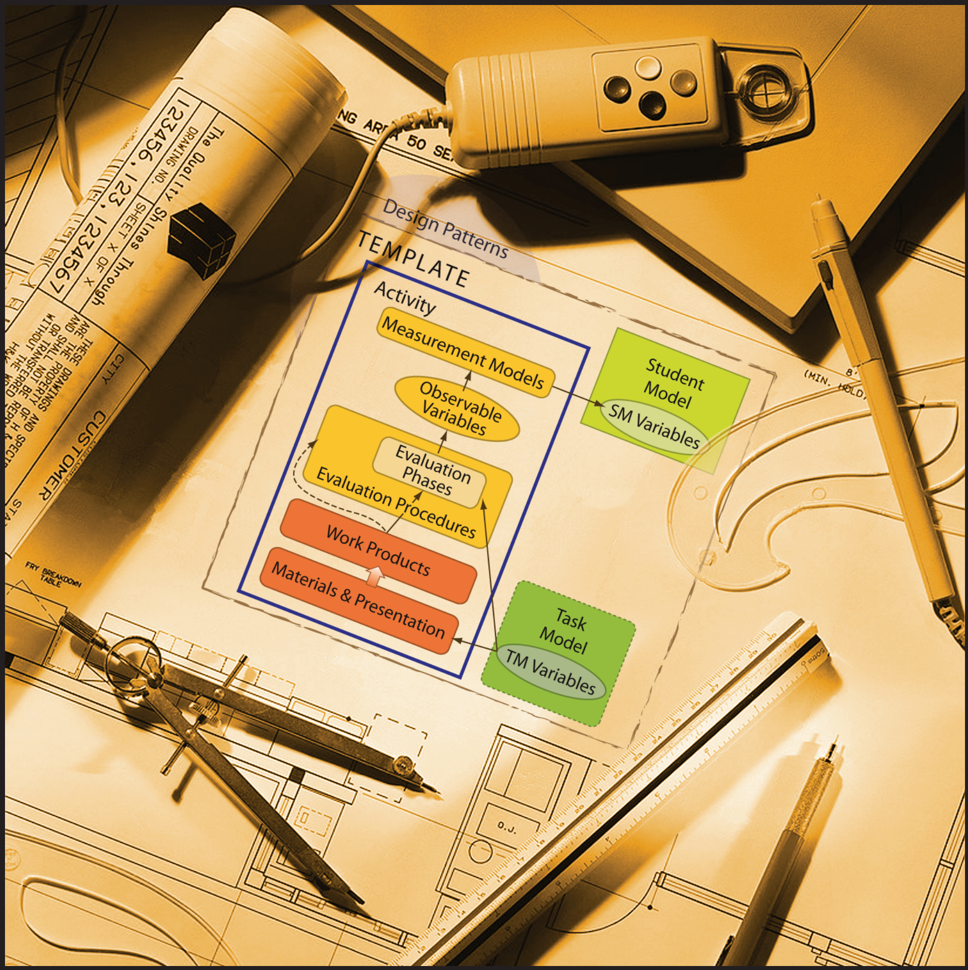


Large-Scale Assessment Technical Report 12|
August 2013

Using the Principled Assessment Design in Inquiry (PADI) System: Some Frequently Asked Questions

Project: Application of Evidence-Centered Design to State Large-Scale Science Assessment

Terry Vendlinski, Geneva Haertel, Britte Chang, Angela DeBarger, Daisy Rutstein, Ron Fried, Eric Snow and Dan Zalles, SRI International
Robert Mislevy, Educational Testing Service
Youngmi Cho, University of Maryland
Dennis Fulkerson and Kevin McCarthy, Pearson Learning
Doreen Finkelstein, Haynie Research and Evaluation

Report Series Published by SRI International

SRI International
Center for Technology in Learning
333 Ravenswood Avenue
Menlo Park, CA 94025-3493
650.859.2000
http://ecd.sri.com

Technical Report Series Editors
Alexis Mitman Colker, Ph.D., *Project Consultant*
Geneva D. Haertel, Ph.D., *Principal Investigator*
Robert Mislevy, Ph.D., *Co-Principal Investigator*
Ron Fried, *Documentation Designer*

D E S I G N T O S T A T E L A R G E - S C A L E S C I E N C E A S S E S S M E N T

T E C H N I C A L R E P O R T 1 2

Using the Principled Assessment Design in Inquiry (PADI) System: Some Frequently Asked Questions

August 2013

Prepared by:

Terry Vendlinski, Geneva Haertel, Britte Cheng, Angela DeBarger, Daisy Rutstein, Ron Fried, Eric Snow and Dan Zalles – SRI International

Robert Mislevy – Educational Testing Service

Youngmi Cho – University of Maryland

Dennis Fulkerson and Kevin McCarthy – Pearson Learning

Doreen Finkelstein – Haynie Research and Evaluation

Acknowledgments

This material is based on work supported by the National Science Foundation under grant DRL-0733172 (An Application of Evidence-Centered Design to State Large-Scale Science Assessment).

Disclaimer

Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

CONTENTS

FIGURES

List Of Tables

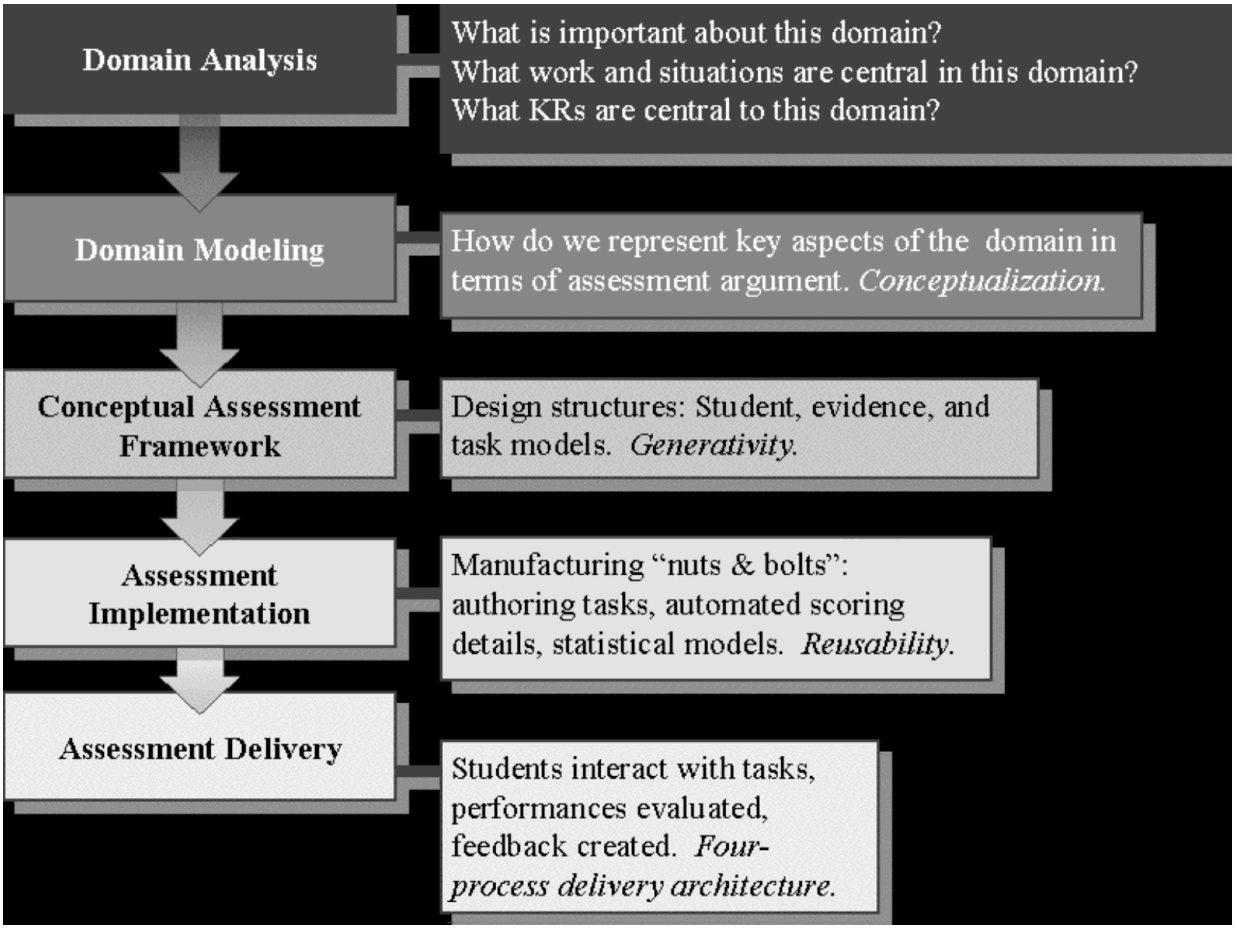
Background

Evidence Centered Design (ECD) arises from Messick's (1994) contention that good assessment design should consider: 1) the knowledge, skills and abilities (KSAs) one desires to assess; 2) the behaviors or work that would be indicative of those knowledge, skills and abilities; and, 3) the situations likely to elicit those behaviors. To move the evidentiary argument from situations to inferences, one must consider how to interpret the data collected in order to make inferences about a student's knowledge, skills and abilities. ECD supports and documents this chain of reasoning, from claims through evidence to interpretation.

The Principled Assessment Design in Inquiry (PADI) system was designed to instantiate the ECD process, to document decisions made while designing assessments using the ECD process, and to make the ECD process accessible to both professional and more novice assessment designers (Mislevy & Haertel, 2006). The ECD model in PADI uses five layers to accomplish these objectives (see Figure 1). Users of the PADI system begin the assessment development process by first analyzing a domain of interest. They document important knowledge, practices and ways of communicating within the domain. After analyzing a domain in detail, designers use this information to specify important attributes about the student, the evidence, and the task models in the PADI system.

The PADI system uses forms called *Design Patterns*, to help task developers think thorough these important attributes in targeted skill areas—particularly ones that are hard to assess, like inquiry and investigation in science. Design Patterns sketch out a design space for such tasks, drawing on experience and research in the domain. They encourage assessment designers to lay out the components of the assessment argument, including the integration of the three models in a narrative manner. Design Patterns also support incorporating Universal Design for Learning (UDL) features that can increase accessibility for special needs students (Rose & Meyer, 2002). In many cases, increasing such accessibility for these students makes the assessments developed more accessible for all students. This Technical Report focuses on the use of Design Patterns within PADI. The PADI system also provides other tools to promote good assessment design (e.g., Task Templates, Conceptual Assessment Framework, Item Writers' Guide, and Template Series Representation).

Figure 1. The Five Layers of ECD



Questions Frequently Asked by New PADI Users and the Answers to Those Questions

This technical report focuses on PADI Design Patterns, but some broader discussion of ECD will be useful to provide context. Its objectives are: 1) to document and answer Frequently Asked Questions (FAQs) about ECD in PADI; 2) to explain how the PADI system is used; and, 3) to address questions about Design Patterns and their use when developing assessments. Table 1 lists the questions addressed here.

Table 1. Frequently Asked Questions (FAQs) while developing assessments in the PADI system.

1. How does PADI implement Evidence Centered Design Principles?
2. What is the quickest way to start using the PADI System? (Quick start guide)
3. What are Design Patterns?
4. What does a Design Pattern look like?
5. What are the attributes of a Design Pattern?
6. What are the differences between Task Specifications and Design Patterns?
7. Why use a Design Pattern?
8. How do Design Patterns integrate the Student, Evidence, and Task models?
9. How have professional item development teams learned to use the PADI System? (An 11 step training process)
10. How can Design Patterns be used to identify construct-irrelevant variance?
11. How can Design Patterns be used to design assessments that measure hard to measure constructs?
12. How can Variable Features be used when designing assessments?
13. How can Design Patterns be used to integrate ECD and Universal Design for Learning?
14. How can Design Patterns be used to ensure the coherence of the assessment argument?
15. How do Design Patterns support the integration of content and practice in a domain?
16. How can Design Patterns be used to support the assessment of a science practice when the science content is not explicit?
17. How can Design Patterns incorporate learning progressions into the assessment design process?
18. How can multiple Design Patterns be used to support scenario development?
19. What is a Library of Design Patterns?
20. How can Design Patterns support the development of scenario-based assessments?
21. How can Design Patterns and Standards be associated? (Using Activation Charts)
22. How can assessment items or tasks be reverse engineered to create a Design Pattern?

Each of the questions in Table 1 is addressed in individual sections of this technical report below. FAQs are addressed in the order they appear in the table.

1. How does PADI implement Evidence Centered Design Principles?

Principled Assessments Designs in Inquiry (PADI) is an online assessment design system that applies the principles of Evidence Centered Design (ECD) as described by Almond, Steinberg, and Mislevy (2002) and Mislevy, Steinberg, and Almond (2003). Development on PADI was initiated in 2001 / 2002 and the software tool was fully operational in 2007. PADI programmers developed a software application to assist in the design and development of assessment tasks from reusable components. The practice of reusing components helps ensure consistency throughout an assessment system and also speeds up the development process.

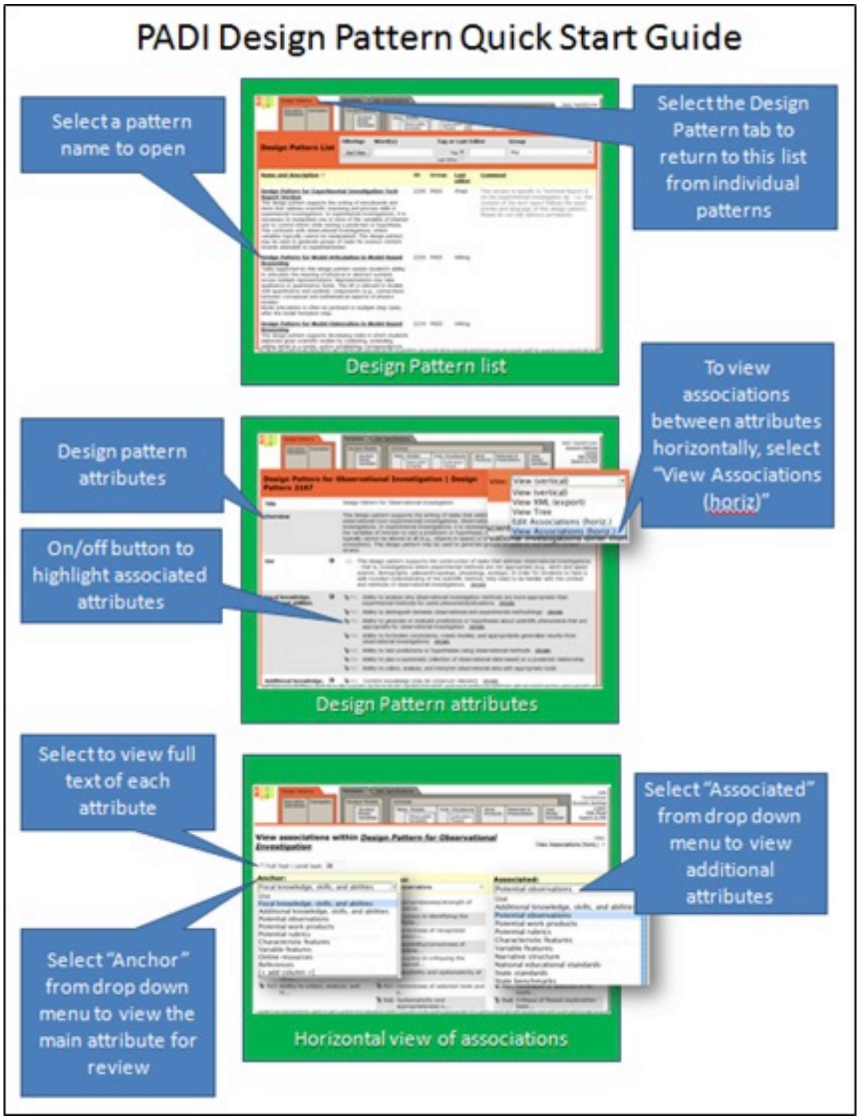
The PADI online system includes tools that support design and documentation of assessments and analysis of assessment data. These tools include: *Design Patterns*, *Task Templates*, and a *Scoring Engine*. Design Patterns capture assessment arguments describing how the alignment of cognitive objectives for performance, observations, and interpretation are operationalized for a specific (narrow or broad) domain of knowledge. Design Patterns, which reside in the Domain Modeling layer of ECD and at the highest level of conceptualization in the PADI design system, are non-technical in nature. Design Patterns are the initial template encountered in the PADI system and foreshadow the specification of the student, evidence, and task models required by assessment designers who are implementing Task Templates. Task Templates are a second layer of specifications that provide the "nuts and bolts" of the technical details required to design an assessment. These details include: rubrics; measurement models; and statistical interdependencies among items and tasks. Task templates are also used to specify relationships among the student model variables and to specify the stimulus materials, presentation logic, and item formats in the task models. A scoring engine, developed by the BEAR Center at UC Berkley, is integrated in the PADI design system. The scoring engine takes advantage of advances in educational measurement by anticipating the need for multidimensional item response modeling (IRM) to draw inferences from the evidence generated from student responses. The use of multidimensional IRM can enhance the interpretability of assessment evidence by relating it to multiple learning goals. It also can improve the reliability and validity of comparisons made over time and among student groups, particularly when students do not complete the same assessment tasks, through the use of consistent scaling at the task level (Rasch, 1960; Wilson, 2005; Wright, 1993).

The framework and tools developed in PADI can be applied to assessment in any subject area, grade level, or for any population (e.g. students with disabilities, English Language Learners, etc.).

2. What is the quickest way to start using the PADI System? (Quick start guide)

The PADI Design System consists of many different components, and can be confusing to users not familiar with the various affordances of the system. The entry point for most users is the Design Pattern. Figure 2 shows a Quick Start Guide to using and viewing Design Patterns in the PADI System. Each of the call-outs in Figure 2 is explained in greater detail below.

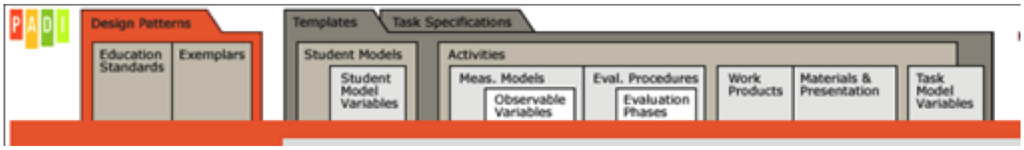
Figure 2. Quick Start Guide



The PADI website is accessible from the link http://ecd.sri.com/. Start viewing sample Design Patterns by clicking on the words "design patterns" near the top right of the screen under the heading "Example Design Patterns".

The PADI Design System has three main tabs; Design Patterns, Templates and Task Specifications (shown in Figure 3).

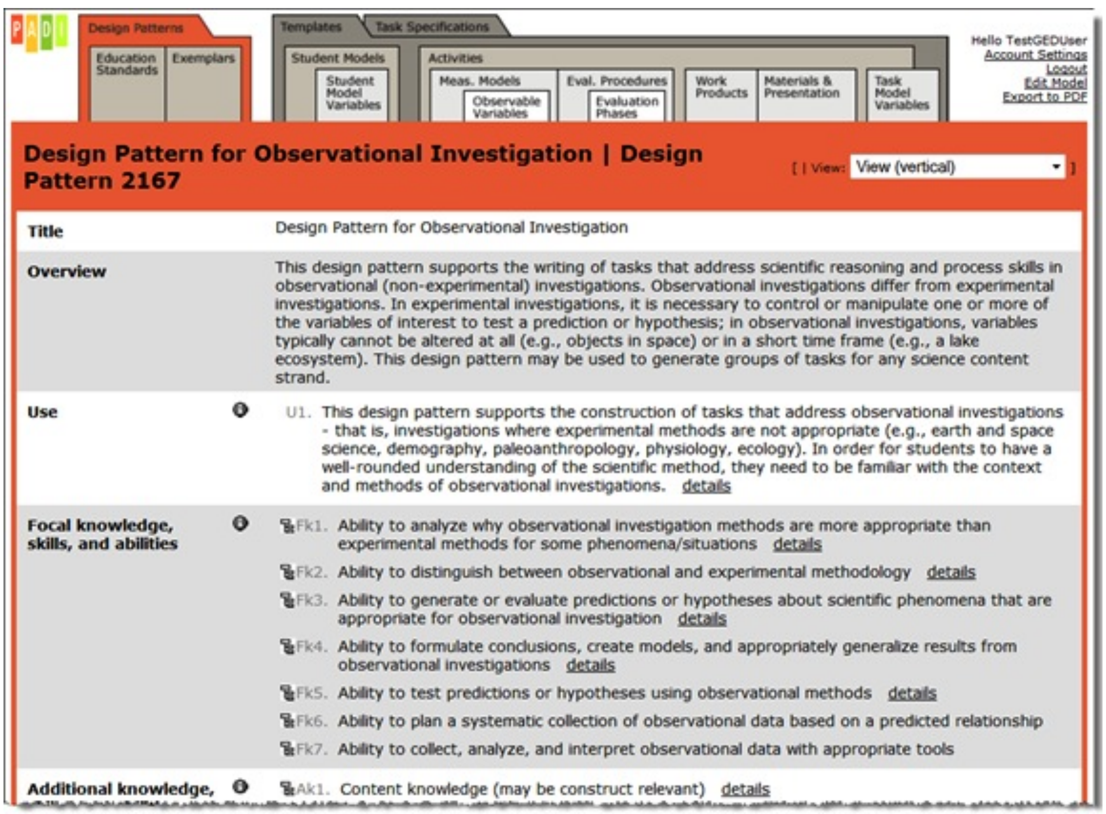
Figure 3. PADI Design System Tabs



Click on the Design Patterns tab to view the list of available Design Patterns. Clicking on this tab will always return a user to the main Design Pattern list.

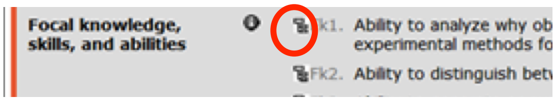
Once the design pattern list is open, a user can select the desired design pattern by clicking on its name. The pattern will open and display its attributes (Title, Overview, Use, Focal Knowledge, Skills, and Abilities, etc.) on the left and the attribute descriptions on the right (see Figure 4). These attributes are explained more fully in Question 0 of this Technical Report.

Figure 4. A Design Pattern showing Design Pattern Attributes on the left side of the frame.



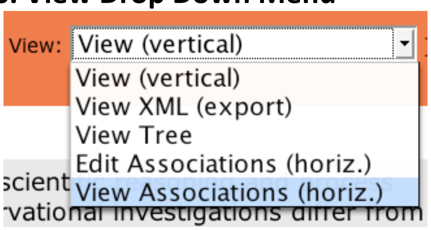
Several design pattern attributes have an icon that appears to the left of each instance of the attribute (e.g., each instance of a Focal Knowledge, Skill and Ability in Figure 5). Clicking once on this icon highlights the instances of other attributes in this design pattern that are associated with one another. For example, clicking on the icon for Focal KSA 1 will make the Potential Observations and Potential Work Products linked to Focal KSA 1 appear in bold font. A user can scroll through the Design Pattern to see the entries that are associated with each other. The associated instances of each attribute will appear in bold font. Clicking this icon again turns off the association function.

Figure 5. On/Off Icon



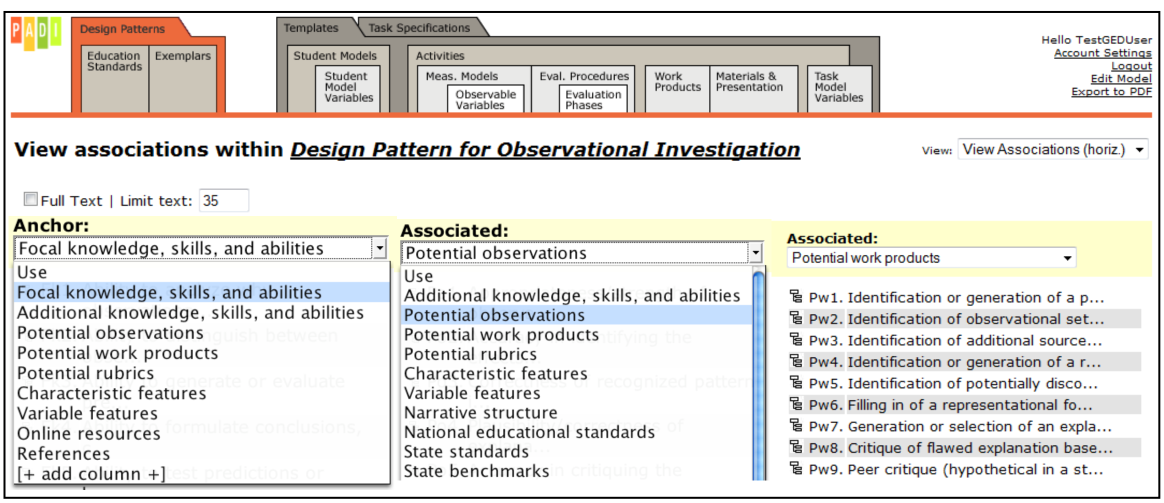
It is sometimes helpful to view associations among attributes side-by-side. This relieves the user of having to scroll up and down the web page. A user can retrieve the side-by-side (horizontal) view by clicking on the “View” drop down menu at the top right of the screen and selecting “View Associations (horiz.)” from the dropdown menu (Figure 6).

Figure 6. View Drop Down Menu



Once in this horizontal view, a user can select the attributes to review by first selecting the “Anchor” attribute (the main attribute for review), and then selecting “Associated” (the additional attributes to view). Figure 7 shows the interface through which a user can select the various Anchor and Associated attributes.

Figure 7. Horizontal View with Anchor and Associated Drop Down Menus



A user can elect to see the full text of each selected attribute by clicking on the “Full Text” box above the “Anchor” column.

3. What are Design Patterns?

A design pattern is a formal representation that addresses both a recurring design problem and the core of the solution to that problem in a particular field of expertise. The design pattern was first introduced in architecture (Alexander, Ishikawa, & Silverstein, 1977) and has been widely adapted in software engineering (e.g., Gamma, Helm, Johnson, & Vlissides, 1994) because of its advantages of reusability and flexibility. A design pattern can be applied repeatedly to resolve a problem in different contexts even though the particulars of the context may change.

The idea of an assessment design pattern was adopted by Mislevy et al. (2003) in the PADI project because this project aimed to provide a practical, theory-based approach to developing high-quality assessments of science inquiry. Designing high-quality assessments of science inquiry has been a challenge largely because it requires the coordination of expertise in different domains. In particular, constructing science inquiry items with acceptable inferential validity requires the input of science content specialists, assessment designers, and science classroom teachers. This challenge has been tackled by introducing Design Patterns into the assessment design process. In PADI, a Design Pattern is used as a schema or structure for conceptualizing the components of assessment arguments and their interrelationships. Thus, the Design Pattern plays an important role in bridging the expertise of content and measurement specialists so that they can communicate their knowledge effectively. It also guides assessment designers to think through the essential elements of assessment in ways that lead to a coherent assessment argument, expressed in a narrative form. The outcome of using Design Patterns is to present the knowledge of these experts in a more systematic and fully developed product.

The benefits of using a Design Pattern as an assessment approach fall into three general categories. First, Design Patterns facilitate decision-making about assessment design. The assessment design process is complex. In order to maximize efficiency and minimize the possibility of invalid inferences, assessment designers can either create a Design Pattern to guide the development of a new assessment for their purposes, or use an existing Design Pattern. Design Patterns are useful in that they provide a framework for documenting the decisions made during assessment design. Once made and documented in the Design Pattern, these assessment decisions provide a pathway through the assessment design process that can be further refined as needed. Second, Design Patterns help explicate the details of the assessment argument. They encourage careful consideration, expressed in a narrative form, of the features of the items to be constructed, which skills the items will assess, and the evidence that will be provided to indicate whether those skills have been attained. Third, Design Patterns offer flexibility. They can be built to address a range of psychological perspectives on learning, in a variety of content areas and at various grade levels and for various purposes. Design Patterns also to vary in their generality and scale. The flexibility of Design Patterns allows them to be easily extended to other content areas (see Wei et al, 2008, for an application of Design Pattern in language assessment).

4. What does a Design Pattern look like?

A Design Pattern specifies the attributes of an assessment argument as they are instantiated in Evidence Centered Design (ECD). Design Patterns help item and task writers move from the propositional knowledge of a domain to a high level conceptualization of the components of an assessment task.

A Design Pattern consists of attributes that can be associated with components of an assessment argument. Examples of attributes include Knowledge, Skills, or Abilities (KSAs) about which assessors want to make a claim, the kinds of data that provide evidence about the degree of student attainment of that KSA (Potential Observations), and features of tasks that can enable students to produce that evidence (characteristic and task variables). Figure 8 is a screen shot of a Design Pattern for experimental investigations in the sciences. Design Pattern attributes are listed on the left side. The right side contains the instances of each attribute.

Figure 8. An example of a Design Pattern

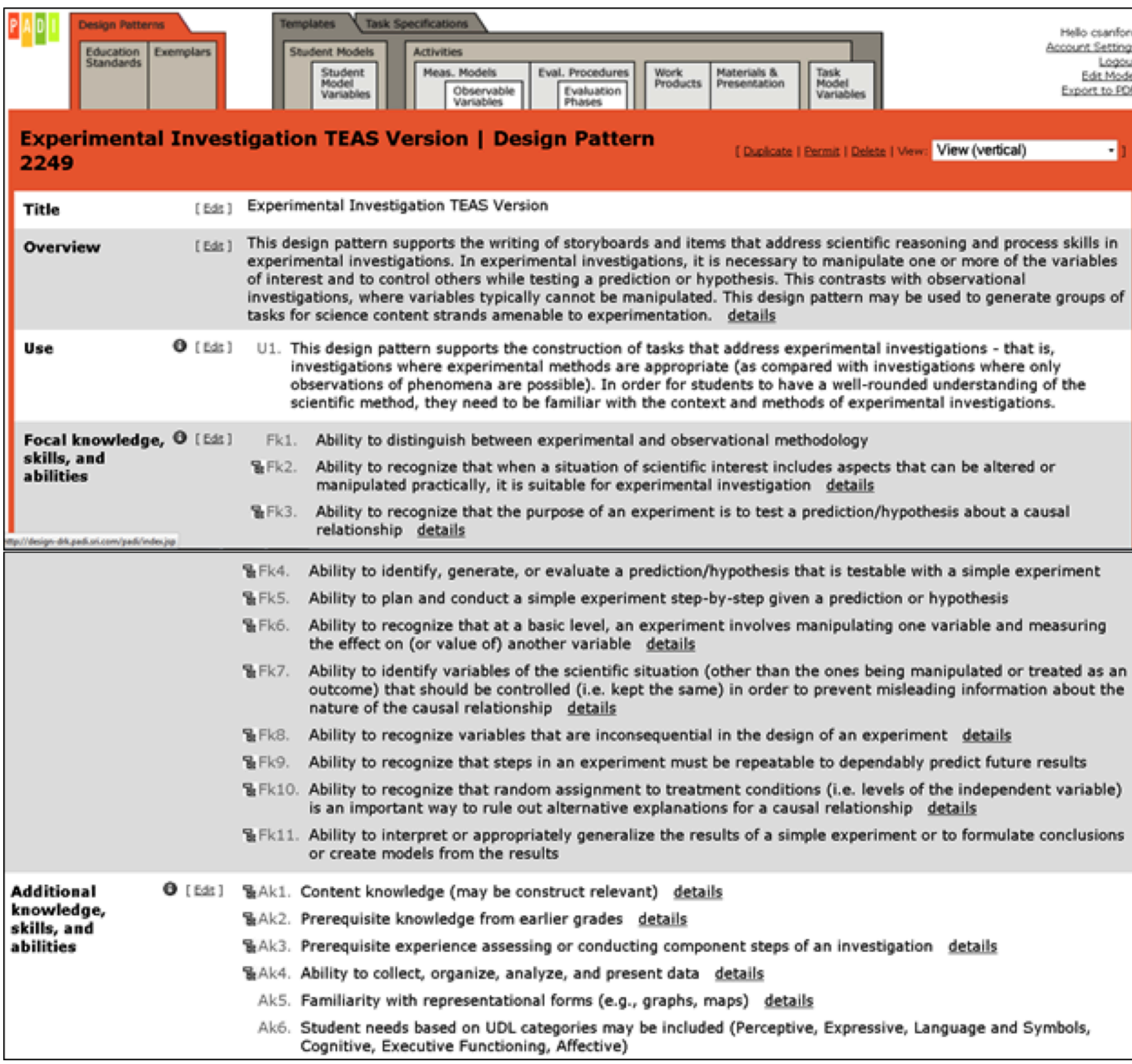


Figure 8 – Continued . An Example of a Design Pattern

Potential observations	[Edit]	Po1.	Accuracy in identifying situation suitable for experimental investigation
		Po2.	Plausibility of a measurable research question being raised
		Po3.	Plausibility of hypothesis as being testable by a simple experiment
		Po4.	Plausibility/correctness of design for a simple experiment
		Po5.	Correct identification of independent and dependent variables
		Po6.	Accuracy in identifying variables (other than the treatment variables of interest) that should be controlled (held constant) or made equivalent (e.g., through random assignment).
		Po7.	Plausability/correctness of steps to take in the conduct of an experiment
		Po8.	Plausibility of plan for repeating an experiment
		Po9.	Correctness of recognized data patterns from experimental data
		Po10.	Plausibility/correctness of interpretation/explanation of experimental results
		Po11.	Accuracy in critiquing the experimental design, methods, results, and conclusions of others
		Po12.	Generate a prediction/hypothesis that is testable with a simple experiment
Potential work products	[Edit]	Pw1.	Select, identify, or evaluate an investigable question details
		Pw2.	Identify or differentiate independent and dependent variables in a given scientific situation
		Pw3.	Identify or differentiate variables that do and do not need to be controlled in a given scientific situation
		Pw4.	Complete some phases of experimentation with given information, such as selection levels or determining steps
		Pw5.	Generate or identify data pattern from results in a simple experiment
		Pw6.	Generate an interpretation/explanation/conclusion from a set of experimental results
		Pw7.	Critiques of peers on their choice of experimental procedures or explanations of experimental results details
		Pw8.	Given an experiment with unexpected or confusing results, identify possible reasons details
Potential rubrics	[Edit]		
Characteristic features	[Edit]	Cf1.	Focus on Nature of Science (Strand I in MCA) benchmarks that relate to experimental investigations at the appropriate grade level
		Cf2.	Presentation of situation of scientific interest where variables can be (or have been) practically altered to address a causal prediction details
		Cf3.	Presentation of situation requiring the design or conduct of a controlled experiment details
		Cf4.	Presentation or representation of an experimental design
		Cf5.	Presentation of observed result from an experiment requiring the development of explanations, conclusions, or models details
Variable features	[Edit]	Vf1.	Content (strand) context details
		Vf2.	Which one of multiple phases of experimental investigation will be addressed
		Vf3.	Qualitative or quantitative investigation or a combination
		Vf4.	Ease or difficulty with which the treatment (independent) variable can be manipulated
		Vf5.	Are manipulated variables given or to be determined?
		Vf6.	The number of variables investigated and the complexity of their interrelationships details
		Vf7.	Number of variables that need to be controlled to unambiguously study the relationship between the manipulated variable and the outcome variable details
		Vf8.	Length of time over which the experiment must be conducted in order to study the potential impact of the treatment variable
		Vf9.	Data representations details
		Vf10.	Variable features may be added to support student needs associated with UDL categories (Perceptual - Screen presentation will include variable font size, Option for altering screen contrast, Option for magnification or zoom, Optional text-to-speech; Expressive - Range of response options required (radio buttons, drag and drop), Range of student support for producing response (speech-to-text); Language and Symbols - Provision of multiple representations of symbols (linguistic labels for symbols, define abbreviations, etc.), Provide definitions of non-construct relevant terminology, Use of studentsâ?? dominant language; Cognitive - Use of a concept map, Use of a response template, Use of context to heighten salience, Highlighting key terms and ideas; Executive Functioning - Breaking task into manageable units, Icons to encourage thinking and reflection, On-screen progress monitoring; Affective - Use of scenario or real-world context to heighten engagement, Age-appropriate materials -Interactive narrative (gaming), Affirmation of participation

Figure 8 – Continued. An Example of a Design Pattern

Narrative structure	[Edit]	Cause and effect. An event, phenomenon, or system is altered by internal or external factors. Investigation. A student or scientist completes an investigation in which one or more variables may be observed or manipulated and data are collected Change over time. A sequence of events is presented to highlight sequential or cyclical change in a system.
National educational standards	[Edit]	NSES 8ASI1.1. Identify questions that can be answered through scientific investigations. Students should develop the ability to refine and refocus broad and ill-defined questions. An important aspect of this ability consists of students' ability to clarify questions and inquiries and direct them toward objects and phenomena that can be described, explained, or predicted by scientific investigations. Students should develop the ability to identify their questions with scientific ideas, concepts, and quantitative relationships that guide investigation. NSES 8ASI1.2. Design and conduct a scientific investigation. Students should develop general abilities, such as systematic observation, making accurate measurements, and identifying and controlling variables. They should also develop the ability to clarify their ideas that are influencing and guiding the inquiry, and to understand how those ideas compare with current scientific knowledge. Students can learn to formulate questions, design investigations, execute investigations, interpret data, use evidence to generate explanations, propose alternative explanations, and critique explanations and procedures. NSES 8ASI1.3. Use appropriate tools and techniques to gather, analyze, and interpret data. The use of tools and techniques, including mathematics, will be guided by the question asked and the investigations students design. The use of computers for the collection, summary, and display of evidence is part of this standard. Students should be able to access, gather, store, retrieve, and organize data, using hardware and software designed for these purposes. NSES 8ASI1.4. Develop descriptions, explanations, predictions, and models using evidence. Students should base their explanation on what they observed, and as they develop cognitive skills, they should be able to differentiate explanation from description, providing causes for effects and establishing relationships based on evidence and logical argument. This standards requires a subject knowledge base so the students can effectively conduct investigations, because developing explanations establishes connections between the content of science and the contexts within which students develop new knowledge.
		NSES 8ASI1.5. Think critically and logically to make the relationships between evidence and explanations. Thinking critically about evidence includes deciding what evidence should be used and accounting for anomalous data. Specifically, students should be able to review data from a simple experiment, summarize the data, and form a logical argument about the cause-and-effect relationships in the experiment. Students should begin to state some explanations in terms of the relationship between two or more variables. NSES 8ASI1.6. Recognize and analyze alternative explanations and predictions. Students should develop the ability to listen and to respect the explanations proposed by other students. They should remain open to and acknowledge different ideas and explanations, be able to accept the skepticism of others, and consider alternative explanations. NSES 8ASI1.7. Communicate scientific procedures and explanations. With practice, students should become competent at communicating experimental methods, following instructions, describing observations, summarizing the results of other groups, and telling other students about investigations and explanations.

5. What are the attributes of a Design Pattern?

A Design Pattern contains attributes that address the necessary elements of an assessment argument (Mislevy, 2003). A total of 22 attributes are specified in a Design Pattern developed in the PADI project; 8 of which are necessary to specify an assessment argument, and 14 additional attributes that provide various kinds of support but are less essential to the assessment argument. Table 2 provides a list of the essential attributes and definition of each (For the details of the exhaustive list of the attributes, the reader is referred to the PADI Technical Report1, *Design patterns for Assessing Science Inquiry*, Mislevy et al, 2003).

Each Design Pattern details three central elements around which all assessments revolve: the student's knowledge, skills, and abilities about which one wants to make an inference (*Focal KSAs*), the salient characteristics of what students say, do, or make that would provide evidence about acquisition of the Focal KSAs (*Potential Observations*), and features of task environment that evoke the desired evidence (*Characteristic Features*). These three attributes are the building blocks that the assessment designers should think through during the entire process of an assessment design in order for the assessment argument to be coherent.

Among the key attributes listed in Table 2, *Rationale* articulates the underlying warrant that justifies the connection between the targeted inferences and the kinds of task and evidence that support them. *Additional KSAs* are other KSAs that may be required in a task that addresses the Focal KSAs. Since Additional KSAs need not be what are intended to be assessed, they can be potential threats to test validity. Therefore, they may need to be minimized or avoided in order not to introduce construct irrelevant variance. Alternatively, if it is known that the examinee group of interest possesses sufficient level of a given Additional KSA, the Additional KSA may be incorporated in the assessment tasks along with the intended KSAs without degrading the validity of inferences for this population. *Potential Work Products* are students' responses or performances that hold clues or evidence relevant to the Focal KSAs. *Potential Rubrics* are links to the rules and instructions that are accessible to evaluate student work products and produce values for Observations. Generic scoring rubrics might be accessible from a PADI library or elsewhere on the Internet so that they can be used if the assessment situation corresponds to the Design Pattern. *Variable Features* of tasks are a primary tool for developers to adjust the difficulty of tasks or focus their evidentiary value on different aspects of the Focal KSA, or to incorporate or circumvent particular Additional KSAs. In addition, each Design Pattern provides links to standards, other Design Patterns, Task Templates and exemplary tasks as appropriate.

Table 2. Essential Attributes of Assessment Design Patterns

Attribute	Definition
Rationale	Nature of the KSAs of interest and why they are important
Focal KSAs	Primary knowledge/skill/abilities of individuals that one wants to make inferences about.
Additional KSAs	Other KSAs that may be required for successful performance on the item or task but are not the target of the assessment.
Potential Observations	Observable activities that provide evidence about the KSA of interest.
Potential Work Products	Different modes or formats in which individuals produce the evidence of the Focal KSAs.
Potential Rubrics	Scoring rubrics that might be useful in evaluating Potential Work Products.
Characteristic Features	Features of an item or task that must be present to elicit the desired evidence.
Variable Features	Features that can be varied in order to shift the difficulty or focus of an item or task.
Narrative Structures	The overall storyline that characterizes multiple prompts in scenario-based tasks.

Mislevy, R. J. (2003). Substance and structure in assessment arguments. *Law, Risk, and Probability, 2*, 237-258.

6. What are the differences between Task Specifications and Design Patterns?

ECD, as applied to the field of assessment, is a design process that includes analyzing and learning domains, specifying assessment arguments, embodying arguments in terms of student, task and evidence models, implementing the assessment, and executing its delivery through operational processes. Design Patterns and Task Specifications are resources that assessment developers create and use as they build assessment arguments and design operational assessments that reflect the purposes, constraints, and resources of their intended use.

A Design Pattern is a tool in the Domain Modeling layer. For a targeted aspect of proficiency (e.g., observational investigation), a Design Pattern provides support for the kinds of features that must be in tasks to assess it and features that can be varied, and kinds of work products and their features that can be used as evidence. It is written broadly, with a focus on the nature of the proficiency, so it can help developers design a variety of kinds of tasks, from short, constrained test items to complex investigations, or interactive oral exams to simulation tasks. . A Design Pattern supports task design for the proficiency in categories that are motivated by the form of assessment arguments, and provide advice for the elements of task and evidence models.

Alternatively, Task Specifications provide a detailed blueprint for designing and writing tasks with specified properties that suit the purposes, constraints, and resources of the particular testing context. A developer can use a Design Pattern to help create Task Specifications, using the support it provides about the content area. Tasks created from the same specification are equivalent in terms of purpose, cost, and can be used interchangeably in a specific context.

7. Why use a Design Pattern?

Design Patterns act as a type of advance organizer for task and items writers, explicating important elements and bounding the design space for these writers. Consisting of a series of interconnected attributes, a Design Pattern helps writers consider choices, see connections among these attributes, and gain awareness of nuances in the design space. In particular, Design Patterns help new item and task writers think through the substantive features of the task and item writing assignment, thereby helping them organize thoughts and ideas prior to and during the writing process. Additionally, in standards-based assessments, Design Patterns help writers consider task and item aspects associated with assigned standards, assisting the writers in achieving alignment to the standards.

Design Patterns are a flexible resource that can be used in many different ways when approaching task and item writing. There are a number of ways Design Patterns are used by writers. Research (Snow, Fulkerson, Feng, Nichols, Mislevy, & Haertel, 2010) suggests that:

- Design Pattern attributes are a source of ideas for writing standards-aligned items,
- Design Patterns help writers plan the content and flow of scenario-based tasks,
- Design Patterns can increase the diversity and richness of a set of items / tasks by reducing redundancy, and
- Design Patterns help ensure that items align to standards and appropriately assess the intended content.

Design Patterns can be used for a number of purposes, including curriculum design, evaluation design, and designing video games.

8. How do Design Patterns integrate the Student, Evidence, and Task models?

Each Design Pattern is organized around the general form of an assessment argument that specifies the knowledge, skill, or ability to be assessed (student model), the kinds of observations that can provide evidence about the attainment of this knowledge, skill, or ability (evidence model), and the features of assessment tasks that allow students to provide this evidence (task model). Table 3 defines the attributes within a standard Design Pattern, rendering explicit an assessment argument (shown according to Messick's student, evidence, and task model components) (Messick, 1994). The Design Pattern is intended to reflect an integrated assessment argument when completed; however, the coherence of the three ECD models can be further enhanced by taking additional steps.

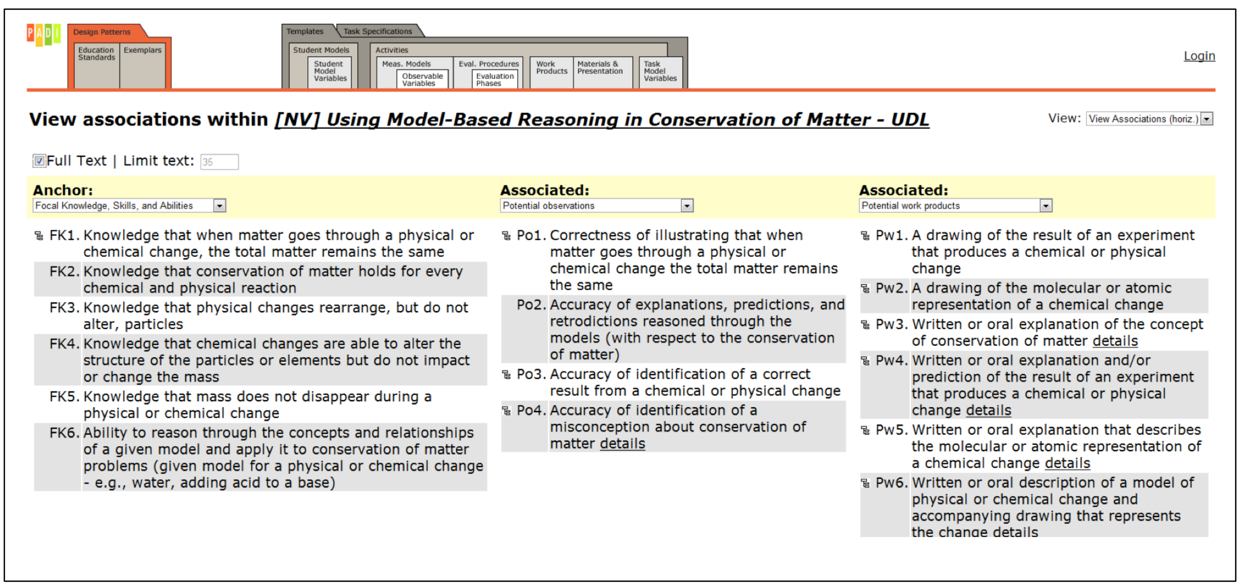
Table 3. Design Pattern Attributes, Definitions, and Corresponding Messick Argument Components

Design Pattern Attribute	Attribute Definition	Messick Assessment Argument Component
Title	Short name for the Design Pattern	
Summary	Brief description of the family of tasks implied by the Design Pattern	
Rationale	Nature of the KSAs of interest and why they are important	
Focal Knowledge, Skills & Abilities (KSAs)	Primary knowledge/skill/abilities of individuals that one wants to make inferences about.	**Student Model/Claim** What construct (complex of student knowledge, skills, or abilities) should be assessed?
Additional KSAs	Other KSAs that may be required for successful performance on the item or task but are not the target of the assessment.	
Potential Observations	Features of student work that provide evidence about the KSA of interest.	
Potential Work Products	Different modes or formats in which individuals produce the evidence of the Focal KSAs.	**Evidence Model/Actions** What behaviors should reveal the construct?
Potential Rubrics	Scoring rubrics that might be useful in evaluating Potential Work Products, to produce values of Observations.	
Characteristic Task Features	Features of an item or task that must be present to elicit the desired evidence.	
Variable Task Features	Features that can be varied in order to shift the difficulty or focus of an item or task.	**Task Model/Situation** What tasks should elicit those behaviors?
Narrative Structures	The overall storyline that characterizes multiple prompts in scenario-based tasks.	
Educational Standards	National standards or state extended standards (if appropriate)	**Student Model/Claim**

In using a Design Pattern to help create a task, it is helpful to begin by identifying components of the student model. This effort can include identifying the educational standards that are the target of the assessment and specifying these as Focal KSAs, which reflect assessable components of the standard. One may also brainstorm Additional KSAs, such as cognitive background knowledge associated with Focal KSAs at the same time. Educational standards may also suggest boundary statements that can be captured as Characteristic Features of tasks. For example, in mathematics, these boundary statements may refer to the types of numbers (e.g., integers, rational, irrational) to be included in tasks to elicit students' skill in performing an operation.

The Focal KSAs can be used to craft Potential Observations and Potential Work Products. The Potential Observations and Work Products help to make concrete the behaviors and performances anticipated as evidence of the Focal KSAs. (Designers often find the horizontal view, provided in the PADI Design System, supports the process of linking the Focal KSAs (Student Model), Potential Observations, and Work Products (Evidence Model). See Figure 9 for an example PADI's Horizontal View.)

Figure 9. Horizontal View Associating Focal KSAs, Potential Observations, and Work Products



After considering the Focal KSAs and evidence required to make inferences about them, it is helpful to revisit the Additional KSAs. Drawing on examples provided by the Potential Observations and Potential Work Products, the designer identifies Additional Knowledge, Skills, and Abilities that may be required for successful performance on the tasks suggested by the Potential Observations and Work Products but are not the target of the assessment. For example, if the knowledge to be assessed is the ability to add two-digit numbers (Focal KSA) and the assessment items include stems that will require the student be a proficient reader, then proficiency reading would be an Additional KSA.

The Student Model and Evidence Model components should be relatively consistent with each other at this time and suggest possible task features. Thinking through the Characteristic Features and Variable Features help to further formalize the Task Model. Considering the Focal KSAs, Potential Observations and Work Products, it is possible to identify what features should be characteristic or consistent of all tasks or items associated with the Design Pattern and what task features can be varied across items or tasks, but still elicit the desired Focal KSAs.

9. How have professional item development teams learned to use the PADI System? (An 11 step training process)

There is not a single way to use the PADI system. The system is a flexible design tool that systematically helps ensure key structures and relationships will be addressed in the assessment design process, but optimal process can vary from one project to another. In various projects, we have developed training materials to help people start using PADI in a way that best suits their purposes. The NSF-funded “An Application of ECD to States Large-scale Science Assessment” project is one such project that shows how this can be done. In this project, GED® item writers were provided an 11-step training process. The background of the item writers and the 11-step process are detailed below.

GED® science scenario and item writers were selected from a pool of highly qualified task / item writers with proven experience writing scenarios and items for various assessment programs. Some of the writers selected for GED® science had previous exposure to ECD and had been previously trained to use PADI Design Patterns in the development of scenario-based tasks and items. Other writers, however, did not have this experience. To accommodate the writers' varied ECD experience and to ensure adherence to the production schedule, Pearson content specialists developed a user-friendly 11-step process for efficiently and effectively incorporating the use of Design Patterns to the development of GED® science scenario-based tasks and items. Writers that followed this process successfully incorporated ECD into the scenario-writing process while minimizing the challenges associated with a less structured process.

Step 1: Become familiar with Design Patterns and the PADI system.

The project team selected a small library of eight Design Patterns for use in developing GED® scenario-based science tasks. Writers were asked to spend time understanding the purpose and attributes of each Design Pattern and learn how to access and use those Design Patterns in the PADI system.

Step 2: Review assigned targets and indicators.

Each writer was given a writing assignment consisting of GED® science content targets and science practice indicators. Writers were asked to review and understand the GED® Assessment Targets document as a whole and, specifically, their assigned targets and indicators.

Step 3: Brainstorm and research a scenario topic idea that encompassed assigned targets and indicators.

Since GED® science items were presented in the context of a scenario, writers were asked to brainstorm a suitable and scientifically sound scenario topic that would appropriately serve as a context for items aligned to the assigned targets and indicators.

Step 4: Refer to an activation chart to connect Design Patterns to assigned science practice indicators.

The project team developed an activation chart that was designed to connect GED® science practice indicators and Design Pattern Focal KSAs. The purpose of this activation chart was to assist with the selection of appropriate Design Patterns for a given scenario and to facilitate the incorporation of Design Pattern attributes into the writers' development process. Activation Charts are described in more detail in Question 21 below.

Step 5: Use Design Pattern attributes (Focal and Additional KSAs, Characteristic and Variable Features, Work Products, and/or Potential Observations) to develop all item ideas (when feasible).

Once the developers were able to recognize the connections between Design Pattern attributes and GED® targets and indicators, these developers were asked to consider Design Pattern attributes when developing item ideas.

Step 6: Modify scenario idea based on item ideas and Design Patterns.

The development of item ideas and the use of Design Patterns occasionally required modification to ensure that the scenario idea remained an appropriate context for the proposed items.

Step 7: Ensure scenario and item ideas are aligned to assigned targets and indicators and selected Design Pattern.

This step asked writers to reflect on the proper alignment of the scenario and item ideas. The writers were instructed to make modifications, if necessary, to properly align the targets, indicators and Design Patterns.

Step 8: Fill out scenario planning form.

Pearson content specialists required writers to complete and submit a scenario planning form before writers populated the scenario and item templates. The planning form asked writers to describe scenario and item ideas and to explicate connections to Design Pattern attributes. This step ensured that scenarios and item ideas were appropriately aligned to GED® targets and indicators and to Design Pattern attributes before investing the time and effort required to write the scenario and items in the appropriate templates. This step formally instantiated the reflective activity in Step 7 above.

Step 9: Get approval from Pearson content specialists.

Writers were required to receive approval on their scenario planning forms by Pearson content specialists to ensure scenario and item ideas were aligned and appropriate.

Step 10: Write scenarios and items in templates and submit for processing.

Once scenario-planning forms were approved, writers were allowed to write the scenario and items in the templates and submit the final forms to Pearson for processing.

Step 11: Resubmit completed scenario planning form.

Once the scenario was completed, writers were asked to fill out additional fields in the scenario planning form and submit the completed scenario planning form.

10. How can Design Patterns be used to identify construct-irrelevant variance?

Construct-irrelevant variance is the variance in assessment scores that can arise when assessment items or tasks require knowledge, skills or abilities that are not the focus of the assessment, centering on the Focal KSAs in the Design Pattern. For example, students' reading ability may impact their ability to comprehend detailed tasks such as those presented in scenario-based tasks about math or science. If the task is meant to assess a student's math or science ability, difficulty with reading directions or stimuli, rather than a lack of math or science knowledge, may result in the student's inability to complete the task. Since the focus of the task was on math or science, not reading, the student's inability to read should be irrelevant to the construct being measured. Additional KSAs, identified in the Design Pattern, can alert a developer to possible sources of construct-irrelevant variance. These Additional KSAs are skills or knowledge that a task might require, but that are not the target of the assessment. Item / task designers should consider each Additional KSA and the impact of those Additional KSAs on potential scores. Designers may choose to alter assessment designs to mitigate the influence of Additional KSAs on scores. Alternatively, designers can support Additional KSAs through the use of Variable Features. For example, struggling readers who have difficulty interpreting extended text presented in scenario-based tasks might benefit from simplified text. Such a support will aid students in terms of the reading demands of the task, but would not provide assistance in construct-relevant aspects of task performance (i.e., performance on Focal KSAs such as using science knowledge in a scenario-based science task). The PADI project has produced more detailed documentation about the use of Variable Features to support Additional KSAs (see Questions 12 and 13).

11. How can Design Patterns be used to design assessments that measure “hard to measure” constructs?

In recent years, assessment designers and researchers have used the term” hard to measure” to describe a class of KSAs that are not easily evaluated using traditional assessment methods because they require problem contexts that are difficult to reproduce in classroom settings (e.g. nuclear fusion, epidemiological outbreaks, volcano eruptions). In addition, these KSAs can have inter-dependencies such as those that might occur when measuring phases of scientific inquiry. These inter-dependencies present both design and statistical challenges. For example, a student’s scientific practices, such as argumentation, can be dependent both on his / her scientific domain knowledge and the features of the assessment task in which argumentation is being observed. Such inter-dependencies challenge interpretation of student performances.

Design Patterns can assist in the conceptualization of the Task, Student, and Evidence Models for hard-to-measure constructs. In the Task Model, designers articulate the Characteristic Features of tasks that are required in order to assess such hard-to-measure KSAs, while Variable Features are identified in order to manipulate task difficulty. The Student Model is expressed through the Focal and Additional KSAs. That is, the relationships among the student model variables – reasoning or skill based constructs and disciplinary content knowledge – are described. Once the Focal and Additional KSAs are identified and their relationships are specified, interpretation of student performance becomes more tractable. In addition, through the use of Potential Observations and Rubrics (the Evidence Model), hard-to-measure KSAs can be evaluated. The PADI project has built design patterns to support assessment design for several “hard to assess” proficiencies in science, including observational and experimental investigation, systems thinking, scientific explanation, and model-based reasoning.

Developers can use Design Patterns to create families of tasks that assess hard-to-measure constructs. Once the assessment argument is articulated in the Task, Student, and Evidence models of a Design Pattern, task variants can be easily created. In addition, design arguments can also be used across Design Patterns. Designers can reuse ideas developed in one design context to structure related design arguments in other contexts.

12. How can I use Variable Features when designing assessments?

When developing a task it is important to consider the features of that task. The Characteristic Features attribute of the Design Pattern describes features that should be part of every task designed using that Design Pattern. Model revision tasks, for example, must always in some way have a model that is inconsistent or inadequate in some way and a need to revise it. Variable Features, on the other hand, can vary across different tasks created under the same design pattern. Variable Features can provide information on how to vary the difficulty of a task by manipulating the amount of scaffolding of a task or some other features without varying its focus.

For example, if the focus of a task was Interpretation of Data, then a Variable Feature of a design pattern that supports creating such tasks is the representational form for the data. Consequently, a task developer will need to make a decision about how to represent the data in a given data-interpretation task. This decision may in turn introduce Additional KSAs. For example, if the developer decides to provide data in a bar graph and interpreting this particular representation is not the focus of the task, then to perform successfully, students must be able to interpret bar graphs. The ability to read data in the form of a bar graph would be an Additional KSA. The developer would need to decide whether to support this Additional KSA by providing the student directions on how to read a bar graph. Another Variable Feature is whether multiple representations of data will be used. For example, it may be useful to provide the data in both a bar graph and a table format. Using multiple representations may reduce the dependency on the Additional KSA, which requires that students be capable of reading a bar graph, by allowing students to choose the representation that lets them to successfully demonstrate their capability to interpret data.

Other Variable Features may be used to influence the difficulty of the task. Continuing the example above, the number of data points and the complexity of the pattern in the data could also be Variable Features, because changing the number of data points and / or data pattern used (e.g. linear, quadratic, exponential) would alter the task difficulty without altering the Focal KSA being assessed.

Item developers make decisions about Variable Features when developing a task. Once an item developer has an overall framework for a task based on the Focal KSAs and the Characteristic Features, then the item developer can use the Variable Features to further define the task and vary its difficulty. Variable Features can be used to adjust the difficulty of a task that addresses one or more Focal KSAs or to support Additional KSAs in order to mitigate the construct-irrelevant variance that appears in the resulting task.

13. How can Design Patterns be used to integrate ECD and Universal Design for Learning (UDL)?

UDL is a conceptual framework that has been extended to assessment design. UDL accounts for individual differences in how students recognize, strategize, and engage in learning and testing situations, and has been integrated synergistically into the ECD framework. Incorporating UDL in the ECD process can enrich the assessment design process by encouraging item writers to consider multiple means of perception, expression, cognition, language and symbol use, executive functioning, and engagement. Moreover, the incorporation of UDL is intended to minimize unintended negative influences that access needs may have on student performance (e.g., the student can read the text or can respond to a test question in a way that fully expresses their understanding of the targeted concept). In this way, UDL, like ECD, is intended to maximize opportunities for students to show what they know and can do in terms of the targeted concepts and skills.

The Additional KSAs and Variable Features detailed in Design Patterns play a key role in integrating ECD and UDL (see Questions 10 and 12 for additional information). While Focal KSAs address the targeted concepts and skills, Additional KSAs can be used to define perceptual (receptive) skills, expressive skills, knowledge of language and symbols, cognitive skills, executive skills, and affective dimensions that are also relevant in an assessment context (CAST, 2011). The role of Additional KSAs is to specify the Knowledge, Skills and Abilities that may be required for a student to perform successfully on a task, but that are not focus of the assessment. Attention to these categories of Additional KSAs (e.g. perceptual (receptive) skills, expressive skills, etc.) increases the likelihood that tasks resulting from a Design Pattern are consistent with principles of UDL (Nagle, DeBarger, Cameto, Haertel, Morrison, Seeratan, & Knokey, 2012). Variable Features associated with these Additional KSAs, in turn, offer strategies for minimizing the demands of these Additional KSAs. Table 4 provides examples of associations between Additional KSAs and Variable Features in each UDL-related category.

Table 4. UDL-Related Categories of Additional KSAs and Variable Features

Category	Definition of Additional KSA	Definition of Variable Features
Perceptual (Receptive)	KSAs associated with perceiving or receiving images, physical objects, and linguistic components of tasks	Ways to vary the delivery mechanisms by which tasks are perceived and task supports for the use of equipment required for assessments
Skill and Fluency (Expressive)	KSAs associated with communicating/expressing a response and using/manipulating equipment and physical materials	Task supports for responding to and composing a response and supports for manipulating equipment and physical materials
Language and Symbols	KSAs associated with decoding, recognizing, and comprehending text, symbols and images, and understanding vocabulary and syntax in which tasks will be presented.	Task options for presenting language and symbols and supporting students in comprehending essential text, symbols, and images.
Cognitive	KSAs associated with cognitive and information processing problems (e.g., ability to access relevant Knowledge, Skills and Abilities and apply them in solving problems), and skills associated with using supports provided as part of the task (e.g., ability to understand the purpose of highlighted features in text or illustrations)	Task options for varying the complexity of tasks; for guiding exploration and information processing (e.g., sequential highlighting); for supporting the identification of critical task features, big ideas and relations (e.g., graphic organizer); and for supporting memory and transfer (e.g., embed task in a scenario)
Executive	KSAs associated with monitoring, planning and sequencing, self-regulating and reflecting, and setting goals and expectations (aka metacognitive skills).	Task options for the provision of guides, checklists, graphic organizers, and templates; for prompts, scaffolds and questions to monitor progress; and for adjusting levels of challenge and support
Affective	KSAs associated with engaging, persisting, and sustaining effort in tasks	Task options for engagement (e.g., enhancing relevance, value, and salience of tasks) and teacher options for supporting student attention and engagement (e.g., prompting the student to engage)

Other elements of the Design Pattern also can be informed by UDL principles. Characteristic Features can set boundaries for tasks by setting the upper and lower limits for task complexity when assessing the Focal KSAs. For example, in the domain of mathematics, Characteristic Features may limit the types of numbers used to assess students' ability to solve word problems. In English Language Arts, Characteristic Features can set bounds for length or contexts of text passages. In considering which task features should be characteristic, it is important to remember that boundaries should not limit what students should show about their understanding of the knowledge and skills that are focal.

Attention to principles of UDL also suggests more thoughtfulness in how Potential Observations and Potential Work Products are defined. A Design Pattern can articulate Potential Observations and Potential Work Products such that multiple modes of response are noted as possible ways to provide evidence of Focal KSAs. Potential Observations can often be expressed in different ways (e.g., verbally, in writing, or behaviorally). Likewise, corresponding Work Products may be defined more broadly (e.g., "expression of an answer"), rather than narrowly (e.g., "written response"). UDL suggests that students should be allowed to demonstrate their Knowledge, Skills and Abilities in a variety of ways. Consequently, when Potential Observations require a specific response mode, it is often important to generate additional Potential Observations that offer other means for students to show what they know about a Focal KSA.

14. How can Design Patterns be used to ensure the coherence of the assessment argument?

Design Patterns are built around attributes organized in the general form of an assessment argument (Messick, 1994), that specify the knowledge, skill or ability about which one wants to make inferences, the kinds of observations that can provide evidence about the attainment of this knowledge, skill, or ability, and the features of assessment tasks that allow students to provide this evidence. Table 5 defines the attributes within a standard Design Pattern template.

Table 5. *Design Pattern* Attributes, Definitions

***Design Pattern* Attribute**	**Attribute Definition**
Summary	Brief description of the family of tasks implied by the Design Pattern
Rationale	Nature of the KSAs of interest and why they are important
Focal Knowledge, Skills & Abilities (KSAs)	Primary knowledge/skill/abilities of individuals that one wants to make inferences about.
Additional KSAs	Other KSAs that may be required for successful performance on the item or task but are not the target of the assessment.
Potential Observations	Features of student work that provide evidence about the KSA of interest.
Potential Work Products	Different modes or formats in which individuals produce the evidence of the Focal KSAs.
Potential Rubrics	Scoring rubrics that might be useful in evaluating Potential Work Products, to produce values of Observations.
Characteristic Task Features	Features of an item or task that must be present to elicit the desired evidence.
Variable Task Features	Features that can be varied in order to shift the difficulty or focus of an item or task.
Educational Standards	National standards or state extended standards (if appropriate)

Design Patterns provide assessment designers with a means to model the high-level thinking that must precede the particular technical decisions required in the development of the actual assessment tasks, identification of psychometric models, and articulation of decision rules required for scoring tasks. They serve as an “in-between” layer that connects the content of an assessment argument to the actual structure of the argument. As such, the designer must take steps during the development of Design Patterns to ensure that the attributes of the Design Pattern are aligned with each other and will support efficient item design and the valid interpretations of scores produced by the resulting items.

For example, in creating the Design Pattern for Observational Investigation (Mislevy, Liu, Cho, Fulkerson, Nichols, Zalles, et al., 2009), assessment designers began by specifying the Focal KSAs including, among others, the *ability to generate or evaluate predictions or hypotheses about scientific phenomena that are appropriate for observational investigation*. In thinking, then, about the Potential Observations that can provide evidence about the *ability to generate or evaluate predictions or hypotheses*, the assessment designers identified several possibilities aligned with the targeted proficiency, including *correctness of recognized patterns in data to support a prediction or hypothesis*, and *appropriateness/strength of observational evidence to help confirm or disconfirm a prediction or hypothesis*. Finally, to complete the basic form of an assessment argument for observational investigation, the assessment designers identified Potential Work Products, or features of assessment tasks, that allow students to provide evidence of the targeted proficiency, including *filling in of a representational form (e.g., a graph, chart, or map) to show the relationship among variables relevant to a prediction or hypothesis*, and *generation or selection of an explanation for observed findings* (see Table 6).

Table 6. ***Alignment of Focal KSAs, Potential Observations and Potential Work Products, Design Pattern for Observational Investigation***

Focal KSA	**Potential Observations**	**Potential Work Products**
Ability to generate or evaluate predictions or hypotheses about scientific phenomena that are appropriate for observational investigation	Correctness of recognized patterns in data to support a prediction or hypothesis, Appropriateness/strength of observational evidence to help confirm or disconfirm a prediction or hypothesis	Filling in of a representational form (e.g., a graph, chart, or map) to show the relationship among variables relevant to a prediction or hypothesis Identification or generation of a prediction or hypothesis that is appropriate to an observational investigation situation

15. How do Design Patterns support the integration of content and practice in a domain?

New standards, such as the Next Generation Science Standards (Achieve, 2013) and Common Core State Standards (National Governors Association Center for Best Practices, Council of Chief State School Officers, 2010), call explicitly for the integration of content and practice. These standards define examples of performance expectations that reflect how students can show what they know about the content through engagement in a practice. Design Patterns can function as a tool for making more explicit an assessment argument reflecting a blend of core ideas and practices, going beyond performance expectations to clearly articulated Focal KSAs reflecting a student model, anticipated Potential Observations, Rubrics and Work Products reflecting the Evidence Model, and Characteristic and Variable Features informing the Task Model. In other words, Design Patterns further specify how the integration of content and practice can be defined for the purposes of assessment.

Suppose an assessment was designed to address the following:

- **Core concepts:** Many characteristics of organisms are inherited from their parents. Other characteristics result from individuals' interactions with the environment, which can range from diet to learning. Many characteristics involve both inheritance and environment. (LS 3.A, NRC, 2012)

- **Scientific practice:** Support scientific arguments drawing on evidence, data, or a model. (NRC, 2012)

- **Performance expectations** (Achieve, 2013):

    - (1) Use evidence to compare characteristics inherited from parents, characteristics caused by the environment, and those resulting from both. [Clarification Statement: Examples of characteristics inherited from parents could be the ability to roll one's tongue or characteristics of domestic animals; characteristics caused by the environment could be a scar or language; and characteristics resulting from both could be height or some health conditions.] [Assessment Boundary: The mechanisms of inheritance are not to be included.];

    - (2) Provide evidence that offspring can inherit different information from their parents. [Clarification Statement: Examples of different information that can be inherited could be different coat colors in dogs of the same litter or one sibling who needs glasses and another who does not.] [Assessment Boundary: The genetic mechanisms of inheritance are not to be included.] ;

    - (3) Use evidence to describe patterns of variation in a trait across individuals of the same kind of organism. [Clarification Statement: Examples of variation in a trait across individuals of the same kind of organism could be different coloration of wolves or thickness of wool in sheep.] [Assessment Boundary: The genetic mechanisms of inheritance are not to be included.]

Although the performance expectations are helpful in focusing assessment design on how the content and practices can work together, sufficient detail to create an assessment is still lacking. In this case, questions such as the following need to be answered when designing the assessment.

- Drawing from the examples in the clarification statements and assessment boundaries, what is the range of the grade-level appropriate contexts for these core ideas, level of detail with respect to characteristics, and the kind of evidence students are expected to engage?
- What kind of evidence is appropriate for each of the performance expectations?
- How will students obtain this evidence to include in their responses?
- What qualifies as a complete and coherent scientific argument?
- How should students express their scientific argument (e.g., writing, drawing, speaking)?
- Which task features are essential for eliciting the knowledge and skills for this content area on inherited traits and the practice of developing scientific arguments?
- What task design features might be varied (e.g., computer-based simulations, computer-based animations, paper-pencil writing and drawing)?
- What specific boundaries need to be clearly communicated to item writers so the appropriate concepts are targeted?

Design Patterns can be a useful tool for helping to sort out answers to these questions. For example, given the third performance expectation, “Use evidence to describe patterns of variation in a trait across individuals of the same kind of organism,” more specific knowledge, skills and abilities might be targeted by defining more focused Focal KSAs such as: (1) *Ability to use evidence to describe patterns of variation of the same species in a generation*, and (2) *Ability to use evidence to describe patterns of variation of the same species across generations*. Design Patterns provide detail about what qualifies as appropriate evidence in student work as Potential Observations. For example, Potential Observations for the second Focal KSA might be something like correctness, accuracy, and completeness of descriptions of varying features in data with variation in one or more features of individuals from the same species across generations. An idea for a particular task that uses this generally-stated observable in a specific context would be the following: *Given evidence that shows variation in stem color in a plant species across two generations, student identifies stem color as a trait that varies in this species of plant and indicates numbers of plants that have different stem colors in the two generations. Student develops an argument by explaining that traits are passed from one generation to the next*.

The Characteristic Features of the Design Pattern can specify that genetic mechanisms (the boundary listed in the NGSS performance expectations) should not be included. Additional features could also be specified such as whether plant and animal contexts must be specified (or must be excluded), and which traits/characteristics about which students can be assessed. Characteristic Features might also include statements that make explicit how detailed arguments should be (e.g., if merely eliciting descriptions of evidence is sufficient or if tasks need to elicit more of a scientific argumentation). Variable Features too offer guidance to item writers about options for what they can change in tasks in terms of content, practice, and format. See Question 12 to learn more about the use of Variable Features.

16. How can Design Patterns be used to support the assessment of a science practice when the science content is not explicit??

The Design Patterns initially produced in PADI often focused on science practices—the ways students are expected to reason, to investigate, to model, to explain, and so on. Each of these Design Patterns addressed Characteristic Features of situations where these science practices might be observed, regardless of the particular scientific content that might be involved, and so the design pattern would help a developer create a task to assess this practice in the context of particular content. For example, one can "critique a model's fit to a situation" in all branches of science and all levels of education, from elementary education to graduate study. Regardless of grade level or scientific content being addressed, it is necessary for students to describe the scientific model, compare what the model would predict for some data or situation, and identify and characterize anomalies. The model might be given or developed by the student; the data might also be given, or the student might have had to determine what data was needed and obtain it; and analyze the lack of fit between the data and the model. By highlighting these and many other facets of a task design space, the Model Evaluation Design Pattern (see Mislevy, Riconscente, & Rutstein, 2009) helps assessment developers craft tasks that provided evidence about a student's ability to evaluate a scientific model, in the context of some particular model(s).

It is possible to write more targeted Design Patterns for assessment tasks associated with particular practices *in* a content domain. The Technical Report, *Design Pattern on Model Use in Interdependence Among Living Systems* (DeBarger, & Snow, 2010), is an example of this type of Design Pattern. It articulates how particular scientific models are used in situations that revolve around interdependence among living systems. Such a Design Pattern provides more support for developing assessment tasks than a Design Pattern that focuses solely on practice. A Design Pattern that integrates content and practice not only helps the assessment developer think through the science practice captured in the assessment task, but it also specifies content. This specificity, however, comes at a cost since, to adequately cover both the content and practices within a particular science domain (e.g., physical science, earth and space science, life science) requires many more Design Patterns than only covering science practices that are common to all three domains.

As part of the work using the PADI Design System, approximately 20 Design Patterns were developed to guide the creation of assessment tasks to measure science inquiry practices as they can arise in all three science domains. Unlike the approach used in the Interdependence Among Living Systems Design Pattern, none of these 20 Design Patterns specified a particular scientific content domain; rather they required test developers to use their own content expertise to create inquiry tasks in a particular targeted domain. If the PADI projects had created Design Patterns that integrated science content as well as practices, then far more than

20 Design Patterns would have been required to build assessments for even a single science-content domain. As such, these 20 Design Patterns facilitated the development of assessment tasks within any science domain

Providing Design Patterns that focus on practices without specifying content does not presume that science practice can be measured independent of content. Science educators may debate the degree to which students' practice skills are bound to models and contexts, but regardless of the structure of students' capabilities, we can identify kinds of thinking and doing related to practices that appear in their own form *across* content areas. We can also support writing tasks around these recurring patterns in each of the science domains. The PADI practice-focused Design Patterns are agnostic on whether scientific content must be incorporated with the science practices within a Design Pattern. Ultimately, however, the assessment tasks developed from either of these Design Pattern approaches will require students demonstrate their ability to practice science within one or more domains of science.

17. How can Design Patterns incorporate learning progressions into the assessment design process?

Learning progressions are defined as the development of student reasoning around specific concepts, topics or practices in one or more content areas. For example, learning progressions concerning the water cycle and the carbon cycle (Gunckel at al., 2012) assume that students' development of understanding about these topics evolves sequentially, starting with recognition of relationships between physically observable objects and forces, and culminating in understanding those relationships at the atomic-molecular level. Learning progressions assume that student progress in a conceptual domain is linear, that it is predictable, and that classroom instruction is aligned to the level of the progression at which students are expected to perform.

Design Patterns can be structured in different ways to reflect the assumptions of a learning progression. The extent to which Design Patterns support the development of items that yield valid inferences about student learning depends on the adequacy of the theoretical foundation for a particular learning progression and the amount of empirical evidence confirming that progression. The use of Design Patterns cannot overcome Domain Analysis that is insufficient in scope, lacking in depth, or theoretically weak.

Provided that a sufficient Domain Analysis is conducted, Design Patterns can be written to document what students should be able to successfully accomplish (Focal KSAs) in various kinds of situations (Characteristic and Variable Features) that capture performances (Potential Observations) at different levels of the learning progression. In ECD, warrants posit how student performances in assessment situations with the desired Characteristic and Variable Features depend on student competencies. The desired situations are provided by the assessment tasks; student competencies are identified by the learning progression research. The task features and desired knowledge, skills and abilities are documented in the Design Pattern.

One way a learning progression is reflected in a Design Pattern is through the use of nested Focal KSAs. A nested Focal KSA groups conceptual understandings and skills at one or more grade levels. Each nested Focal KSA consists of several statements about what a student at that level of the progression knows, says, or can do. These same understandings and skills are represented at each increasingly sophisticated level of the progression until the complete progression, from the simplest conceptual understandings to its most advanced form, are described. An assessment task is constructed to present a situation, in which a student must display the conceptual understanding or skills described in the Focal KSA for that level of the progression. We would infer that, a student who successfully demonstrates such conceptual understanding or skills is at or above that level of the progression.

The notion of Learning Progressions has been challenged in at least two ways. First, it is difficult to disentangle a student's developmental readiness to learn from the opportunity to learn content in an instructional program. The fact that a student is not able to demonstrate grade appropriate understanding or skills may reflect the fact that they are not developmentally ready to learn or that the instructional program in which they are enrolled inadequately addresses the content of the progression. Second, learning progressions may not be generalizable to all learners. Further research is necessary to resolve these challenges.

18. How can multiple Design Patterns be used to support scenario development?

In order to answer this question, it is first necessary to address why it is sometimes desirable to use multiple Design Patterns to develop a scenario-based assessment. In our work with the GED® science assessment, item writers used multiple Design Patterns when they were unable to find a single Design Pattern that supported all the science standards they had been asked to assess within the scenario. However, tying the different Design Patterns together could be a difficult task. As one anonymous item writer said:

"You've got to come up with [a scenario-based assessment] that all the [assigned standards] will apply to. So you're trying to look for a unifying piece of information. And when you go to the Design Patterns, and those standards aren't [under one Design Pattern], you've pulled your scenario out in different directions. And that was not helpful to me. I couldn't figure out how that was going to bring it into one cohesive scenario."

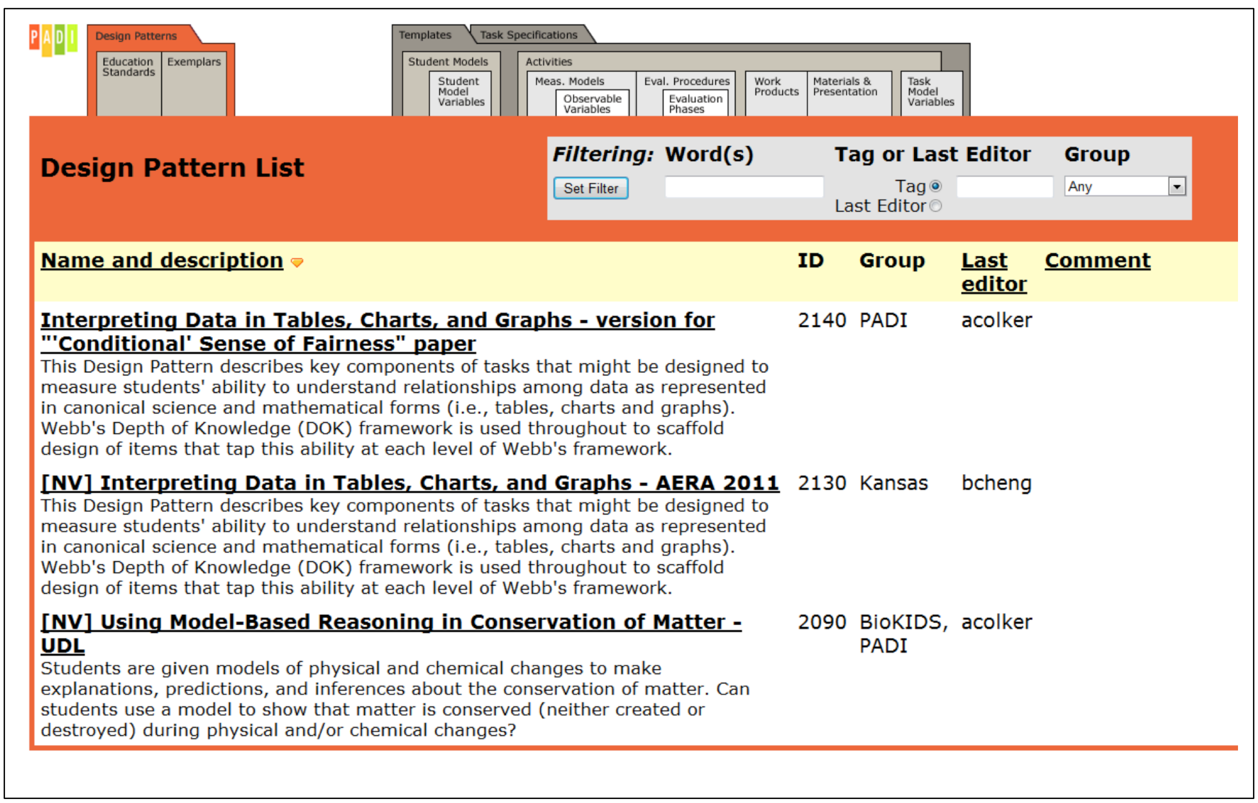
Some item writers had more difficulty than others, and this seemed to be related to how they approached their writing task. Writers who began their items by focusing on the items reported less difficulty with integrating Design Patterns together than did writers who began by focusing on the overall scenario and its narrative structure. When different items were drawn from different Design Patterns – a frequent occurrence – these writers were able to move from the items, or the preliminary item ideas, to the development of a scenario that encompassed those items. The multiple Design Patterns could then be used to guide the scenario development that would encompass all the items they had developed, and thus all the assessed standards.

19. What is a Library of Design Patterns?

For every application of the PADI design system, a library of Design Patterns exists. The library for any given application comprises a set of Design Patterns focused around a topic of interest. For example, there are libraries of Design Patterns associated with topics in: Science, Mathematics, English Language Arts, and Economics. Figure 10 is a screen shot of a library of Science Design Patterns. The Design Patterns in a library are ordered chronologically from those developed earliest to those developed most recently. Across all the applications of PADI, there are over 300 Design Patterns at grade levels ranging from Grades 3 – 16.

The PADI libraries of Design Patterns contain five broad types of patterns. These five types of patterns include those based on: 1) educational standards in specific domains; 2) unifying themes / inquiry, such as model-based reasoning, experimental investigation (NSES, 1996); 3) big ideas within disciplines (Chung, Niemi, & Bewley, 2003); 4) learning progressions that reflect increasingly sophisticated levels of learning; and 5) language proficiencies.

Figure 10. A PADI Library of Design Patterns



Users can scan the Design Pattern Library associated with a particular project when they first enter the PADI system. By clicking on the hyperlink associated with each Design Pattern in the library, the user calls up the appropriate Design Pattern and is able to examine its content. Design Patterns themselves in the PADI system can only be edited by permission.

20. How can Design Patterns support the development of scenario-based assessments?

GED® science-item writers were asked to use Design Patterns to specifically inform the development of items embedded in scenarios. By definition, a scenario-based assessment is a group of assessment items that can be embedded within a single narrative storyline. The scenario context is provided within the narrative storyline. For example, an assessment scenario about an invasive species like the Burmese Python might include both the fact that the python is frequently released in a non-native habitat and that the habitat of interest is the Florida Everglades (the context). In science, narrative storylines may focus on cause-and-effect, change over time, or conducting an investigation, etc. The number of items and the format of the items in a single scenario can vary depending on the complexity of the content assessed.

The experience of GED® item-writers using Design Patterns to construct scenario-based assessments suggest that Design Patterns can inform item development on two levels: at the item level, and at the scenario level. The case of one item writer, who wrote two GED® scenarios and used Design Patterns very differently in each, is described below.

In the development of a scenario-based assessment and its associated items, the item writer used Design Patterns to develop discrete, standards-aligned items that were suitable given a minimally developed scenario narrative structure. Items were each aligned to Focal and Additional KSAs and to a GED® content target and science practice indicator. The Design Pattern activation chart (see Question 21 below) was used to find reasonable connections among Design Patterns and the science practice indicators. Once the items were written, the item writer was able to embellish the scenario narrative to provide an appropriate context in which to embed the discrete items.

For another writing assignment, the same writer used a Design Pattern to inform the development of a scenario context before the items were written or even conceptualized. Once the scenario was complete, standards-aligned items were then written to fit the context of the scenario. In this situation, the Focal and Additional KSAs, Potential Work Products, and Potential Observations of the Experimental Investigation Design Pattern were used to create a suitable scenario that depicted an experimental investigation. The Design Pattern helped generate the scenario idea and flow, and guided the conceptualization and writing of the scenario. Once the scenario was developed, the items were written to fit the scenario context and the assigned GED® science practice indicators. Because the items were written to the context of a Design Pattern-informed scenario, it can be suggested that the items themselves were also well-aligned to the Design Pattern and the construct(s) represented. Examination of the items confirmed alignment to the Design Pattern.

21. How can Design Patterns and Standards be associated? (Using Activation Charts)

Item writers are often given specific science practices or standards and asked to create items based on these standards. The Activation Chart is a means by which item writers can quickly access Design Patterns related to the standards and science practices they are using to create items.

An Activation Chart links standards in the content or practice area to the Focal KSAs identified in the Design Patterns of interest. This linkage prevents item writers from having to review many different Design Patterns to determine which are linked to the standard or practice of interest.

In the Activation Chart illustrated below, the item writer first identifies the science practice (the GED Science Indicators in the first column of Figure 11) that is the focus of their item writing assignment. The item writer connects the science practices and standards with one or more Focal KSAs that are being assessed by moving across the chart (the second column of Figure 11). These Focal KSAs are then linked to Design Patterns (in the third column of Figure 11) that include these KSAs. The Activation Chart provides hyperlinks to the associated Design Patterns that allows the item writer to quickly access the PADI Design System and the appropriate Design Patterns.

Figure 11. Activation Chart

Design Pattern Activation Spreadsheet - Science Practice 3

GED Science Indicators	PADI Focal KSAs	Associated Design Patterns
SP.3 Analyzing events and ideas		
a) Determine which explanation best accords with evidence.	FK11 Ability to interpret or appropriately generalize the results of a simple experiment or to formulate conclusions or create models from the results	Design Pattern for Experimental Investigation Tech Report Version - 2245
	FK4 Ability to formulate conclusions, create models, and appropriately generalize results from observational investigations	Design Pattern for Observational Investigation - 2167
	FK1 Ability to reason through the concepts and relationships of a given model to make explanations, predictions and conjectures	Design Pattern for Model Use in Model-Based Reasoning - 2218
	FK1 In broad terms, ability to determine the appropriateness of a model for reasoning about a situation, for a given purpose	Design Pattern for Model Evaluation in Model-Based Reasoning - 2221
	FK2 Knowledge of types of system interactions	Design Pattern for Systems Thinking and Complexity - 2195
b) Analyze in detail a series of event or results described in a stimulus; determine whether earlier events/results caused later ones or are simply correlated with later events/results.	FK3 Ability to recognize that the purpose of an experiment is to test a prediction/hypothesis about a causal relationship	Design Pattern for Experimental Investigation Tech Report Version - 2245
	FK6 Ability to recognize that at a basic level, an experiment involves manipulating one variable and measuring the effect on (or value of) another variable	Design Pattern for Experimental Investigation Tech Report Version - 2245
	FK7 Ability to identify variables of the scientific situation (other than the ones being manipulated or treated as an outcome) that should be controlled (i.e. kept the same) in order to prevent misleading information about the nature of a causal relationship	Design Pattern for Experimental Investigation Tech Report Version - 2245
	FK8 Ability to recognize variables that are inconsequential in the design of an experiment	Design Pattern for Experimental Investigation Tech Report Version - 2245
	FK3 Ability to generate or evaluate predictions or hypotheses about scientific phenomena that are appropriate for observational investigation	Design Pattern for Observational Investigation - 2167

22. How can assessment items or tasks be reverse engineered to create a Design Pattern?

Reverse engineering from an assessment or set of assessment items to a Design Pattern means designing "in reverse". One begins with a specific set of items/tasks about particular knowledge, skills or abilities and creates a higher-level description of the assessment argument in the form of a Design Pattern. Reverse-engineering is particularly desirable if one wants to better understand the design principles underlying an assessment and then use the Design Pattern to "forward-engineer" or generate new items and tasks that assess the same knowledge, skills and abilities.

Reverse-engineering can begin by identifying the student model components of the Design Pattern. Given a set of related items from an assessment, it is possible to distill which knowledge, skills, and abilities are the targets of the assessment. These will become the Focal KSAs in the Design Pattern. For example, a set of items about concepts related to identifying forces may include Focal KSAs such as: (1) *Ability to identify the sources of forces on an object*, (2) *Ability to identify the direction of a force*; and (3) *Ability to compare the relative sizes of forces on static objects*. The items may also suggest what cognitive background knowledge may be required but not directly assessed, as well as expectations of the perceptual skills students will need to access and respond to the items. These item features give rise to the Additional KSAs in the Design Pattern. For an assessment on identifying forces, Additional KSAs associated with background knowledge may include: (1) *Knowledge that forces can make objects move* and (2) *Knowledge that forces can have different magnitudes*.

The next step in reverse-engineering is to identify the evidence model components of the Design Pattern. To define Potential Observations, one examines the items associated with each of the newly defined Focal KSAs and their rubrics (if available) to identify correct responses. Potential Observations, however, will not be lists of correct responses for specific items, but rather a general description of the nature of a correct response for each Focal KSA. For example, reverse-engineering from a set of items about identifying forces, could yield the Potential Observation: *Given an image with an object at rest on a surface, student determines that both the object and the surface exert forces by drawing force arrows for both objects.* Specific examples of correct responses may be added to a Potential Observation for clarification. Potential Work Products are relatively straightforward statements of the kinds of products students will have produced from the items (e.g., selected response, constructed response). Potential Rubrics can highlight principles behind rubric design in the set of items (e.g., use of partial credit).

The final step in reverse-engineering is to identify the task model components of the Design Pattern. Characteristic Features serve to help define task features that are similar across all items that are included in the reverse-engineering process. For example, it may be the case that all of the multiple-choice items must include distractors that reflect misconceptions. There may be bounds or limitations in terms of the concepts and skills assessed or item formats. Variable Features, in contrast, note the range of variation observed in items. Variations may refer to differences across items in terms of complexity of concepts, item formats, inclusion of supports for cognitive background knowledge (e.g., including definitions to support recall), and use of contexts or scenarios in items.

Conclusion

Two objectives of the PADI online, assessment design system are: 1) to improve the quality of assessments by instantiating the principles of Evidence Centered Design (ECD); and 2) making the principles of ECD accessible to all assessment designers. The PADI system is useful for both professional and novice assessment designers. PADI can be used to create different types of assessments including summative, formative, interim, diagnostic, and benchmark examinations in any content area, based on any theory of learning or cognition. The PADI tool can design assessments to be delivered in a number of modalities including pencil and paper, technology based, and oral administrations. As discussed above, the PADI tool can be used to both forward and reverse engineer assessment items and tasks and can integrate the principles of Universal Design for Learning into the assessment design process. As a result, the PADI system improves the quality of assessments by structuring and systematically documenting the design and development process.

The documentation and development of assessments can seem daunting even within the structure of PADI. This technical report is an initial attempt to answer questions that designers new to the PADI system often ask. It is intended to encourage new users to explore the system on their own. Questions about the system arising from such explorations should be addressed to Geneva Haertel (geneva.haertel@sri.com), Robert Mislevy (rmislevy@ets.org), or Terry Vendlinski (terry.vendlinski@sri.com).

References

Achieve, I. (2013). Next Generation Science Standards Washington, D.C., Achieve, Inc.

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). Enhancing the design and delivery of assessment systems: A four-process architecture. *Journal of Technology, Learning, and Assessment, 5*. Retrieved from http://escholarship.bc.edu/ojs/index.php/jtla/article/viewFile/1671/1509

CAST (2011). *Universal Design for Learning Guidelines version 2.0.* Wakefield, MA: Author.

Chung, G. K. W. K., Niemi, D., & Bewley, W. L. (2003). Assessment applications of ontologies. Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL.

DeBarger, A. H., & Snow, A. (2010). *Design Pattern on Model Use in Interdependence among Living Systems (Large-Scale Assessment Technical Report 13).* Menlo Park, CA: SRI International.

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

Gunckel, K.L., Mohan, L., Covitt, B.A., Anderson, C.W. (2012). Addressing challenges in developing learning progressions for environmental science literacy. In Alonzo, A.C. and Gotwals, A.W. (Eds). *Learning progressions in science: Current challenges and future directions.* 39-76. Boston, MA: Sense Publishers.

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

Mislevy, R. J., Chudowsky, N., Draney, K., Fried, R., Gaffney, T., Haertel, G., et al. (2003). *Design Patterns for assessing science inquiry* (PADI Technical Report 1). Menlo Park, CA: SRI International.

Mislevy, R. J. (2003). Substance and structure in assessment arguments. *Law, Risk, and Probability, 2*, 237-258.

Mislevy, R., & Haertel, G. (2006). *Implications of Evidence-Centered Design for Educational Testing (Draft PADI Technical Report 17).* Menlo Park, CA: SRI International.

Mislevy, R.J., Riconscente, M.M., and Rutstein, D.W. (2009) Design Patterns for Assessing Model-Based Reasoning. *Application of Evidence Centered Design to State Large-Scale Science* Assessment (Technical Report 6). Menlo Park, CA: SRI International. Retrieved from http://ecd.sri.com/downloads/ECD_TR6_Model-Based_Reasoning.pdf on March 18, 2013.

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

Mislevy, R., Liu, M., Cho, Y., Fulkerson, D., Nichols, P., Zalles, D., et al., (2009). A Design Pattern for Observational Investigation Assessment Tasks (Large-Scale Assessment Technical Report 2). Menlo Park, CA: SRI International.

Nagle, K., DeBarger, A. H., Cameto, R., Haertel, G., Morrison, K., Seeratan, K., & Knokey, A. M. (2012). Alternate Assessment Design - English Language Arts Technical Report 8: Implementing Evidence-Centered Design to Develop Assessments for Students with Significant Cognitive Disabilities: Procedural Guidelines for Creating Design Patterns and Development Specifications and Exemplar Task Templates for English Language Arts (Alternate Assessment Design - English Language Arts/Reading Technical Report 8). Menlo Park, CA: SRI International.

National Governors Association Center for Best Practices, Council of Chief State School Officers. (2010). Common Core State Standards. National Governors Association Center for Best Practices, Council of Chief State School Officers, Washington D.C.

National Research Council [NRC], (1996). *National Science Education Standards* . Washington, DC: The National Academies Press.

National Research Council [NRC], (2012). A Framework for K-12 Science Education: Practices, Crosscutting Concepts, and Core Ideas, The National Academies Press.

Rasch, G. (1960). *Studies in mathematical psychology: I. Probabilistic models for some intelligence and attainment tests.* Copenhagen, Denmark: Danmarks Paedogogiske Institut.

Rose, D. H., & Meyer, A. (2002). Teaching every student in the digital age: Universal Design for Learning. Retrieved July 12, 2007, from the Association for Supervision and Curriculum Development http://www.cast.org/teachingeverystudent/ideas/tes.

Snow, E., Fulkerson, D., Feng, M., Nichols, P., Mislevy, R., & Haertel, G. (2010). Leveraging Evidence-Centered Design in Large-Scale Test Development (Large-Scale Assessment Technical Report 4). Menlo Park, CA: SRI International.

Wei, H., Mislevy, R. J., & Kanal, D. (2008). An introduction to Design Patterns in language assessment. *Design Patterns for assessing science inquiry* (PADI Technical Report 18). Menlo Park, CA: SRI International.

Wilson, M. (2005). *Constructing measures: An item response modeling approach*. Mahwah, NJ: Erlbaum.

Wright, B.D. (1993). Equitable test equating. *Rasch Measurement Transactions*, 7, 298-299.

!

Sponsor
The National Science Foundation, Grant DRL - 0733172

Prime Grantee
SRI International. *Center for Technology in Learning*

Subgrantees
University of Maryland
Pearson