant text

Vf30. Executive Features (6): Supports for Monitoring Progress
- Guided questions for self-monitoring
- Representations of progress (e.g. before and after photos, graphs and charts)
- Templates that guide self-reflection on quality and completeness
- Differentiated models of self-assessment strategies

Vf31. Affect Features (1): Supports for Intrinsic Motivation (Challenge and/or Threat)
- Offer individual choice
- Enhance relevance, value, authenticity (e.g. contextualize to students' lives, provision of an example)
- Options to vary level of novelty and risk (e.g. options in peer and adult support, alternatives to competition, alternatives to public display or performance, alternative consequences)
- Options to vary sensory stimulation (e.g. shortened work periods, frequent breaks, noise buffers, optional headphones, alternative settings, presentation of fewer items at a time)

Vf32. Affect Features (2): Supports for Sustaining Effort and Persistence
- Maintain salience of goals (e.g. explicit display of goals, periodic reminders, replacement of long-term goals with short-term objectives, prompts for visualization)
- Adjustable levels of challenge and support
- Encourage collaboration and support
- Communicate on-going, mastery-oriented feedback

Vf33. Affect Features (3): Support for Self-regulation
- Guide motivational goal-setting
- Scaffold self-regulatory skills and strategies
- Develop emotional self-assessment and reflection

I am a kind of [Edit]

These are kinds of me [Edit]

These are [Edit]

parts of me		
Educational standards	[Edit]	NV (3) Inquiry Standard N.8.A.1. Students know how to identify and critically evaluate information in data, tables, and graphs
Templates	[Edit]	
Exemplar tasks	[Edit]	
Online resources	[Edit]	
References	[Edit]	

Tags [Add Tag]

(No tags entered.)

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]).

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

[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.

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.

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

---

[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.

– 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, Summary Task Templates, and Development Specifications and Exemplar Task Templates. 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. The products are described briefly below; the steps for developing the two major products, Design Patterns and Development Specifications and Exemplar Task Templates, are described in detail in the next two sections.

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] A detailed description of the process used to create Design Patterns is provided in the next section.

---

[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*

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.

Design Pattern Development Guidelines

The nine steps in the following pages describe the process to complete a *Design Pattern*. However, it is possible for the process to be more iterative than implied by these steps; that is, prior steps may be revisited and the *Design Pattern* refined accordingly to further specify attributes or make the assessment argument more explicit. The example described in the steps that follow was developed by the AAD-M project for the Number and Operations expectation: "**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.**"

Step 1: Create a Title and Summary

The title is a name for the *Design Pattern* that briefly describes the content or skills addressed in it. It is important to adopt a naming convention and to consistently use it. For the AAD-M project, the title

(Baxter, G., & Mislevy, R., 2005, Menlo Park, CA: SRI International); and Technical Report 8, *An 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.

was comprised of three elements: the content area subdomain, the label or code of the standard or "expectation" addressed, and the grade-level range. For example, the title "Number and Operations A3 (grades3–5)" was created from the expectation of the NCTM *Principles and Standards for School Mathematics* mentioned above. The content area subdomain was Number and Operations. The next element in the title was the code for the NCTM expectation being addressed, A3, in which "A" referred to the second standard in the Number and Operations subdomain, and the "3" referred to the third expectation within that first standard.[5] The final element of the title was the grade-level range, grades 3–5.

The summary provides more detail about the scope or breadth of knowledge and skills to be addressed in the *Design Pattern*. To operationalize this attribute, the AAD-M project used the verbatim wording of the expectation from the NCTM *Principles and Standards for School Mathematics*. For instance, the summary for Number and Operations A3 (grades3–5) was "**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,**" which is the exact wording of the NCTM expectation.

Note that although the AAD-M project chose to use the NCTM standards and expectations to guide the work, other standards, including the state's extended standards or the Common Core State Standards, could serve as the base for the ECD approach.

Step 2: Add Relevant Educational Standards

In the AAD-M project, prior to developing the *Design Pattern*, the co-design team engaged an expert in mathematics to create a crosswalk linking NCTM expectations to state extended standards. Extended standards from Utah, Idaho, and Florida related to the NCTM expectation were included in the *Design Pattern* as a reference to show what mathematics content and skills each state identified as essential for students with significant cognitive disabilities.

Step 3: Develop Rationale Statement

The Rationale identifies why the construct(s) identified in the summary are important to assess. Creating a rationale statement requires input from a math content expert, who can situate the mathematics constructs targeted by the *Design Pattern* within the broader domain of mathematics. For instance, the rationale statement for Number and Operations A3 (grades3–5) was "**Fractions represent a significant extension of children's knowledge about numbers. When children possess a sound understanding of fractions, they can use this knowledge to describe real world phenomena and apply it to problems involving measurement, probability, and statistics.**"

Step 4: Identify Focal Knowledge, Skills, and Abilities (Focal KSAs)

Standards are often written at a grain-size that is too large for assessment purposes. Focal KSAs reflect the standard when it is further unpacked into its essential, assessable elements. The content expert(s) on the co-design team draft Focal KSAs by reviewing the standard. In co-design meetings, Focal KSAs are discussed and further refined.

The focus and grain-size of the Focal KSAs need to be agreed upon by the co-design team. It is possible to generate multiple standards-based Focal KSAs, each of which only addresses one facet of a standard. For example, Number and Operations A3 (grades3–5), included six Focal KSAs each of which addresses a component of the expectation:

[5] The NCTM does not label specific standards and expectations within or across domains using the naming conventions described above (i.e., A3). This naming convention was a creation of the AAD-M project to distinguish among various standards and expectations within an NCTM subdomain (e.g., Number and Operations).

1. Ability to recognize a whole and divide it into or recognize equal parts (e.g., halves, thirds, or quarters)
2. Ability to identify fractions by the number of parts in the whole and in the fractional amount
3. Ability to identify a collection as a whole, and consider groups of objects in the collection as parts of the whole
4. Knowledge that fractions are numbers and identify points on the number line corresponding to particular fractions, between 0 and 1, and greater than 1
5. Knowledge that a division operation may not have a whole number result
6. Ability to solve problems involving fractions

It is critical to take the time with the co-design team to consider how the content or skills in the standard should be parsed because Focal KSAs will influence other attributes of the *Design Pattern*. Focal KSAs may be refined or deleted after their influence with respect to other attributes of the *Design Pattern* becomes more apparent. For example, in Step 4 Focal KSAs are “operationalized” when observations of student behaviors that are likely to provide evidence of each Focal KSA are specified. At this point the co-design team may realize that a Focal KSA is too vague or too complex to create these observations. If this occurs, the co-design team should revisit Step 3 and refine the Focal KSA. The co-design may also go back and add new Focal KSAs as they progress through the design steps.

Step 5: Develop Potential Observations and Potential Work Products

Potential Observations help to make each Focal KSA more concrete by describing the evidence (in the form of a specific student behavior) that indicates that a student has acquired the KSA. Potential Observations are phrased to describe the highest quality of student performance that demonstrates evidence of the Focal KSA. Qualifiers such as “accurate” and “correct” are used in all Potential Observation statements. Co-design teams also may find it helpful to generate specific examples for each Potential Observation (i.e., given a particular mathematics problem or context, describe the observed behavior). In constructing *Design Patterns* for the AAD-M project, the extended standards from each participating state also were considered when determining the range and qualities of behaviors that would likely be observed for students with significant cognitive disabilities. See Table 3 for examples of Potential Observations.

Potential Work Products are descriptions of the form of the information that can be gathered from students (e.g., written explanation or selection of a response). When possible, work products should be stated such that they do not reflect bias in how students express their response. Often, Potential Observations can be expressed in multiple ways (e.g., in speech or in writing). Thus, the Potential Work Product “Expression of a mathematical pattern” is preferable to “Student writes the mathematical pattern,” since not all students can write. However, in some cases, a Potential Work Product must be specific to a particular mode of expression for a Potential Observation. In these cases, additional Potential Observations and associated Potential Work Products should be specified that reflect alternate modes of expression. See Table 3 for examples of Potential Work Products.

A “horizontal view” of the *Design Pattern* is used during co-design meetings to make the connections among each Focal KSA and its associated Potential Observations and Potential Work Products explicit. Table 3 shows an excerpt of the “horizontal view” for Number and Operations A3 (grades3–5).

Table 3. "Horizontal View" of Excerpt from Number and Operations A3 (grades 3–5) Design Pattern Focal KSAs, Potential Observations and Potential Work Products

Focal KSAs	Potential Observations	Potential Work Products
Ability to recognize a whole and divide it into or recognize equal parts (e.g., halves, thirds, or quarters)	Student correctly divides an object into a specified number of equal parts Student correctly identifies a pictorial representation of a fraction Student correctly distinguishes a whole from fractions of a whole	Student worksheet with multiple pictorial representations of fractions A whole object divided into fractional pieces Selection of a whole orange and fractional pieces of it

Step 6. Develop Characteristic Features of Tasks

In reviewing the Focal KSAs, Potential Observations and Potential Work Products, the co-design team identifies the key features of tasks that will be developed using a particular *Design Pattern*. These Characteristic Features must apply to all tasks created from a *Design Pattern*. For example, one Characteristic Feature developed for the Number and Operations A3 (grades3–5) *Design Pattern* is "**All problems will involve the use of fractions.**" In addition, Characteristic Features can define ways to constrain tasks in relation to the content (e.g., limitations on which numbers should be used). Characteristic Features also can pertain to more general task features desired in tasks associated with a *Design Pattern*. These may include task features such as prompting for individual student responses (not group responses), allowing accommodations, and involving a test administrator who knows the student's comprehensive/response abilities.

Step 7. Identify Cognitive Background Knowledge Additional KSAs

Steps 2–6 make explicit relationships among the standard (or, in the case of the AAD-M project, the NCTM expectation), the Focal KSAs, student behaviors and work products that provide evidence of the Focal KSAs, and characteristic features of tasks to elicit the desired student behaviors. In Step 7 the co-design team describes the Additional KSAs that are not construct relevant but may be required for successful performance on tasks associated with a particular *Design Pattern*.

To determine the Cognitive Background Knowledge Additional KSAs, the co-design team must consider the prerequisite knowledge and skills that may be needed for each Focal KSA. For example, the Number and Operations A3 (grades3–5) *Design Pattern* includes the Focal KSA, **Ability to recognize a whole and divide it into or recognize equal parts (e.g., halves, thirds, or quarters)**. In order for students to be able to demonstrate this ability, the co-design team determined that students may need additional background KSAs, such as:

- Ability to count using whole numbers
- Ability to use a number line to model whole numbers and operations on them
- Knowledge that there is "space" on the number line between each whole number
- Ability to perform division operations (e.g., grouping)

Step 8. Create Cognitive Background Knowledge Variable Features of Tasks

In order to prevent Cognitive Background Knowledge Additional KSAs from impinging on a student's ability to demonstrate what they know about the Focal KSAs, the co-design team considers how these Additional KSAs may be supported. These supports are Cognitive Background Knowledge Variable Features. For example, for Number and Operations A3 (grades3–5), the following Cognitive Background Knowledge Variable Features were identified:

- Provision of a table, chart, or tactile reminder of the numbers to support understanding that numbers occur in a specified sequence
- Supports for use of a number line (supported, unsupported, degree of support) (e.g., provide a number line, model use of a number line, re-teaching use of a number line just prior to assessment)
- Supports for division skills (supported, unsupported, degree of support) (e.g., provide a calculator, use number line to provide visual representation of division, counters, counter mats)

Step 9. Review and Select UDL Additional KSAs and Variable Features

In the AAD-M project six categories of UDL were used: (1) Perceptual (Receptive), (2) Skill and Fluency (Expressive), (3) Language and Symbols, (4) Cognitive, (5) Executive, and (6) Affective. UDL Additional KSAs are nonconstruct relevant knowledge, skills, and abilities in these categories that may be required for successful performance on tasks associated with a *Design Pattern*. UDL Variable Features are used to support student abilities associated with perceiving task stimuli, expressing responses to tasks, comprehending linguistic components of tasks, information processing, executive functioning, and engagement. Unlike the Cognitive Background Knowledge Additional KSAs and Variable Features, which are created afresh for each Design Pattern, the UDL Additional KSAs and associated Variable Features have been standardized and are prepopulated in each Design Pattern (for a full list of these associations, see *Implementing Evidence-Centered Design to Develop Assessments for Students with Significant Cognitive Disabilities: Guidelines for Creating Design Patterns and Development Specifications and Exemplar Task Templates for Mathematics*). The co-design team is responsible for reviewing this standardized list and selecting those Additional KSAs and associated Variable Features that are most relevant for the task.

Development Specifications and Exemplar Task Template Development Guidelines

Once a *Design Pattern* has been reviewed and finalized, the creation of the *Development Specifications and Exemplar Task Template* can commence. In this section the methodology involved in this enterprise is described. Specifically, the following section provides guidelines and suggestions for the development of tasks.[6] Appendix B includes an example *Development Specifications and Exemplar Task Template* for Number and Operations A3 (grades 3–5).

Step 1. Pre-populate Section A of the Development Specifications and Exemplar Task Template

Step 1 involves pre-populating some of the attributes within the *Development Specifications and Exemplar Task Template* with information taken directly from the *Design Pattern*. The first three attributes (Title, Summary, and Rationale) come directly from the associated *Design Pattern*.

Grade-Level Standards are included to provide background information about what general education children should be able to do in light of specific standards or expectations and at certain grade levels. For example, the following Grade-Level Standards information was included for Number and Operations A3 (grades 3–5), which focuses on understanding fractions:

- Understand the structure of numbers and the relationships among numbers
- Explore a variety of models of fractions focused on familiar fractions: halves, thirds, fourths, fifths, sixths, eighths, and tenths

---

6 The methodology we describe here is a result of our experiences in developing the AAD-M project's task design and development specifications template. Although there may be minor variations among different co-design teams in their implementation, these are the general guidelines that were followed.

- Develop strategies for ordering and comparing fractions using benchmark fractions such as ½ and 1
- Use parallel number lines to show a unit fraction and its multiples

Whether this information is extracted from a standards document or generated for the project, the informed perspective of a math education or content expert is required. The co-design teams refer to this information as they build items that are accessible and appropriate for the 1% population, while considering alignment to grade-level expectations for the general education population.

Step 2. Pre-populate Section B of the Development Specifications and Exemplar Task Template

Step 2 involves pre-populating the attributes within Section B of the *Development Specifications and Exemplar Task Template* with information taken directly from the *Design Pattern*. In this step all Focal KSAs, Cognitive Background Knowledge Additional KSAs, Potential Observations, Potential Work Products, and Characteristic Features are copied from the associate *Design Pattern* into the *Development Specifications and Exemplar Task Template*.

Finally, the UDL Variable Features selected in the *Design Pattern* as most relevant for the task are copied into the Potential Variable Features section of the *Development Specifications and Exemplar Task Template.* These Variable Features will be "set" as part of the item development process to document precisely how task features are manipulated to influence item difficulty. For instance, in Number and Operations A3, grades 3–5, the Potential Variable Features included:

- Number of representations presented to the student
- Models (fraction circles, card board representations, other manipulatives) and pictures
- Types of representation (fractions or wholes)
- Presentation of fraction (verbal, symbolic)
- Size of the denominator (2, 3, or 4)
- DOK of the content (e.g., fractions used [halves, thirds, quarters, etc.])

Step 3. Review and/or Revise the Pre-populated Attributes in Section B

It is an important and necessary step to review the pre-populated components of the template, as well as reflect again on the extended standards aligned to the standard/expectation being addressed. This reflection provides a sense of how and if the participating states are currently assessing content related to the expectation and also helps to shed light on how the emphases placed on a particular expectation may vary by state.

From among the Focal KSAs, the co-design team will select the Focal KSA that will serve as the foundation for Items 1 and 2. The choice of the Focal KSA can depend on several factors:

- Alignment of the particular focal KSA to the intended emphases of the state's extended standards.
- Complexity of the KSA (e.g., number of steps involved, level of cognitive skill required, and whether this level is appropriate for the target population). During the ECD design pattern process, the expectation is deconstructed into a set of distinct focal KSAs. Some co-design teams may prefer to select more fine-grained or more comprehensive Focal KSAs.
- Clarity or relative simplicity of the intended KSA to be assessed.
- Feasibility for developing tasks that can be "worked down"[7] (Browder et al., 2007) to encourage content accessibility for a wider spectrum of the target population.

[7] Browder uses the phrase "work it down" to describe how to develop alternate assessments (AA) for students with significant cognitive disabilities that are linked to grade-level academic content standards. She suggests starting with

Once the Focal KSA is selected, the next step is to determine the Potential Observations and Potential Work Products that will be targeted for Items 1 and 2. Within the design pattern each Focal KSA is associated with one or more Potential Observations (i.e., which represent different ways of gathering evidence of the focal KSA) and one or more corresponding Potential Work Products. A decision must be made about which Potential Observation and Potential Work Product will be used to provide evidence about the chosen Focal KSA. Although it is usually the case that the Potential Observation for Items 1 and 2 is selected from the list of Potential Observations detailed in the *Design Pattern* for the chosen Focal KSA, the co-design team may identify others at this point. If there is not a Potential Observation and/or Potential Work Product within the list from the *Design Pattern*, then a more appropriate Potential Observation and/or Potential Work Product that embodies the Focal KSA can be suggested, selected, and subsequently added to the *Development Specifications and Exemplar Task Template*. The selection of the Potential Observation and Potential Work Product may depend on several factors including:

- Cognitive complexity of the observed behavior for the target population (e.g., number of steps or skills involved in providing an answer)
- How characteristics of students from this population might limit their ability to demonstrate evidence about their knowledge in a specific way

Once the Focal KSA, Potential Observation, and Potential Work Product are decided upon, Characteristic Features are reviewed to remind the co-design team about the critical task features that must be present. Potential Variable Features are also reviewed so that the co-design team can consider possible ways to vary the four items. It is possible that the co-design team will propose additional Characteristic Features and Potential Variable Features. If it is determined that a proposed Characteristic Feature (not already within the *Design Pattern*) applies to all tasks created from a *Design Pattern*, it should be added.

For consistency the co-design team should update the *Design Pattern* by adding any new Potential Observations, Potential Work Products, Characteristic Features, and Variable Features that are generated during the task development process. Consistency of content between the *Design Pattern* and *Development Specifications and Exemplar Task Template* is critical. Note that this reconsideration or revision to the *Design Pattern* illustrates the iterative nature of the ECD process for developing both *Design Patterns* and *Development Specifications and Exemplar Task Templates*.

Step 4. Determine the Task Requirements for the Item

As items are created it is important to keep the following considerations in mind:

Presence of context—A decision must be made about whether to include a context or to present the task in a decontextualized fashion. For example, if the Focal KSA aims to assess the students' ability to calculate summary statistics, a contextualized item can be developed, "In a recycling contest, students collected aluminum cans. This data table shows how many aluminum cans each student collected. What is the mean number of cans collected?" Alternatively, a decontextualized item can be developed, "Using these 10 data points, calculate the mean." Including context can make an item more interesting and engaging to students, but it can also increase the cognitive demand in a nonconstruct relevant way. If the decision is to have context present, here are further considerations in choosing one that is appropriate:

---

content standards at grade level then considering how items can be translated so that students at different levels of functioning or communication would be able to access it.

- Choose a context that is grade-level appropriate and respectful. For instance, when targeting students in the grade 9–12 range, a recycling contest was the chosen context for students to demonstrate their ability to answer a question about data by identifying, creating, and using a graphical display, and calculating and using a summary statistic. Although the use of a marbles contest could allow the assessment of the same mathematical skill, it would not have been grade-level appropriate.
- Establish a context that is realistic where possible. For example, if inches of rainfall during the year is the chosen context, the data points included should reflect what is typical and realistic.
- Ensure concrete examples are used in the context where possible. For example, discuss mathematical relations in the context of everyday situations.
- Use a context that is generalized where possible. For example, when discussing rainfall, instead of referring specifically to rainfall within a particular state (e.g., Hawaii or Florida), it is important to discuss rainfall in general so that the technical accuracy of the information (i.e., knowing the amount of rainfall that occurs in a particular state) is not the subject of the question.
- Choose a context that is clear and unambiguous.

Student response mode—A decision must be made about whether students will be asked to select the correct response from a set of response options or whether the student will be asked to construct the correct response on their own. If the student is asked to construct the correct response, another decision must be made about whether students would be asked to construct a verbal response, a graphical representation, a computer generated response, a concrete representation of their response, or a written response. The assessment designer must consider the relationship between the response mode required and the specific cognitive limitations of the students. It is possible that although an item may be designed with a particular response mode in mind, it may need to be modified by the test administrator at the time of administration given a particular student's capabilities.

Presence of data—A decision must be made about whether data will accompany the text, and if so, the following questions should be considered:

- Should the data be presented within a table, graphically, or in a list?
- If data is presented graphically, what type of display should be used (e.g., line graph, pie chart, bar graph)?
- Should the data be rich enough to allow the assessment designers to ask a range of nontrivial or interesting questions or should the data be limited to a specific question without extraneous information, relationships, or variables illustrated?
- How many data points should be presented?
- Should single and/or double-digit numbers be included (e.g., 9, 14)?
- Should categorical and/or numerical data be presented? ;
- How complex should the highlighted relationship be in the data distribution?

Number of questions within an item—A decision must be made about whether one question or multiple questions should be asked of the student. This may depend on the complexity of the Focal KSA and on the approaches states are using and whether item interdependency can be addressed in their measurement model.

- If multiple questions are asked, should they be asked in the same context and/or data set or should multiple contexts and/or data sets be progressively built into the items?
- Should an overall framing or thematic question be included when multiple questions are asked?

***Number of steps to the solution*—**A decision must be made about how many steps should be involved in getting to the final solution.

Step 5. Develop the Item Directive

In Section C of the template specific task information is generated and recorded. This information will be recorded for each of the 4 items within the task. It is suggested that co-design teams work through steps 5–9 for Item 1, then go back and repeat these steps for Item 2, and finally go through them again to create Items 3a and 3b.

The Item Directive segment of the template includes the item prompt or question, the item description and distracters when applicable, and specific instructions that will be presented to students for each item. For the AAD-M project the convention was adopted that text in bold was to be read aloud by the examiner. The Item Directive does not detail specific individual adjustments that can be made (and that are acceptable) in the task administration. This information concerning individual adjustments is presented in the Variable Features for Administration to Individual Students section of the template (described in detail in step 9).

The mathematics expert within the co-design team typically suggests an idea for the Item Directive, taking into consideration the Focal KSA, the decisions made about the task requirements, their experience in the classroom, and the best way to assess the mathematics concepts targeted.

After drafting an initial representation of the idea for the Item Directive, the team discusses and modifies the Item Directive based on insights from differing perspectives, such as the principles of ECD, mathematics education, and classroom experience with special education students. The concerns addressed in these discussions should include:

- Capabilities of students in the target population
- Construct relevant and irrelevant details elicited by the proposed Item Directive:
  - Whether the proposed Item Directive adheres to the Focal KSA
  - What Additional KSAs might be required by the task
  - How to minimize or support the Additional KSAs within the design of the Item Directive
- Evaluation of the content of the Item Directive:
  - Context (see criteria in step 4)
  - Data presentation (see criteria in step 4)

To illustrate this process, the following is an example of the Item Directive for Item 1 of Number and Operations A3 (grades 3–5), which will be further elaborated in steps 6–7. The examiner presents students with three drawings of pizzas/pies and says, "**Here are three drawings of parts of a pizza**." The examiner then presents a card with the numeric fraction "¾", places it on the table in front of the student, and then asks, "**Which drawing shows three-fourths of a pizza?**" (Note: Some items may include multiple options for context information. These options are placed within square brackets [ ]. They are provided within the Item Directive to allow for maximum flexibility and appropriateness according to specific characteristics of the population. For instance, a board game [as opposed to a video game] may be a more appropriate example of a prize for populations from lower socioeconomic backgrounds; hence, it is provided as a possible replacement option).

Step 6. Document the Correct Answer

After the co-design team has reached consensus on the Item Directive, they next document the Correct Answer. The answer can be a number, graph, or description. The team should also specify whether alternative versions of the stated correct answer are also acceptable. For example, for the item created for Number and Operations A3, grades 3–5, the Correct Answer to the Item Directive (e.g., Which drawing shows three-fourths of a pizza?) is "**Student indicates the picture of ¾ of a pizza**."

Step 7. Describe the Stimulus Items and Materials for the Examiner

The Description of the Stimulus Items is a depiction or detailed description of the graphics, objects, or tools to be used in task administration. This might include a table of data presented to the student with which they must create graphics or interpret, synthesize, and/or calculate statistics. If there are multiple questions within an item, there will be a description of the stimulus materials for each question. The Stimulus Materials for Item 1 of Number and Operations A3, grades 3–5 include:

- Illustration of three pizzas divided into quarters; one has two quarters remaining, one has one quarter remaining, and the third has three quarters remaining
- Note card with the numeric fraction ¾

The Materials for the Examiner is a description of the materials examiners will need to administer, document, and score an item (e.g., worksheet, camera with which to take a picture of product, or a manipulative). It includes the task worksheet that describes the item and delivery instructions and task data sheet or other method to record the student's response.

Step 8. Update Selected Variable Features

The co-design team must return to the Selected Variable Features to update the information based on the selections made for the finalized Item Directive. The team first decides on the DOK level for the item. Using the 6-point DOK scale (Flowers, et al., 2007), the team decides which level best exemplifies the DOK required by the Item Directive created for the item. This decision is based on a number of factors including:

- ***Understanding of the structure*** of the DOK levels and the verbs used to exemplify each level, including how each level and verb can be operationalized generally in the context of mathematics and more specifically in the context of the item. For instance, an item that asks students to explain and/or make a conclusion is considered to be at the comprehension level.
- ***Determining the mathematical sophistication*** of what is elicited by the item based on the abstract nature of the mathematics concept being probed based on (1) the amount of prior mathematics knowledge that has to be drawn upon, (2) the number of mathematical principles required for the solution, and (3) whether the question can be answered with a procedure or routine.
- ***Determining the complexity*** of what is elicited by the item based on (1) whether the student has to extend or produce novel findings, (2) whether the item has multiple questions or requires multiple or integrated skills, and (3) whether the answer is a constructed response or selected response. In addition, the distracters in a selected response item can be written to impact the item's complexity.

If the DOK assigned to Item 1 is lower than desired, the team may decide to use Item 1 as an Item 2 or may revise the Item Directive to increase the DOK level of the item.

The co-design team should explicitly detail the decisions made for each Variable Feature selected to create the Item Directive. For instance, if the co-design team chooses to ask students to create a histogram (rather than a scatter plot or box plot), then they must document this decision.

Step 9. Document Variable Features for Administration to Individual Students

Variable Features for Administration to Individual Students specify task features that could be changed to impact item accessibility according to individual student needs (e.g., large print, Braille for those with visual impairments). Although the Item Directive will not be modified, it is possible that certain students will require specific accommodations in addition to the accessibility and scaffolding features built in to the design of the item. The boundaries of this category will be determined in part by accommodation policies in individual states. However, it is certain that these Variable Features should not compromise the construct (Focal KSA) targeted. Currently, two types of Variable Features for Administration to Individual Students have been consistently noted in the *Development Specifications and Exemplar Task Template*: (1) the freedom to vary the format of the question presentation (e.g., presented in sign language with Braille, auditory, or with or without gestural prompts) and (2) the students' response format individualized based on their communication system. States need to specify which accommodations or formats are and are not allowed

Step 10. Repeat Steps 5–9 to Develop Item 2

The co-design team should repeat Steps 5–9 to develop Item 2. Item 2 must assess the same Focal KSA as Item 1, but it involves skills that are considered to be at a lower DOK level. In addition, Item 2 is typically less complex, more narrow in scope, and more heavily scaffolded or supported. In creating Item 2, the modifications below should be kept in mind. These modifications help to ensure that the DOK and scope have been appropriately decreased and that supports or scaffolding have been appropriately increased relative to Item 1 while still preserving the Focal KSA.

- ***Reduce DOK Levels:***
  - If Item 1 required students to *construct* a response (a higher DOK level), in Item 2 students can be asked to *select* the appropriate answer from a set of response options (a lower DOK level).
- ***Reduce Complexity:***
  - If Item 1 asked students to *create a scatter plot or box plot*, Item 2 can ask for the *creation of a histogram*, which is technically less sophisticated. A histogram is focused on the frequency of one variable, while a scatter plot is about the relationship between two variables.
  - If Item 1 presents *20 data points* to be mathematically represented, Item 2 could present only *10 data points*.
  - If Item 1 contains *4 subquestions* (i.e., a, b, c, and d), Item 2 could contain only *2 questions* (i.e., a, b).
- ***Narrow the Scope of Content to Be Assessed:*** If Item 1 assessed *a composite set of skills* (e.g., students determine the appropriate representation to be used to answer a research question, create that representation, and then use the representation to answer the research question), then Item 2 should *assess fewer components of those skills* (e.g., perhaps students just create and use the representation).
- ***Increase Scaffolding or Support:*** If the Focal KSA is about creating mathematical representations, Item 1 might ask students to create the representation with little support. Item 2 will increase the amounts and kinds of scaffolding within the design of the item. For instance, students could be provided graph paper to support the creation of a graphical representation (e.g., histogram) and/or students could be provided with key elements of the graph already completed (e.g., axes, labeled axes, and bins).

Step 11. Repeat Steps 5–9 to Develop Items 3a and 3b

Steps 5–9 also should be followed to complete Items 3a and 3b to ensure systematic development and documentation of design decisions for these items. However, recall that for these items an Additional KSA (not the Focal KSA) is targeted.

Some important considerations developing Item 3a are as follows:

- For consistency, select an Additional KSA that is aligned to the selected Focal KSA.

- The choice and use of an Additional KSA (or prerequisite skill) that is narrowly focused increases the likelihood that the item is less sophisticated than Items 1 and 2.
- Ensure that students at the lower functioning end of the spectrum of students with significant disabilities are taken into account in the design of this item.

Item 3b targets the attention DOK level. This usually involves removing all distracters from Item 3a and leaving only the correct answer for the student. The student is asked to point to or otherwise indicate the remaining stimulus item. This item is included in an effort to ensure that all students, including those with the most severe cognitive disabilities, will be able to participate in the testing experience and encounter some success.

Design for Pilot Task Tryouts

The collaborating states pilot-tested the newly developed assessment tasks with teachers administering them to students eligible to take state AA-AAS. Grant funds were be used to reimburse teachers for their time in administering the pilot tests and to incentivize students to participate. 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 are fully described in a technical report (Technical Report 7: Pilot Task Tryout Design[8]). To pilot the 120 newly designed items (based on 30 design patterns with a suite of 4 items for each design pattern), each state gathered information by administering the items using common instructions and a common data collection system. In addition to item scores, survey data was collected on each item's characteristics as they were administered, whether students had an opportunity to learn the knowledge and skills assessed, and the characteristics of the students and teachers involved in the piloting. The pilot survey analysis focuses on items variability and the appropriateness of the items to measure a range of student performance levels.

Task viability. Teachers who administered the pilot task tryouts to students judged the viability of the each of the items. Can the four items associated with a design pattern be administered as designed? Are the item instructions and materials clear to the teacher? Are they clear to the student. Data will inform future improvement of the tasks.

Appropriateness of tasks to measure a range of student performance levels. Tasks were 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 most complex item and at what level of symbolic functioning? Which students successfully perform the decreasingly complex items and at what level of symbolic functioning? Data will inform future modification of tasks so all or most students can gain access to at least one task associated with each design pattern.

(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

[8] Technical Report Series available at http://alternateassessmentdesign.sri.com

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 described.

Website. SRI has developed and maintains a website to post study reports and selected products: http://alternateassessmentdesign.sri.com. The site includes links to other 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.

Webinar. SRI hosted a free 2-hour web-based interactive seminar presentation to disseminate project procedures and outcomes to researchers and local, state, and national educators.

Technical Report Series including Procedural Guidelines. Each report in the Technical Report series is 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 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

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