SRI International

Alternate Assessment Design–Mathematics

Technical Report 4: Design Patterns

The Background and Role of Design Patterns in the Evidence-Centered Design Process

May 6, 2011

SRI International
Geneva Haertel
Renée Cameto
Kathryn Morrison
Angela Haydel DeBarger
Katherine Nagle
Kavita Seeratan
Reina Fujii
Anne-Marie Knokey

SRI Project No.: P18879

This report was supported in part by the Elementary and Secondary Education Act, Enhanced Assessment Grant (ESEA grant No. # S368A08008) through a partnership with the Utah (Lead), Idaho, and Florida State Departments of Education and SRI International as the research entity. Any opinions, findings, or conclusions are those of the authors and do not necessarily reflect the view of ESEA or the U.S. Department of Education.

SRI International 333 Ravenswood Avenue • Menlo Park, California 94025-3493 • 650.859.2000 • www.sri.com

Contents

Figure

Tables

Introduction and Background

Evidence-Centered Design (ECD) is a view of assessment as evidentiary argument: it is an argument from what we observe students say, do, or make in a few circumstances to inferences about what they say, do or make more generally (Mislevy, Steinberg & Almond, 2003). ECD can serve as a cornerstone of test validation, providing items that are well-matched to the domain definition and inferences that can be drawn from students' performances. As the ECD process is implemented and the test is developed, the domain from which the content is drawn is delineated at both general and specific levels and items are created to assess the key aspects of the domain. Thus, both content and construct evidence for validity is built in during the development of the items (Ebel & Frisbie, 1991; Fuhrman, 1996). This technical report lays out the basic ideas of ECD and then focuses intensively on the second stage of ECD, referred to as domain modeling and the theory and use of Design Patterns. Below the layers of ECD are introduced and some details about each layer are presented. Attention is paid, in particular, to the domain modeling layer during which Design Patterns are created.

Layers in Evidence-Centered Assessment Design

ECD is organized around the five layers described in Table 1. The layers are referred to in terms of the roles they play in the assessment design and development process: Domain Analysis, Domain Modeling, Conceptual Assessment Framework, Implementation, and Assessment Delivery. Each layer involves the use of key concepts and entities, knowledge representations, workflow, and communications tools.

Because ECD enables test developers to refine, document, and implement the functions and design decisions within each of the five layers independently, the developers can carry decisions through the other layers to guarantee that the eventual pieces of the operational assessment are consistent with each other and with the intended assessment argument. Not all elements of all layers may be detailed in a given assessment; different assessments, depending on their nature and purpose, will focus more attention on some layers than others. Each layer in the assessment design process is briefly described below.

Domain Analysis (Layer 1)

In Domain Analysis, the assessment designer gathers information about concepts, terminology, representational forms, and ways of interacting in the domain to be assessed. Lists of content and process standards, statements of "big ideas," classroom experience, and cognitive research are examples of sources that can be collected and examined during the assessment design process.

Domain Modeling (Layer 2)

In Domain Modeling, information that is gathered during Domain Analysis is organized along the lines of an assessment argument. This layer articulates the argument that connects observations of students' actions in various situations to inferences about what they know or can do. In Domain Modeling, the assessment argument takes a narrative form—the assessment designer may sketch descriptions of proficiencies of interest, observations that provide evidence of those proficiencies, and ways of arranging situations in which students can provide evidence of their proficiencies. This is the layer of ECD in which the knowledge representation referred to as a Design Pattern is created. Design Patterns were developed originally in an NSF-funded project, "*Principled Assessment Designs for Inquiry* (PADI)" (Mislevy, et al., 2003).

Table 1. Layers of Evidence-Centered Design for Educational Assessments

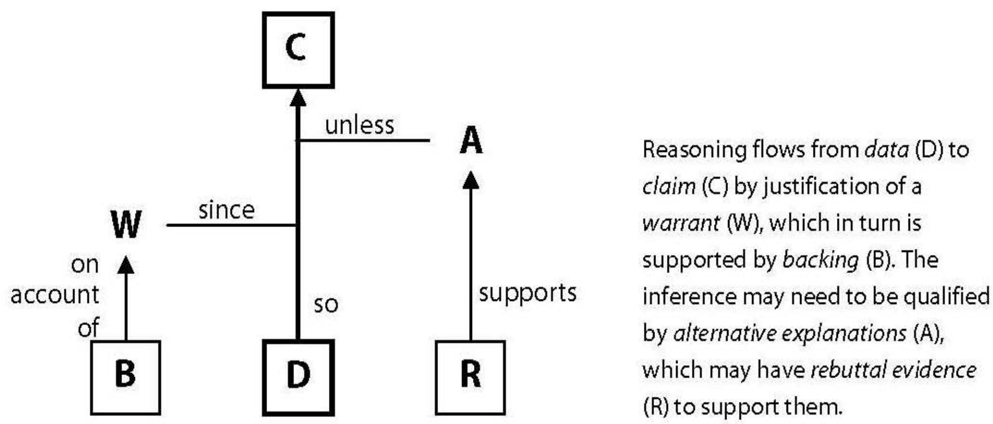
Layer	Role	Key Entities	Selected Knowledge Representations
Domain Analysis	Gather substantive information about the domain of interest that has direct implications for assessment; how knowledge is constructed, acquired, used, and communicated.	Domain concepts, terminology, tools, knowledge representations, analyses, situations of use, patterns of interaction.	Representational forms and symbol systems used in domain (e.g., algebraic notation, Punnett squares, maps, computer program interfaces, content standards, concept maps).
Domain Modeling	Express assessment argument in narrative form based on information from Domain Analysis.	Knowledge, skills, and abilities; characteristic and variable task features, potential work products, potential observations.	Toulmin and Wigmore diagrams, PADI design patterns, assessment argument diagrams, "big ideas" of science.
Conceptual Assessment Framework	Express assessment argument in structures and specifications for tasks and tests, evaluation procedures, measurement models.	Student, evidence, and task models; student, observable, and task variables; rubrics; measurement models; test assembly specifications; PADI templates and task specifications.	Algebraic and graphical representations of measurement models; PADI task template; item generation models; generic rubrics; algorithms for automated scoring.
Assessment Implementation	Implement assessment, including presentation-ready tasks and calibrated measurement models.	Task materials (including all materials, tools, affordances); pilot test data to hone evaluation procedures and fit measurement models.	Coded algorithms for rendering tasks, interacting with examinees and evaluating work products; tasks as displayed; IMS/QTI representation of materials; ASCII files of item parameters.
Assessment Delivery	Coordinate interactions of students and tasks: task-and test-level scoring; reporting.	Tasks as presented; work products as created; scores as evaluated.	Renderings of materials; numerical and graphical summaries for individual and groups; IMS/QTI results files.

The original PADI project ended in 2008, but the online assessment design system that was developed through the NSF grant continues to be used by others involved in designing assessments. Approximately 162 Design Patterns have been developed—some of these Design Patterns were created during the original PADI project, but others during subsequent assessment design projects. These subsequent projects encountered Design Patterns either as a software template that was part of the PADI online assessment design system or as a stand-alone word document that was completed during an assessment design process implemented outside of the PADI system.

A core of Design Pattern attributes are specified in each Design Pattern developed; these core attributes are introduced later in this technical report. They are associated with Messick's (1994) conceptualization of the assessment argument. Oftentimes, however, a particular assessment design project may add an attribute to the standard set in order to address a particular feature needed for the design work. For example, in the NSF-funded project titled, "*An Application of Evidence-Centered Design to a State's Large Scale Science Assessment*," the Design Pattern attribute called Narrative Structures was added to the Design Pattern. Narrative Structures are helpful in thinking through the storyline that had to be presented as part of each scenario-based assessment task to be used in the assessment being designed. In the AAD-M project, no additional attributes were added to the Design Pattern form, although the list of Additional Knowledge, Skills, and Abilities (KSAs) to be identified was extended to include not only Cognitive Background Information (e.g., prerequisite knowledge about the content being assessed), but also many types of Universal Design for Learning (UDL) knowledge and skills that might be required for successful performance. Thus, the Design Pattern form and attributes for any particular assessment design process may vary to reflect project requirements.

The concern of ECD at this domain modeling layer is to articulate the key elements of an assessment argument schema. Toulmin's (1958) diagram for argument structures provided a general structure that captures the features of arguments, in terms of claims, data, and warrants. These argument features provide a starting point for domain modeling. (See Figure 1 for the basic structure of Toulmin's argument form.) The claim (C) refers to the target of the assessment, such as level of proficiency in scientific problem-solving, or ability to use language appropriately in varying contexts. Data (D) refers to the quality of responses to questions, or behaviors observed in particular situations and are provided to support the claims. The warrant (W) is the logic or reasoning that explains why certain data should be considered appropriate evidence for certain claims. Much of the information for constructing the argument will have been marshaled during Domain Modeling, although cycling across layers is the norm in practice.

Figure 1: Toulmin's (1958) Structure for Arguments



Design Patterns are an example of a knowledge representation that supports work in the domain modeling layer (Mislevy et al. 2003; Mislevy & Haertel, 2006; Mislevy, Behrens et al., 2010). Analogous to design work in architecture (Alexander, Ishikawa, & Silverstein, 1977) and software engineering (Gamma et al., 1994), users of ECD in assessment rely on Design Patterns to help organize information from domain analysis into the form of an assessment argument. Design Patterns function like design objects.

Design Patterns help assessment designers complete an assessment argument around some theme in the domain of interest, such as model-based reasoning in science (Mislevy, Riconscente, & Rutstein, 2009), negotiating apology situations in language testing, or interpreting fractions in mathematics. The structure of the Design Pattern is organized around the structure of an assessment argument. Thus, filling in the Design Pattern renders explicit the relationships among the information that is required to guide the development of assessment tasks in a particular domain. The information entered into the Design Pattern is related to the components that comprise the assessment argument—the student, evidence and task models. The student, evidence, and task models are foreshadowed in the Design Pattern attributes and further specified by the assessment experts in the third layer of ECD, the Conceptual Assessment Framework (CAF); while the domain content to be assessed is contributed by the assessment designer and content experts as they complete the Design Pattern.

Table 2 shows: (1) the attributes of a Design Pattern, (2) definitions of the attributes, (3) the connection of attributes to the Toulmin assessment argument (claims, actions, and situations), and (4) connections of the attributes to the student, evidence and task models which comprise the Conceptual Assessment Framework (CAF), the third layer of ECD. (The CAF is discussed in the following section of this report) Centered on the knowledge, skills and abilities (KSAs) in a content domain, a Design Pattern offers approaches for gathering evidence about those capabilities, organized in such a way as to lead toward the design of particular tasks, scoring rubrics, measurement models, and other more technical elements required in a well-designed assessment.

Table 2: Assessment Argument Elements and Design Pattern Attributes

Assessment Argument Elements & Guiding Questions	Design Pattern Attribute	Definition of Design Pattern Attribute
	Title	Short name for the Design Pattern
	Summary	Brief description of the family of tasks implied by the Design Pattern
	Rationale	Nature of the KSA of interest and how it is manifest
Student Model/Claim What construct (complex of student attributes) should be assessed?	Focal KSAs	The primary knowledge/skills/abilities targeted by this Design Pattern
	Supported Benchmarks	State benchmarks that this Design Pattern supports
	Additional KSAs	Other knowledge/skills/abilities that may be required by tasks motivated by this Design Pattern
Evidence Model/Actions What behaviors should reveal the construct?	Potential Observations	Things students say, do, or make that can provide evidence about the Focal KSAs
	Potential Work Products	Features of Work Products that encapsulate evidence about the Focal KSAs
Task Model/ Situation What tasks elicit those behaviors?	Characteristic Features	Aspects of assessment situations likely to evoke the desired evidence.
	Variable Features	Aspects of assessment situations that can be varied in order to control difficulty or target emphasis on various aspects of the KSAs

Conceptual Assessment Framework (CAF) (Layer 3)

The work at Layer 3, Conceptual Assessment Framework, focuses on technical specifications for the "nuts and bolts" of the assessment. Three models comprise the CAF: student, evidence, and task. These three models are specified by the assessment designer and are linked via the student-model variables, observable variables, work products, and task model variables (Mislevy & Riconcente, 2005). Detail about task features, measurement models, structures, and stimulus materials are expressed in terms of representations and data structures.

The Student Model identifies aspects of student proficiencies. The number, character, and grain size are determined to serve the purpose of the assessment. The Task Model describes the environment in which students say, do or make something. It specifies the forms and key features of directives, stimulus materials, and features of the presentation such as simulation capabilities in technology-based tasks. A key decision is specifying the work products – the assessment designer may choose among alternative formats such as multiple choice, open-ended items, performance tasks, oral presentations, or essays. Other examples of task related decisions that assessment designers make include specifying the number, sequence, and complexity of steps to be completed in a multipart task such as an investigation, specifying the "look and feel" of the graphical interface that is used in online assessment tasks, or the degree of scaffolding provided for a task.

The Evidence Model bridges the student and task models. It consists of two sub-models: the evaluation component and the statistical component. The first component is task-level scoring: identifying and evaluating salient aspects of student work, to produce values of observable variables. The component, test-level scoring, synthesizes data across tasks using a measurement model, such as simple number-right scores, Item Response Theory (IRT) modeling, or Rasch analyses. In the AAD-M project, a formalized version of the CAF was not built. A summary task template was created to document key information about the scoring and measurement model used in each state participating in the project, but detailed scoring specifications and modeling parameters were not pursued. Since the goal of the AAD-M project was to illustrate the use of Design Patterns and their support for assessment task development, the specification of the CAF was beyond the resources for this project.

Assessment Implementation (Layer 4)

The work at the Assessment Implementation layer includes authoring tasks, finalizing rubrics or automated scoring rules, estimating parameters in the measurement models, and producing fixed test forms or algorithms to assemble tailored tests**.** Because of the compatible data structures developed in the prior layers, the assessment designer can leverage the value of the design system for authoring or generating future tasks, calibrating items, presenting materials, or interacting with examinees.

Assessment Delivery (Layer 5)

In the Assessment Delivery layer of ECD, the test taker interacts directly with tasks, performances are evaluated, and feedback and reports are produced. The delivery system architecture that has been incorporated in ECD is the Four Process Delivery System (Almond, Steinberg & Mislevy, 2002).

The Design Pattern

Design Patterns bridge knowledge about aspects of a domain that an individual wants to assess and the structures of a coherent assessment argument in a format that can guide task creation and assessment implementation. The focus at the Design Pattern level is on the substance of the assessment argument rather than on the technical details of operational elements and assessment delivery systems, which are addressed at subsequent layers of the ECD process (i.e., the Conceptual Assessment Framework (CAF) layer, the Implementation layer, and the Delivery layer). In this section of the technical report, the nature and role of Design Patterns in assessment design is considered. Appendix A contains a Design Pattern in the area of Data Analysis and Probability. It is a fully developed example of a Design Pattern developed for the AAD-M project. This Design Pattern is titled "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." Other AAD-M technical reports contain examples of additional Design Patterns that were used to guide the development of exemplar mathematics tasks for students with severe cognitive disabilities as part of the AAD-M project. Appendix A also contains a Development Specifications and Exemplar Task Template. This template is used to support the development of the assessment tasks.

The Role Design Patterns Play in Assessment Design

As stated in the prior section, the domain modeling layer of ECD specifies the relationships among the knowledge and skills in the domain to be assessed. Design Patterns are an example of a domain modeling tool. In the case of the AAD-M project, all of the Design Patterns generated are all within the domain of mathematics (grades 3-8) and each Design Pattern falls within one of the five strands of NCTM—Numbers and Operations, Algebra, Geometry, Measurement and Data Analysis and Probability. The NCTM standards and expectations, as well as the extended mathematics standards for the states of Utah, Idaho, and Florida were analyzed by a mathematics educator to determine which extended standards were common to the three states and to which NCTM expectations these common standards were associated. After this crosswalk of relationships was completed, a set of common expectations were identified for the purposes of generating Design Patterns and exemplar tasks. This domain analysis of the mathematics content to be assessed was followed by the generation of Design Patterns. For each of the common expectations that were identified, the AAD-M project developed a Design Pattern which bridged the mathematics content, measurement expertise, and special education expertise. All of these types of expertise are needed to create operational assessment tasks in the domain of mathematics for students with severe cognitive disabilities.

As assessment diagrams, like the Toulmin diagram displayed in Figure 1, provide graphic support for understanding the structure of an overall assessment argument, Design Patterns provide support for detailing the substance of the assessment argument for the purposes of the assessment task development. Expertise research has provided common themes in the ways increasingly proficient people structure and use their knowledge in areas as diverse as chess, architecture, volleyball, shipboard navigation, and emergency room medicine (Ericsson, 1996). Identifiable kinds of things people do in certain kinds of situations are observed in domains and at levels of education quite different in their particulars. An example is the phenomenon of "design under constraint," which is clearly at the heart of engineering and architecture but is equally apropos in creative domains such as writing a story or play and everyday activities such as planning a vacation. Being able to recognize constraints, use strategies to address them, and monitor how one is progressing are common skills required for developing proficiency in any domain where one must "design in the face of constraints."

"Designing under constraints" is a schema that assessment designers may want to recognize in any domain that is the target of assessment. Design Patterns can be developed for different purposes. For example, a pattern for "designing under constraints" can be created to "flesh" out the attributes of an assessment argument to be applied ***across domains*** of expertise. A Design Pattern for "designing under constraints" also could be used to create a family of assessment tasks ***within a specific domain***. Many times assessment designers are asked to develop tasks that evince this aspect of proficiency in the context of the domain's particulars. A Design Pattern also can be created to develop an assessment argument that would generate a family of tasks ***within a standard***. For example, an assessment designer can create a Design Pattern about "describing the shape and important features of data" and in so doing lays out the underlying assessment argument structure. As part of creating a Design Pattern, several KSAs can be identified for a given standard, objective, domain of expertise or cross-domain theme. Then using the potential variable features, which are part of the task model, and identified in the Design Pattern, the assessment designer can change the surface features of tasks and generate new tasks associated with the particular KSA. Thus, a Design Pattern can be used to guide the creation of a family of assessment tasks in a systematic and reliable way and ensure that the tasks will be closely aligned to the standard, learning objective, domain of expertise or cross-domain theme that they are intended to measure. Thus, Design Patterns are analogous to design objects in other fields. They organize experience across many particular situations in ways that help a designer recognize and tackle challenges such as planning work flow in a kitchen, generating software objects, creating clothing with intricate details and features on a large scale, or building a complex structure with several interdependent processes. Design Patterns for assessment design likewise help domain experts and assessment specialists "fill in the slots" of an assessment argument built around recurring themes in learning (Mislevy et al., 2003).

"Filling in the Slots" of a Design Pattern to Create an Assessment Argument

Design Patterns are intentionally non-technical, "centered around some aspect of KSAs, a Design Pattern is meant to offer a variety of approaches that can be used to get evidence about that knowledge or skill, organized in such a way as to lead toward the more technical work of designing particular tasks" (Mislevy & Riconscente, 2006, p. 72).

When a Design Pattern is completed, it specifies elements that can be assembled into an assessment argument:

- **Focal Knowledge, Skills, and Abilities** (KSAs) indicate the main claims about students that tasks created from the Design Pattern address. In the AAD-M Project, a focal KSA is related to the particular knowledge, skills and abilities associated with one of the NCTM expectations. Here is a Focal KSA from the Design Pattern on the following expectation, "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" One of the Focal KSAs associated with this expectation is "Ability to identify and/or calculate summary statistic to answer a question when given a data set and a question about the data."

- **Additional KSAs** may also be required to complete a task, such as whether familiarity with certain representational forms or mathematical operations is presumed. Additional KSAs are typically dealt with by supporting a student's performance through the provision of content, skills, or examples that are not related to the targeted Focal KSA or by providing accommodations. The Additional KSA attribute makes task authors aware of assessment design choices and their implications—including possible explanations for poor performance due to the task requiring knowledge or skills other than the targeted KSA. These additional KSAs, if not addressed, may become sources of construct-irrelevant variance in Messick's (1989) terminology. An example of an Additional KSA associated with Focal KSAs specified in the preceding paragraph is "Ability to read graphs, e.g., line graphs and bar graphs." This Additional KSA is a required skill in order to demonstrate the Focal KSA specified above.
- **Potential Work Products** are things students might say, do, or make that provide information about the Focal KSAs. Two Potential Work Products associated with the Focal KSA specified above are: (1) Selection from a list of phrases that describe the data; and (2) Description of data sets.
- **Potential Observations** are the aspects of the work products that constitute evidence. An example from the Design Pattern on probability and data analysis is "Student accurately describes one feature of a data set (e.g., which quantity occurs most frequently, the average quantity, the maximum and minimum values)."
- **Potential Rubrics** are ways that the test administrator or scorer might evaluate work products in order to produce values along the dimension being observed.

All of the Design Pattern attributes described above concern ways of getting evidence about the targeted proficiency (the Focal KSA)—and the wider the array of ways to get evidence, the better, so assessment designers can choose among a variety of possibilities to obtain evidence to suit the resources, constraints, and purposes of their particular situation. Characteristic and Variable Features of assessment tasks described specify aspects of the situation in which students act and produce work products.

- **Characteristic Features** are those that all assessment tasks motivated by the Design Pattern should possess in some form, because they are central to evoking evidence about the Focal KSAs. For example, all tasks inspired by the AAD-M Design Pattern, "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" must involve graphs.
- **Variable Features** address aspects of the assessment that the assessment designer can use to affect difficulty of the tasks or the focus of attention. In the "Describe the shape and important features of a set of data…." for example, the type of data representation that a student receives is a key Variable Feature since the NCTM expectations about data analysis and probability stress learning to work with data representations that vary in difficulty, such as line graphs, bar graphs, and pictographs.

As the assessment designer "fills in the slots" in the Design Pattern, the components or elements of the assessment argument described by Samuel Messick (1994) are foreshadowed. The Focal KSAs identify the proficiencies that will be included in the student model. The Additional KSAs identify the threats to the validity of the claims or inferences that can be drawn from the evidence acquired. The Potential Observations, Potential Work Products, and Potential Rubrics make clear the kinds of evidence that will be gathered and scored. Eventually these three types of information will be used to construct an Evidence Model, including evaluation decision rules and a measurement model. Finally, the Task Model will be constructed drawing on: (1) the Characteristic Features that must be present in all tasks, and (2) Potential Variable Features that can be manipulated to make the tasks vary in difficulty or focus. Completing the Design Pattern (filling in the slots) is not the same as articulating the student, task and evidence models, but it is the first step in the process and greatly supports the development of assessment tasks aligned to the Design Pattern.

Conclusion

Work at the domain modeling layer is important for improving the practice of assessment, especially for the reasoning and capabilities for situated actions that cognitive psychology calls to our attention. Experience with experimental tasks is valuable, but it is confounded with particular domains, psychological stances, knowledge representations, and delivery vehicles. Because proficiencies are the primary organizing category in Design Patterns, the assessment designer is able to keep a focus on the proficiency of interest and make sure a coherent assessment argument results. The specifics of response types, stimulus materials, measurement models, and delivery modes are then determined in light of the particular constraints and resources of the application.

Liu and Haertel (2011) argue that Design Patterns are an epistemic form, similar to those catalogued and described by Collins and Ferguson (1993) and further illustrate the value of such tools in addressing complex design tasks. Collins and Ferguson chose the term “epistemic form” to underscore how a representation that builds around important principles can be a powerful cognitive tool, to help people organize work, coordinate their activities, and even construct new knowledge. Collins and Ferguson point out that epistemic forms range from simple lists to more complex forms such as blueprints and financial reports. Design Patterns are such a tool: the domain is assessment design, the underlying idea is the essential structure of assessment arguments, and the task at hand is to write assessment tasks.

Collins and Ferguson assert that to use an “epistemic form” to full advantage, the user must learn to play the “epistemic game” required by the form. In the case of assessment design, the games one must learn to play with Design Patterns concern how to use the support they provide for relating aspects of task features and scoring with validity argumentation into the larger design process—which includes deep knowledge of the content area, the students to be assessed, and the constraints and the resources that characterize the assessment project at hand.

Design Patterns are particularly useful in guiding the development of complex assessments, including those used in the AAD-M assessment tasks. The Design Pattern has attributes that can be used to guard the validity of these complex assessments, which must take into account a large number of disabilities that can make it difficult for a student’s performance to be solely attributed to the knowledge and skill required by the Focal KSAs. The Design Pattern is seen to be a construct-oriented support tool, rather than simply just an organizational or procedural support tool. With the affordances of the Design Pattern, the assessment designer can guard against the introduction of construct irrelevant variance being inadvertently introduced into the assessment task due to the absence of skills that range from executive processing to perceptual capabilities to the use of language and symbols. While the Design Pattern tool does not eliminate all construct irrelevant variance, it certainly can mitigate against the "noise" introduced into assessment tasks through less systematic design processes.

The implementation of NCLB has called for the design of assessments that can validly measure domain content and skills for all students, including those with significant cognitive disabilities. This demand challenges both expert and novice assessment designers alike. We remain hopeful that Design Patterns are a support tool, derived under the framework of ECD that can usher in a new era of alternate assessment for these students. Design Patterns can document and make available the tacit knowledge that characterizes the work of experienced and talented assessment task developers With Design Patterns, assessment developers can remain focused on the proficiencies of interest and apply the cognitive knowledge more readily to the situational contexts required for the particular student population. In sum, Design Patterns are re-usable, generative, and sharable—with these documents, the advances in assessment design for special needs populations can be supported, documented, and communicated in such a way that the difficult thinking required to build valid tasks for these students can be made available to other designers.

References

Alexander, C., Ishikawa, S., & Silverstein, M. (1977). *A pattern language. Towns, buildings, construction.* New York: Oxford University Press.

Almond, R. G., Steinberg, L. S., & Mislevy, R. J. (2002). A four-process architecture for assessment delivery, with connections to assessment design. *Journal of Technology, Learning, and Assessment, 1*(5). Online journal; article at http://www.bc.edu/research/intasc/jtla/journal/v1n5.shtml

Collins, A., & Ferguson, W. (1993). Epistemic forms and epistemic games: Structures and strategies to guide inquiry. *Educational Psychologist, 28,* 25-42.

Ebel, R. L. & Frisbie, D. A. (5th ed). (1991). *Essentials of Educational Measurement*. Englewood Cliffs, NJ: Prentice Hall.

Ericsson, K. A. (Ed.) (1996). *The road to excellence: The acquisition of expert performance in the arts and sciences, sports, and games.* Mahweh, NJ: Erlbaum.

Fuhrman, M. (1996). Developing good multiple choice tests and test questions. Journal of Geoscience Education, 44, 379-384.

Gamma, E., Helm, R., Johnson, R., & Vlissides, J. (1994). *Design Patterns: Elements of Reusable Objective-Oriented Software*. Addison-Wesley, Reading, MA.

Liu, M. & Haertel, G. D. (2011, April). Design patterns: A tool to support assessment task authoring. Paper presented at the annual meeting of the American Educational Research Association, New Orleans.

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

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

Mislevy, R. J., Behrens, J. T., Bennett, R. E., Demark, S. F., Frezzo, D. C., Levy, R., Robinson, D. H., Rutstein, D. W., Shute, V. J., Stanley, K., & Winters, F. I. (2010). On the roles of external knowledge representations in assessment design. *Journal of Technology, Learning, and Assessment, 8*(2). http://escholarship.bc.edu/jtla/vol8/2

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

Mislevy, R., Hamel, L., Fried, R., 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.

Mislevy, R. J., & Riconscente, M. M. (2006). Evidence-centered assessment design: Layers, concepts, and terminology. In S. Downing & T. Haladyna (Eds.), *Handbook of Test Development* (pp. 61-90). Mahwah, NJ: Erlbaum.

Mislevy, R., & Riconscente, M. (2005). *Evidence-centered assessment design: Layers, structures, and terminology (PADI Technical Report 9).* Menlo Park, CA: SRI International. http://padi.sri.com/downloads/TR9_ECD.pdf

Mislevy, R. J., Riconscente, M. M., & Rutstein, D. W. (2009). Design patterns for assessing model-based reasoning *(An Application of Evidence-Centered Design to a State's Large Scale Science Assessment. Technical Report 6).* 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.

Riconscente, M., Mislevy, R., Hamel, L., & PADI Research Group (2005). *An introduction to PADI task templates (PADI Technical Report 3).* Menlo Park, CA: SRI International.

Toulmin, S. E. (1958). *The Use of Argument*. Cambridge University Press, Cambridge, UK.

Appendix A: Exemplar Design Pattern and Development Specifications and Exemplar Task Template:

Data Analysis and Probability

Design Pattern
Data Analysis and Probability B1 (grades 3-5)

	Attribute	Definition	Design Pattern (DP)	Notes/Guidance
1	Title	Short name for the DP	Data Analysis and Probability B1 (grades 3-5)	
2	Summary	Brief description of the **family of tasks** implied by the DP	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	
3	Rationale	Nature of the KSAs of interest and why they are important	In grades 3-5, students make an important transition from examining individual pieces of data to understanding that data come in sets and data are distributed over some range. This leads to more sophisticated ideas in data analysis in math.	This expectation reflects extended standards for all 3 states as a critical element in the mathematics curriculum
4	Focal KSAs	The primary KSAs targeted by this DP	• Ability to describe the shape and important features of a data set showing understanding without necessarily using technical terminology (e.g., How many are in the set, which item/value is most frequent, which values are missing?) • Ability to compare two data sets using features of the set (e.g., Do the sets have the same number of "items"? Do they have the same mean?)	• Link to grade level academic content • Include variety in depth of knowledge so that all students are appropriately challenged • Do not include prerequisite KSAs • Note: While the extended content standards have been taken into account, the Focal KSAs have been selected to represent the content in the NCTM expectation being addressed. NCTM expectations represent the commonality between the extended standards of the 3 consortium states.
5	Additional KSAs	Other KSAs that may be required by tasks from this DP, some of which can be supported by Universal Design for Learning (UDL) and accommodations	**Cognitive Background Knowledge** • Knowledge of what data are (numbers that represent quantities that are qualities of objects or situations) • Ability to quantify • Ability to compare two or more things • Knowledge of the concepts "more" and "less" • Ability to read graphs, e.g., line graphs and bar graphs **Perceptual (Receptive)** • Ability to perceive the linguistic components of the question (e.g., through print, objects, audio, Braille/Nemeth code)	• May include prerequisite background knowledge (KSAs) • Additional KSAs organized by 6 UDL categories • Content related Additional KSAs are addressed in the Cognitive Background Knowledge category

	Attribute	Definition	Design Pattern (DP)	Notes/Guidance
			• Ability to perceive images in the question (e.g., through print, objects, holistic description, through Braille/Nemeth code description) • Ability to perceive physical objects required for the task (e.g., see hatch marks on ruler) **Skill and Fluency (Expressive)** • Ability to communicate response • Ability to compose or express a response in text e.g., by writing, using Braille/Nemeth code) • Ability to express a response verbally or by signing • Ability to manipulate physical materials (e.g., dexterity, strength and mobility) • Ability to manipulate digital/electronic equipment • Knowledge of how to use physical materials or digital/electronic equipment (e.g., familiarity) **Language and Symbols** • Ability to recognize text, symbols, or images • Ability to decode text, symbols, or images • Ability to comprehend text, symbols, or images • Ability to understand English vocabulary and syntax **Cognitive** • Ability to attend to stimuli • Ability to recall related background knowledge • Ability to perform (e.g., answer questions, solve simple problems) • Ability to provide an explanation • Ability to organize information • Ability to synthesize information • Ability to understand the meaning of an example • Ability to process multi-step problems • Ability to recall and use information presented in a task/item (working memory) • Ability to understand the structure of "organizers" used to present information or to scaffold responses (e.g., how to complete a table)	

	Attribute	Definition	Design Pattern (DP)	Notes/Guidance
			• Ability to understand the purpose of highlighted features in text or illustrations **<u>Executive</u>** • Ability to set goals and expectations • Ability to monitor goals and progress • Ability to plan and sequence • Ability to self-regulate and reflect during problem solving **<u>Affective</u>** • Ability to engage (e.g., task-specific motivation) • Ability to persist and sustain effort	
6	Potential Observations	Observed behaviors of students that can provide evidence of Focal KSAs	• Given a context that frames the interpretation of a data set and list of descriptions about the data, the student correctly chooses the most accurate description (e.g., "Most of the data are on the right side of the graph.") • Student accurately describes one feature of a data set (e.g., Which quantity occurs most frequently, the average quantity, the maximum and minimum values?) **NOTE: There was some discussion that this PO is more towards reading and interpreting the graph and this can serve as an Item 2 in the Item Template** • Given a context that frames the interpretation of two data sets and list of phrases that compare the two data sets, student correctly chooses the most accurate phrase • Student accurately compares one feature of two data sets (e.g., Which quantity occurs most frequently in each data set? The maximum and minimum values in each data set?) **NOTE: There was some discussion that this PO is more towards reading and interpreting the graph and this can serve as an Item 2 in the Item Template.**	• Each Potential Observation includes a qualifier (e.g., correctly, accurately, appropriately) that specifies the judgment about a behavior that will provide evidence about a student's knowledge, skill, or ability
7	Potential Work Products	What students say, do, or make that provides evidence about the Focal KSAs	Products may be written, verbal, demonstrations, and may be audio or video recorded or recorded by teacher, e.g.: • Selection from a list of phrases that describe the data • Description of data set	• Illustrate types of work products that could be gathered as part of the assessment to provide evidence of Focal KSAs • Create concrete examples of work products

	Attribute	Definition	Design Pattern (DP)	Notes/Guidance
				• **Menu** of options – not required work products (e.g., if user only uses paper/pencil tasks, why include multiple options for work products? We want users to think broadly; this document can be used if assessments change in the future) • Do not include qualifying words (e.g., appropriate)
8	Potential Rubrics	Some evaluation techniques that may apply	• Dichotomous (0,1) • Partial credit (0,1, 2, 3) • Consistency of proficiency (# times student gets correct response over multiple trials) • Scores based on independence of student's response	• Rubrics unique to states should be identified • Ways tasks may be scored • How to apply so assessment is rich, not confined
9	Characteristic Features	Aspects of assessment situations likely to evoke the desired evidence	• All tasks will include graph(s) • Graphs are accompanied by contextual information (a scenario) • Information presented in graph has to be complete (i.e., axis have to be labeled, title, sufficient number of data points are plotted) • Tasks are individually administered by a teacher or trained administrator • Accommodations allowed • Test administrator knows student and his/her comprehensive/response abilities • Periodic collection of work samples (for portfolios) • Word problems	• Features tasks must include to evoke the desired response • Consider cost/benefit of adding "story" information or authentic context to problems (increase relevance) vs. limiting extraneous information (minimize ambiguity and reduce cognitive load)
10	Variable Features	Aspects of assessment situations that can be varied in order to control difficulty or target emphasis on various KSAs	**Cognitive Background Knowledge** • Provide example of a data set and instructional review about how to interpret data in a graph • Modeling quantification • Remind student of materials or activities used to teach comparisons in math • Remind student of materials or activities used to teach the concepts of "more" and "less" in math as well as definition of the terms • Modeling reading a graph	• Special consideration required of the variable features of "story" problems. Adding story information can increase relevance but also adds ambiguity and increases cognitive load. • For multi-step problems, use of executive management supports will be essential.

	Attribute	Definition	Design Pattern (DP)	Notes/Guidance
			<u>Perceptual (Receptive)</u> • Delivery mechanisms by which the question is perceived (e.g., read aloud verbatim/read aloud paraphrase, pictures, large print, printed text, Braille/Nemeth code, signing, auditory amplification, symbols, concrete objects, description of objects or images, text to speech, CCTV – close circuit TV, to increase size of font, vary contrast, etc.) • Supports for the use of objects required for the task (e.g., speaking calculator, size of calculator, size of number line) **<u>Skill and Fluency (Expressive)</u>** • Response mode options (e.g., pointing, speech and verbalization, writing, Braille/Nemeth code, signing, switch or other assistive device/augmentative communication device, scanning software, eye gaze, for lowest functioning students – predictable behavioral response, tolerate assistance – e.g., hand over hand) • Supports for composing a response in text (e.g., written by student, speech to text, written by teacher, keyboarding) • Supports for manipulating physical materials (e.g., use of velcro, size of materials, teacher manipulation of materials) • Supporting for manipulating digital/electronic equipment (pointers, teacher manipulation of equipment, spoken commands, stylus for input, larger keyboard/buttons, adaptive mouse) • Practice tutorials with unfamiliar physical materials or digital/electronic equipment **<u>Language and Symbols</u>** • Level of abstraction required of student (e.g., concrete objects, images, text) • Embedded support for vocabulary, numbers, and symbols (e.g., technical glossary, hyperlinks/footnotes to definitions, illustrations, background knowledge, number line) • All key information in the dominant language (e.g.,	

	Attribute	Definition	Design Pattern (DP)	Notes/Guidance
			English) is also available in prevalent first languages (e.g., Spanish) for second language learners and in sign language for students who are deaf • Use of multiple representations (e.g., physical models, demonstrations, acting out scenarios) • Alternate syntactic levels (simplified text) • Highlight essential elements, words, or phrases • Digital text with automatic text to speech • Digital Braille with automatic Braille to speech **Cognitive** • Depth of knowledge of the content • Level of complexity of the content [e.g., use of calculation (e.g., median and mean) vs. description; use of technical terminology; number of data points represented in graph; values of data points (single vs. multiple digits); number of categories in graph; number of data features; types of data features (e.g., spread); number of graphs to compare; comparison of same or different types of graphs; types of distribution of data illustrated (e.g., 0-point, outliers)] • Prompts to explain sequential steps used to solve the problem • Item/task format (selected response vs. constructed response, performance, etc.) • Adjustable levels of challenge (teacher able to adjust) • Options for supporting background knowledge: o Pre-teach background content o Provide analogies and examples o Provide hyperlinks to multi-media o Provide links to related information o Provide links to familiar materials o Provide concept maps o Remind student of prior experiences o Remind student of materials or activities used to teach foundational math skills • Options for supporting critical features, big ideas, and relations:	

	Attribute	Definition	Design Pattern (DP)	Notes/Guidance
			o Provide graphic organizers o Outline information o Highlight information o Provide alternative forms of key concepts o Provide multi-media glossaries o Provide translation tools o Provide modeled prompts o Provide a response template • Options for guiding exploration and information processing: o Provide multiple entry points o Allow viewing of stimuli from previous stages and parts o Use familiar materials o Use consistent signals/cues o Provide sequential highlighting o Chunk information into smaller elements o Mask part of the information o Mask incorrect answer options o Provide modeled prompts o Provide a practice item or task o Provide a calculator o Provide a number line o Provide an abacus • Options for supporting memory and transfer: o Note-taking o Mnemonic aids o Locate items near relevant text o Reread question o Present items as a discrete unit or embed in a scenario **<u>Executive</u>** • Prompts and scaffolds to estimate effort, resources, and difficulty • Prompts, scaffolds, and questions to monitor progress, to “stop and think”, and for categorizing and systematizing • Representations of progress (e.g., before and after photos, graphs and charts) • Guides, checklists, graphic organizers, and/or templates for goal setting, prioritizing, breaking long-term objectives	

	Attribute	Definition	Design Pattern (DP)	Notes/Guidance

			into reachable short-term goals, self-reflection, and self-assessment
			• Adjust levels of challenge and support (e.g., adjustable leveling and embedded support, alternative levels of difficulty, alternative points of entry)
			
			<u>Affective</u>
			• Teacher options for providing supports for attention and engagement:
			  • Cover up part of text so student isn't overwhelmed
			  • Prompt student to re-engage
			  • Provide verbal/gestural prompts
			  • Provide feedback to support engagement
			  • Provide supports to reduce student frustration (e.g., noise reduction, extended test taking time, contingencies, number of items administered at one time)
			  • Provide varied levels of challenge and support
			  • Provide optimal student positioning (positions which encourage alertness, not recumbent)
			  • Administer assessment at optimal time of day for student engagement
			• Task options for engagement:
			  • Provide students with choices for personal control of age-appropriate content when construct is not impacted (e.g., choice of topic or theme)
			  • Provide students with choices for personal control of task context when construct is not impacted
			  • Enhance relevance, value, and authenticity of tasks
			  • Heighten salience
			  • Variety of stimuli
			  • Vary amount of context supporting tasks (e.g., discrete tasks vs. scenarios)
			  • Item/task format (selected response vs. constructed response, performance, etc.)	

11	Educational Standards	State extended math standards related to selected	**<u>Florida:</u>** Grade 3:NA Grade 4: NA	

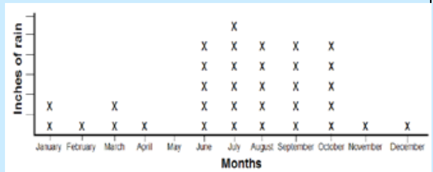
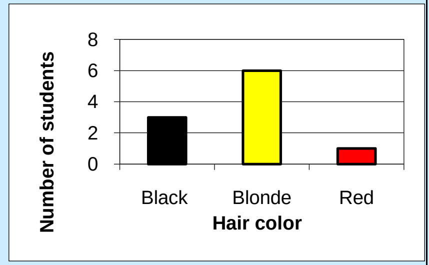
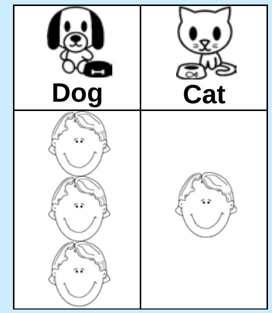
	Attribute	Definition	Design Pattern (DP)	Notes/Guidance
		NCTM expectation	Grade 5: S.7.In.b. Describe the meaning of data in a three-category pictograph or bar graph. S.7.Su.b. Identify the meaning of data in a two-category object graph or pictograph. **Idaho:** Grade 3: NA Grade 4: NA Grade 5: 5.3.2 A Find the end points of the range of a set of data using whole numbers 1-10. **Utah:** Grade 3: NA Grade 4: NA Grade 5: Va. Identify the minimum and maximum value in a set of data.	

Task/Item Development
Data Analysis and Probability B1 (grades 3-5)

Attributes	General Information
Title	Data Analysis and Probability B1 (grades 3-5)
Summary	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
Rationale	In grades 3-5, students make an important transition from examining individual pieces of data to understanding that data come in sets and data are distributed over some range. This leads to more sophisticated ideas in data analysis in math.
Grade level standards (from NCTM)	A reasonable objective for upper elementary and middle-grades students is that they begin to regard a set of data as a whole that can be described as a set and compared to other data sets (Konold forthcoming). As students examine a set of ordered numerical. As students examine a set of ordered numerical data, teachers should help them learn to pay attention to important characteristics of the data set: where data are concentrated or clumped, values for which there are no data, or data points that appear to have unusual values. Much of students' work with data in grades 3–5 should involve comparing related data sets. Noting the similarities and differences between two data sets requires students to become more precise in their descriptions of the data. In this context, students gradually develop the idea of a "typical," or average, value. Building on their informal understanding of "the most" and "the middle," students can learn about three measures of center—mode, median, and, informally, the mean. Students need to learn more than simply how to identify the mode or median in a data set. They need to build an understanding of what, for example, the median tells them about the data, and they need to see this value in the context of other characteristics of the data.

		Item 1	Item 2	Item 3a/3b
Attributes	Definition	Application/Comprehension/ Performance	Performance/Recall	Recall/Attention
Focal KSA	Focal KSA from DP for Items 1 & 2; Add'l KSA from DP for Item 3	• Ability to describe the shape and important features of a data set showing understanding without necessarily using technical terminology (e.g., how many are in the set, which item/value is most frequent, which values are missing)		• Ability to read graphs, e.g., line graphs and bar graphs
Potential Observations from DP	Observed behaviors of students that can provide evidence of the Focal KSA	• Student accurately describes one feature of a data set (e.g., which quantity occurs most frequently, the average quantity, the maximum and minimum values) NOTE: There was some discussion that this PO is more towards reading and interpreting the graph and this can serve as an Item 2 in the Item Template		*Not addressed in DP*
Potential Work Products	What students say, do, or make that provides evidence about the Focal KSA	• Selection from a list of phrases that describe the data • Description of data set		*Not addressed in DP*
Characteristic Features	Aspects of assessment situations likely to evoke the desired evidence	• All tasks will include graph(s) • Graphs are accompanied by contextual information (a scenario) Information presented in graph has to be complete (i.e., axis have to be labeled, title, sufficient number of data points are plotted)		

Potential Variable Features/ Scaffolding	Features that could be changed to impact item difficulty	• DOK level • Type of item response format (Constructed response vs. selected response) • Use of technical terminology • Number of data points represented in graph • Number of digits represented in data points (single vs. multiple digits) • Number of categories on x-axis in graph • Number of data features (an item can ask for more than one important feature of the data) • Types of data features targeted (e.g., spread, outlier, clumps) • Type of data features targeted • Type of data representation • Number of data features in the graph • Number of data features targeted		
Selected Variable Features/ Scaffolding for the Item	From Item 1 to Item 3: • Reduce DOK • Reduce scope • Increase scaffolding	• DOK level: Comprehension (Understand; Group) • Type of item response format: Constructed response • Use of technical terminology: No • Number of data points represented in graph: 34 • Number of digits represented in data points: N/A • Number of categories on x-axis in graph: 12 months • Number of data features in the graph: 1 (clump) • Number of data features targeted: 1 • Type of data features targeted: Clumps • Type of data representation: line graph	• DOK level: Performance (Read) • Type of item response format: Constructed response • Use of technical terminology: No • Number of data points represented in graph: 10 • Number of digits represented in data points: Single • Number of categories on x-axis in graph: 3 hair colors • Number of data features in the graph: 1 (frequency) • Number of data features targeted: 1 • Type of data features targeted: Frequency • Type of data representation: bar graph	• DOK level: Recall (Match, Recognize, Identify) • Type of item response format: Selected response • Use of technical terminology: No • Number of data points represented in graph: 4 • Number of digits represented in data points: Single • Number of categories in graph: 2 • Number of data features in the graph: 1 (frequency) • Number of data features targeted: 1 • Types of data features targeted: Frequency • Type of data representation: pictograph
Item Directive	The stem or question (includes description and number of distractors if applicable)	Examiner presents a line plot about the inches of rain that fell each month during the year and says, **"Rainy season occurs during months of the year that have the most rain. Which month do you think the rainy season starts? Which month do you think the rainy season ends?"**	Examiner presents a bar graph about the number of students in a class with different color hair. The examiner points to the graph and says, "**In this graph, we see the number of students that have each hair color. What is the hair color of most students?"**	3a. Examiner presents a pictograph about the number of dogs and cats that students have as pets. Examiner says, **"Here is a pictograph that shows the number of students who have dogs and the number of students who have cats. How many students have dogs?** Examiner presents each response on its own card and says, **"Do three students have dogs** [presents card with three students] **or does one student have a dog** [presents card with one student]**?"**

				3b. Examiner covers column that shows number of students who has a cat, and says, **“Look at/touch the students who have dogs.”**
Correct Answer	Correct answer for the item	Students indicates June through October	Students indicates blonde hair	3a. Student indicates 3 3b. Student looks/touches pictograph
Description of Stimulus Items	Description of the *graphics or objects* used in administration of the task	• A line plot about the inches of rain that fell each month during the year  	• Bar graph about the number of students in a class with different color hair.  	• Pictograph about the number of dogs and cats that students have as pets  • Two pictures cards, each response on its own card 
Materials for <u>Examiner</u>	Materials required to administer, document, and score the task (e.g., worksheet, camera to take picture of product, manipulatives)	Task worksheet that describes the item and delivery instructions for examiner. Task sheet or other method to record student response for scoring.	Task worksheet that describes the item and delivery instructions for examiner. Task sheet or other method to record student response for scoring.	Task worksheet that describes the item and delivery instructions for examiner. Task sheet or other method to record student response for scoring.

Variable Features for Administra-tion to Individual Students	Features that could be changed to impact item accessibility for individual student needs (e.g., as specified in the student's IEP)	• Question presentation individualized (e.g., related in sign language; verbal/gestural prompts individualized) • Response format individualized based on student communication system	• Question presentation individualized (e.g., related in sign language; verbal/gestural prompts individualized) • Response format individualized based on student communication system	• Question presentation individualized (e.g., related in sign language; verbal/gestural prompts individualized) • Response format individualized based on student communication system

Updated Flowers/Browder Math DOK[1]:

1. **Attention:** touch, look, listen, repeat what the teacher said, vocalize, respond, attend, recognize
2. **Memorize/recall:** list, describe (facts), state math facts, identify, state, define, match, recognize, label, follow a pattern
3. **Performance:** answer, follow 1 step directions, find answer, present, read, separate, spell, tell time, map, model demonstration, perform, demonstrate, follow, choose, count, locate, group by given attributes, solve simple (one computation skill) problems, measure
4. **Comprehension:** understand, extend a pattern, sketch, ask and answer questions, categorize/group by unknown attributes, explain, conclude, group, restate, review, translate, classify/sort with understanding, simplify (equivalent forms)
5. **Application:** compute, organize, collect (such as data), apply, revise, construct, solve complex (multiple computation skills) problems, use given formulas in novel situations (formula may or may not be identified), explain a process, conduct research
6. **Analysis, Synthesis, Evaluation:** create a complex pattern, analyze, compare, contrast, compose, predict, plan, judge, evaluate, interpret data, generalize findings, create hypotheses

[1] Bechard, S., Almond, P., Karvonen, M., Wakeman, S., Turner, C., Bowen, T., & Turner, L. (2009). *Hitting a moving target: A discussion of ten alignment studies for AA-AAS.* Paper presented at the National Conference on Student Assessment. Los Angeles, CA June 23, 2009.