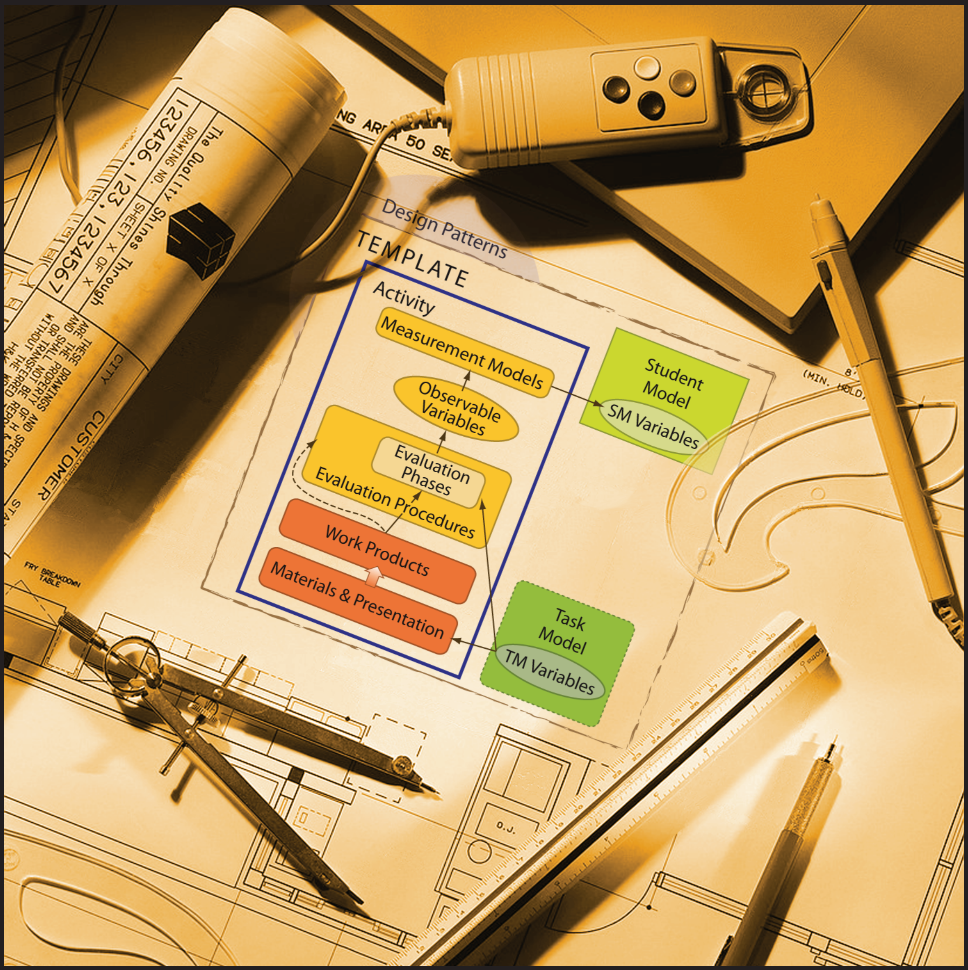

Large-Scale Assessment Technical Report 13| March 2010

Design Pattern on Model Use in Interdependence among Living Systems

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

Angela Haydel DeBarger and Eric Snow, SRI International

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*

DESIGN TO STATE LARGE-SCALE SCIENCE ASSESSMENT

TECHNICAL REPORT 13

Design Pattern on Model Use in Interdependence among Living Things

March 2010

Prepared by:

Angela Haydel DeBarger and Eric Snow, SRI International

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).

CONTENTS

Figures

LIST OF TABLES

ABSTRACT

Design patterns are an important resource for supporting the development of unique and complex assessments, such as technology-supported scenario-based tasks. *Design patterns* encourage coherence in task design by making explicit relations among critical knowledge/skills/abilities (KSAs) to assess, evidence in student work that demonstrates proficiency in KSAs, and features of tasks that can elicit KSAs. The "Model Use in Interdependence among Living Systems" *design pattern* was developed to facilitate the design of storyboards and items for the Science Minnesota Comprehensive Assessment. This *design pattern* articulates an assessment argument describing families of tasks to elicit evidence of students' abilities to use models in the context of life science. The report outlines the key elements of the *design pattern*, describes the *design pattern* development process, and illustrates with examples how the *design pattern* supports storyboard and item design for interactive science assessments.

PURPOSE & GOALS OF THE REPORT

The purpose of this technical report is to describe the development of a hybrid *design pattern* (i.e., mixing hard-to-assess area of science inquiry with specific science content area) that articulates an assessment argument describing families of tasks to elicit evidence of students' abilities to use models in the context of life science, as well as to illustrate the *design pattern's* utility in facilitating the creation of storyboards and items for the Science Minnesota Comprehensive Assessment. We begin by describing the major conceptual foundation underlying this work - evidence-centered assessment design (ECD) and *design patterns* and their application in the Principled Assessment Design for Inquiry (PADI) project. Next, we describe the process and key elements underlying the development of the Model Use in Interdependence among Living Systems hybrid *design pattern*, as well as illustrate with examples how it supports storyboard and item development in the context of the Minnesota Comprehensive Assessment. Finally, we close with a discussion of the benefits and limitations of hybrid *design patterns* in helping facilitate the assessment design, development, and delivery processes, particularly the design of storyboards and items for the Science Minnesota Comprehensive Assessment.

EVIDENCE-CENTERED DESIGN & PADI

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

ECD can be depicted as five interrelated layers, shown in Table 1. Each layer focuses on the substantive domain (Layer 1); the assessment argument (Layer 2); the structure of assessment elements such as tasks, rubrics, and psychometric models (Layer 3); the implementation of these elements (Layer 4); and the way they function in an operational assessment (Layer 5).

Table 1: Layers of ECD

Layer	Role
Layer 1 ***Domain Analysis***	Gather substantive information about the domain of interest that has direct implications for assessment; how knowledge is constructed, acquired, used, and communicated
Layer 2 ***Domain Modeling***	Express assessment argument in narrative form based on information from Domain Analysis
Layer 3 ***Conceptual Assessment Framework***	Express assessment argument in structures and specifications for tasks and tests, evaluation procedures, and measurement models
Layer 4 ***Assessment Implementation***	Implement assessment, including presentation-ready tasks and calibrated measurement models
Layer 5 ***Assessment Delivery***	Coordinate interactions of students and tasks, task-and test-level scoring, reporting

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

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

DESIGN PATTERNS

In the Domain Modeling layer of ECD, information and relationships discovered in domain analysis are organized in a narrative form that serves as a high-level introduction to the assessment argument that will support the new assessment being designed. The work in this layer is a transition from specialized knowledge about the domain to the specialized knowledge about the more technical machinery of assessment, which takes place in the next layer. Toulmin diagrams (1958) are examples of tools for organizing assessment arguments at a narrative level (e.g., Kane, 1992), as are *design patterns*. As discussed in this section, *design patterns* are meant to guide the design of families of assessment tasks organized around aspects of proficiency, which could be implemented in many different ways depending on the particulars of the testing contexts.

Although each assessment application is to some degree unique in its contents, purposes, and contexts, there are certain principles and relationships that all will share simply because all are assessments. For this reason one may gain advantage by embedding these principles in processes and knowledge representations. Architect Christopher Alexander and colleagues (Alexander, Ishikawa, & Silverstein, 1977) coined the term *design pattern* in the mid-1970s. A *design pattern* is the core of a solution to a problem that occurs repeatedly in our environment — but at a level of generality that the approach can be applied in many situations while adapting to the particulars of each case. The same idea was adapted by software engineers to help designers tackle programming problems that recur in different guises (Gamma et al., 1994). For these engineers, *design patterns* provide structured insights into conceptual problems and solutions above the level of specific programming languages and implementation environments.

Analogous forms called assessment *design patterns* were developed by Mislevy et al. (2003) to support the design of tasks for assessing science inquiry in the Principled Assessment Designs for Inquiry (PADI) project. Like designing tests of communicative competence, designing science inquiry tasks is a challenge to standard assessment development practice (i.e., inquiry is regarded in the assessment community as a construct that is hard to assess). It calls for extended performances, cycles of hypothesis generation and testing, and, often, technologies such as automated scoring and computer-based simulation environments. *Design patterns* provide assessment designers with a high-level approach to tackle challenging issues by scaffolding the thinking that must precede the particular technical decisions required in the development of the actual tasks, identification of psychometric models, and articulation of decision rules required for scoring tasks. Assessment *design patterns* organize information about the targeted proficiencies, performance, and use situations in terms of the structure of assessment arguments. They serve as an in-between layer that connects the content of an assessment argument to the structure of the argument.

In particular, each *design pattern* builds around the general form of an assessment argument, concerning the knowledge or skill one wants to address (examples in science inquiry include model-based reasoning and designing experiments), the kinds of observations that can provide evidence about acquisition of this knowledge or skill, and the features of task situations that allow students to provide this evidence. Explicating the assessment structure in a narrative form with slots to be filled, *design patterns* arrange an underlying assessment argument into attributes that can subsequently be instantiated in particular operational tasks. Because the structure of a *design pattern* implicitly contains the structure of an argument in general, and an assessment argument in particular, filling in the *design pattern* slots simultaneously renders explicit the relationships among the pieces of the *design pattern* attributes in terms of the roles they play in argumentation based on Messick's components (see Table 2, adapted from Mislevy and Haertel, 2006). Assessment designers working with the PADI design system use the web-based design interface illustrated for *design patterns*. (See Design Pattern Template in Figure 1.)

Table 2: Design Pattern Attributes, Definitions & Corresponding Messick Argument Components

Design Pattern Attribute	Attribute Definition	Messick Assessment Argument Component
Rationale	The connection between the focal KSAs and what people do in what kinds of circumstances	**Student Model/Claim** What construct (complex of student attributes) should be assessed?
Focal Knowledge, Skills & Abilities	The primary KSAs targeted by the Design Pattern	
Additional Knowledge, Skills & Abilities	Other KSAs that may be required by tasks written using this Design Pattern	
Potential Work Products	Possible things one could see students say, do, or make that would provide evidence about the KSAs	**Evidence Model/Actions** What behaviors should reveal the construct?
Potential Observations	Features of the things students say, do, or make that constitute the evidence	
Characteristic Task Features	Aspects of assessment situations that are necessary in some form to elicit desired evidence	**Task Model/Situation** What tasks should elicit those behaviors?
Variable Task Features	Aspects of assessment situations that can be varied in order to shift difficulty or focus	
Narrative Structures	Describe overall storyline of prompt(s); Help to categorize and may help to generate tasks	

* Note. Narrative structures are unique to PADI *design patterns* associated with this PADI project.

Figure 1: Design Pattern Template

Blank Design Pattern Template | Design Pattern 2236 [View Tree | View Horiz | Duplicate | Permit | Export | Delete]

Title	[Edit]	Blank Design Pattern Template
Overview	[Edit]	
Use	[Edit]	
Focal Knowledge, Skills, and Abilities	[Edit]	
Additional Knowledge, Skills, and Abilities	[Edit]	
Potential observations	[Edit]	
Potential work products	[Edit]	
Potential rubrics	[Edit]	
Characteristic features	[Edit]	
Variable features	[Edit]	
Narrative Structure	[Edit]	
State Benchmarks	[Edit]	
I am a kind of	[Edit]	
These are kinds of me	[Edit]	
These are parts of me	[Edit]	
National Educational standards	[Edit]	
Templates	[Edit]	
Exemplar tasks	[Edit]	
Online resources	[Edit]	
References	[Edit]	

The NSF-funded project, "Application of Evidence-Centered-Design to a State's Large-Scale Science Assessment" (DRL-0733172) is supported within the NSF Discovery Research K-12 initiative. The project is designed to explore opportunities to leverage principles and structures from ECD in the context of the Science Minnesota Comprehensive Assessment for the middle-school level.

A high level overview of the Science Minnesota Comprehensive Assessment development process is depicted in Table 3. It begins with Storyboard Development and culminates in the Operational Test Administration. This development process is informed by Minnesota Department of Education's (MDE's) Guidelines for Test Construction and Minnesota Comprehensive Assessment Test Specifications for Science.

Table 3: Overview of the Science Minnesota Comprehensive Assessment Development Process

Development Stage	Assessment Development Activity
Stage 1	**Storyboard Development** - Authored by MN science teachers and processed by Pearson - Reviewed by MDE content and bias advisory panels - Selected for development
Stage 2	**Item Development** - Authored by MN science teacher and processed by Pearson - Reviewed by MDE content and advisory panels - Selected for field test
Stage 3	**Electronic Development** - Animations/audio created; items converted to electronic format - FR items programmed; all items reviewed in preview applications
Stage 4	**Field Test Administration** - Scenarios/items embedded in operational test - ECR/SCR items reviewed by range finding advisory panel - Items scored; reviewed by data review advisory panel - Items selected for operational test
Stage 5	**Operational Test Administration** - Scenarios/items embedded in operational test - Items scored and forms equated - Standards set by MDE advisory panel - Items released to public or reserved for future operational use

We have identified storyboard development and item design and development to be leverage points in the assessment development process. Leverage points can be defined as those opportunities and processes that can be refined, using the lens and data structures of ECD, in order to streamline the assessment design, development, and delivery processes (Snow, Haertel, Mislevy, Fulkerson, Feng, & Nichols, 2010).

Storyboard development and item development are leverage points that we can impact by applying ECD *design patterns*. Our expectation is that the newly developed *design patterns* will support the Minnesota Comprehensive Assessment and improve the efficiency of storyboard and item development, assist writing teams address hard-to-assess benchmarks, and make explicit the validity argument. We also expect that being able to view individual storyboards and items as instances motivated by *design patterns* will help storyboard and item review teams before assessments move to production and after they are implemented electronically.

DEVELOPING A HYBRID DESIGN PATTERN: MODEL USE IN INTERDEPENDENCE AMONG LIVING SYSTEMS

The first PADI project (DRL- 0733172) focused on creating *design patterns* to guide the design of tasks for science inquiry and, as such, served as suitable starting points for the present *design pattern* work. We continued this line of work in the current project by developing *design patterns* for hard to assess science topics. In order to fully achieve project expectations of *design pattern* use by storyboard and item writers, we soon realized the need to support storyboard and item writers in making connections between state benchmarks and *design patterns*. As a result, we are investigating a process for adapting *design patterns* to be more content-specific, while still functioning as a framework for developing families of tasks related to both scientific content and inquiry or reasoning. Our first content-inquiry hybrid *design pattern* incorporates benchmarks aligned with interdependence among living systems with the core components specified in the Model Use *design pattern* (Mislevy, Riconscente, & Rutstein, 2009).

DESIGN PATTERN AND CONTENT IDENTIFICATION PROCESS. In deciding where to begin with creating this hybrid *design pattern*, we first selected which inquiry *design pattern* to use. We agreed that model-based inquiry in science was an important and difficult-to-assess skill and reviewed the *design patterns* associated with model-based reasoning. A brief description of each *design pattern* is provided in Table 4. We then conducted a preliminary mapping of each *design pattern* to middle school physical, life, and Earth science content benchmarks in the 2009 Minnesota Academic Standards in Science. This mapping revealed which benchmarks have potential to address knowledge, skills, and abilities identified in the model-based reasoning *design patterns*. Table 4 also shows the number of middle school benchmarks that appear to be related to each of the model-based reasoning *design patterns*.

We considered our own domain knowledge expertise when deciding whether to target physical, life or Earth science in this hybrid *design pattern*. Dr. Snow has previously taught middle school life science, and so we focused on the life science benchmarks. Within life science, we believed that knowledge and skills associated with “Model Use” would be worth further explication in a *design pattern*, not only because several of the benchmarks address components of Model Use, but also because Model Use may be assessed in particularly interesting ways with Minnesota’s technology-based assessment environment.

We made an explicit decision that the grain size of this hybrid *design pattern* should be at the Substrand level so that the *design pattern* could inform the design of tasks associated with one or more benchmarks. We believed that a *design pattern* focused at the benchmark level may be too narrow for the storyboard and item writing purposes of this project. Thus, we selected Interdependence among Living Systems as a focal Substrand of interest for the *design pattern*, entitled *Model Use in Interdependence among Living Systems*.

Table 4: Suite of Design Patterns (DPs) for Model-Based Reasoning

Title	Description	# of Matched Benchmarks*
Model Articulation	Tasks supported by this DP assess students' ability to articulate the meaning of physical or abstract systems across multiple representations. Representations may take qualitative or quantitative forms. This DP is relevant in models with quantitative and symbolic components (e.g., connections between conceptual and mathematical aspects of physics models). Model Articulation is often be pertinent in multiple-step tasks, after the Model Formation step.	58
Model Elaboration	This DP supports developing tasks in which students elaborate given scientific models by combining, extending, adding detail to a model, and/or establishing correspondences across overlapping models. This DP can be considered a special case of Model Formation in that the aim is to develop a modeled conception of a situation. The emphasis is what is happening in the model layer with respect to extensions of models or connections between models. Model Elaboration is also similar to Model Revision, in that a given model or a set of unconnected models does not account properly for the target situation and reformulation is required.	4
Model Evaluation	This DP supports developing tasks in which students evaluate the correspondence between a model and its real-world counterparts, with emphasis on anomalies and important features not accounted for in the model. This DP is tied closely with Model Use, and is also associated with Model Revision and Model Elaboration.	16
Model Formation	This DP supports developing tasks in which students create a model of some real-world phenomenon or abstracted structure, in terms of entities, structures, relationships, processes, and behaviors. The Model Formation DP can be viewed as a subpart of the Model-Based Inquiry *design pattern*, and many tasks combine Model Formation with Model Use. The Model Formation DP also overlaps with those for Model Elaboration and Model Revision.	40
Model Revision	This DP supports developing tasks in which students revise a model in situations where a given model does not adequately fit the situation or is not sufficient to solve the problems at hand. Because its centrality, Model Revision is difficult to assess in isolation from other aspects of Model-Based Reasoning. Model Revision is prompted only by Model Evaluation, and then Model Formation must be used to propose alternatives or modifications.	1
Model Use	This DP supports developing tasks that require students to reason through the structures, relationships, and processes of a given model. Model Use is often combined with Model Formation in the same tasks, and most tasks that address Model Evaluation and Model Revision also involve Model Use.	53

*Many benchmarks related to multiple Model-Based Reasoning *design patterns*.

FOCAL KSAS

As in the original Model Use *design pattern*, we continue to conceive of Model Use as reasoning with models. All of these benchmarks have the potential to require students use models to reason about and explain what they know about ecosystems and energy flow. Ecosystems are models of populations and these benchmarks indicate that students need to demonstrate that they can effectively describe and reason about the relationships in ecological models. The Focal KSAs reflect a restatement of the benchmarks to identify the desired ability or skill with an emphasis on Model Use (see Table 5). That is, these Focal KSAs require that the benchmarks be assessed with explicit interaction of some sort with an ecological model.

Like the benchmarks, the Focal KSAs are written at a general level so that student knowledge and skills can be assessed in relation to a variety of specific ecosystems or models (e.g., food web, energy flow). As a result, for most benchmarks, there is a one-to-one correspondence between benchmark and Focal KSA. The exception is the first benchmark (MCA III: 7.4.2.1.1. Identify a variety of populations and communities in an ecosystem and describe the relationships among the populations and communities in a stable ecosystem). This benchmark is associated with two KSAs, each of which relates to a component of the benchmark (i.e., the ability to *identify* a variety of populations and communities and the ability to *explain relationships among* populations and communities). Table 5 shows the relationships among Benchmarks, Focal KSAs, and Potential Observations.

POTENTIAL OBSERVATIONS AND POTENTIAL WORK PRODUCTS

Potential Observations help to make each Focal KSA more concrete by describing the evidence (in the form of a specific student behavior) that indicates that a student has the intended knowledge, skill or ability. Potential Work Products are descriptions of the form of the information that can be gathered from students (e.g., written explanation or selection of a response).

POTENTIAL OBSERVATIONS.

As illustrated in Table 5, the Potential Observations reflect more concretely the behaviors that may be observed by students if they have attained the desired level of proficiency for each Focal KSA. In all of these Potential Observations, correct model use is inferred by students' correct identification of elements in the model or high quality explanations about the phenomena represented in the model. For example, if a student can provide an accurate and complete explanation of how entities in a model are related, it indicates that the student understands how to use and reason with the model.

Similar to the benchmarks and Focal KSAs, these Potential Observations are general statements of student behaviors. They provide the storyboard developer or item writer with enough information to think about how high levels of student knowledge and skill may be communicated, and it is fairly easy to think about how they might be applied in the context of a specific ecological model.

WORK PRODUCTS. Student explanations may be gathered in a variety of ways, as indicated in the Potential Work Products. Work Products for this *design pattern* include:

- Selection of hypotheses, predictions, retrodictions, explanations, and/or missing elements of real world situation
- Constructed hypotheses, predictions, retrodictions, explanations, and/or missing elements of real world situation, via creation of one or more representational forms; filling in given, possibly partially filled- in, representational forms
- Intermediate products developed in selection/construction of hypotheses, predictions, explanations, and/or missing elements
- Written/oral explanation of the hypotheses, predictions, explanations, and/or missing elements.
- Trace of actions taken in solution
- Talk- aloud of solution
- Critique of a given solution
- Completion and description of a flow chart showing how energy flows in an ecological model.
- Completion of a flow chart showing how matter flows in an ecological model
- Description of how producers in an ecological system use the energy from sunlight to make sugars from carbon dioxide and water in a process called photosynthesis

The appropriateness of a given Work Product depends on the purpose of the assessment and on the nature of the explanation one intends to observe. For example, selections and constructions of predictions and explanations and descriptions missing elements in models may be appropriate in both large-scale and classroom assessment contexts. Oral explanations may be most useful as part of a formative classroom assessment. Talk-alouds are difficult to manage in classroom and large-scale assessment contexts but may be useful if an examiner has an opportunity to administer a task individually with a single student.

Table 5: Connections Among Benchmarks, Focal KSAs, and Potential Observations

Benchmark	Focal KSA	Potential Observation
MCA III: 7.4.2.1.1. Identify a variety of populations and communities in an ecosystem and describe the relationships among the populations and communities in a stable ecosystem	FK1. Ability to use an ecological model to explain the relationships among populations and communities	High quality explanation* of how communities and populations represented in an ecological model interact
MCA III: 7.4.2.1.2. Compare and contrast predator/prey, parasite/host, producer/consumer/decomposer relationships	FK2. Ability to use an ecological model to explain similarities and differences among types of interdependent relationships	High quality explanation* of how one or more interdependent relationships represented in an ecological model are similar to or different from other interdependent relationships represented in the model
MCA III: 7.4.2.1.3. Explain how the number of populations an ecosystem can support depends on the biotic resources available as well as abiotic factors such as amount of light and water, temperature range and soil composition	FK3. Ability to use an ecological model to explain how populations in an ecosystem are dependent on biotic and abiotic resources	High quality explanation* of how populations represented in an ecological model are dependent on the biotic and abiotic resources shown in the model
MCA III: 7.4.2.2.1. Recognize that producers use the energy from sunlight to make sugars from carbon dioxide and water through a process called photosynthesis. This food can be used immediately, stored for later use, or used by other organisms	FK 4. Ability to use an ecological model to explain how producers make, use, and store food	High quality explanation* of how producers make, use, and store food in an ecological model
MCA III: 7.4.2.2.2. Describe the roles and relationships among producers, consumers and decomposers in changing energy from one form to another in a food web within an ecosystem	FK5. Ability to use an ecological model to explain how energy changes form in a food web	High quality explanation* of how energy represented in an ecological model, such as a food web, changes form.
		Accurate completion and description of a flow chart showing how energy flows in an ecological model
MCA III: 7.4.2.2.3. Explain that the total amount of matter in an ecosystem remains the same as it is transferred between organisms and their physical environment, even though its form and location change	FK6. Ability to use an ecological model to explain how the amount of matter stays the same as it is transferred between organisms and their physical environment	High quality explanation* of how the amount of matter stays the same as it is transferred between the organisms and components of the physical environment shown in an ecological model
		Accurate completion and description of a flow chart showing how matter flows in an ecological model

Note (*). Explanations can, depending on the assessment context, take many different forms (i.e., explanations can be represented in a variety of Work Products). A high quality explanation is accurate, relevant, complete, organized/systematic, and efficient.

CHARACTERISTIC FEATURES

The tasks resulting from this *design pattern* must have certain characteristics to elicit these types of explanations and reasoning with ecological models. The essential components of tasks associated with this *design pattern* are reflected in the Characteristic Features. All tasks intending to elicit this constellation of Focal KSAs and associated Potential Observations must first include ecological models that represent real-world situations. Ecological models that reflect real-world situations are important because they represent a situation that is scientifically accurate, which is important for science assessments. Real-world situations also tend to be complex, and these kinds of complexities are necessary for students to be able to develop the high-quality explanations described in the Potential Observations.

A second Characteristic Feature of these tasks is that they present an ecological model appropriate to the situation. Ecological models must be present for students to reason with and explain them. Although this seems fairly obvious, it illustrates one of the functions of *design patterns*—to make explicit all of the design principles applied in the development of an assessment task.

A third Characteristic Feature of tasks associated with *design patterns* is that questions require students to reason through the schema and relationships in the model. Questions must go beyond simple identification of components of an ecological model.

ADDITIONAL KSAS

Additional KSAs are knowledge, skills, and abilities that are not the target of the assessment, but may be required in tasks that provide evidence of students' proficiency on the Focal KSAs. To develop the Additional KSAs for this *design pattern*, we started with the Additional KSAs in the general Model Use *design pattern* and made them relevant in the context of ecological model use (see Table 6). Table 6 also describes how a particular Additional KSA or set of Additional KSAs is significant in the context of ecological Model Use. For example, domain area knowledge, such as knowledge of what a population is, is needed for students to be able to reason about relationships among populations represented in a particular ecological model, but tasks resulting from this *design pattern* would not inquire directly about students' understanding of the concept of population.

A subset of Additional KSAs is related to students' experience with and familiarity with models. Some of these Additional KSAs may be supported or not in tasks. For example, some tasks may inform students about whether a particular model is stable or unstable. Other tasks may require students to figure out for themselves whether a model is stable in order to answer questions. Likewise, students may be explicitly directed about how to use a model to determine interdependencies among entities. In other tasks, this skill will not be scaffolded.

Table 6: Relationship between Additional KSAs in Model Use Design Pattern and Model Use in Interdependence among Living Systems Design Pattern

Additional KSAs in *Model Use* Design Pattern	Additional KSAs in *Model Use in Interdependence among Living Systems* Design Pattern	Significance of Additional KSAs
Domain area knowledge (declarative, conceptual, and procedural)	Knowledge of what a population is	Students need to have a fundamental understanding about the components represented in an ecological model in order to be able to use and reason about their relationships in the model.
	Knowledge of what a community is	
	Knowledge of what an ecosystem is	
	Ability to recognize producers in a food web	
	Ability to recognize consumers in a food web	
	Ability to recognize decomposers in a food web	
	Ability to distinguish between biotic and abiotic resources in an ecological model	
Familiarity with real-world situation	Knowledge of entities (e.g., plants and animals) represented in the ecological model	Familiarity and knowledge of the specific kinds of entities in an ecological model are important for being able to interpret the model. At a minimum, students may need to recognize which entities are plants or animals.
Knowledge of model at issue	Knowledge of different ecological models (e.g., food webs, water cycle)	Students need basic knowledge of the type of ecological model used in a task.
	Ability to recognize whether an ecosystem is stable	Recognition of whether an ecosystem is stable is important in how students use and interpret ecological models.
	Understanding that when two entities are related or interdependent, manipulating one will affect the other	Using models requires that students understand that entities in a model can be related or unrelated, and what it means for two entities to be related.
	Ability to determine interdependencies in a model by holding constant some entities while varying others	If model use requires model manipulation, students need to understand how to run proper experiments, by manipulating one entity and holding others constant.
Familiarity with symbolic representations and procedures	Knowledge of required symbolic representations associated procedures (e.g., chemical equations, mathematical notation)	Students need to be familiar with the symbolic representations and how they function in models in order to be able to reason about and use the ecological models (e.g., arrow direction in a food web or chemical notation related to photosynthesis).
Familiarity with modeling tool(s)	Knowledge of how to use and interpret required modeling tool(s) (e.g., online state assessment interface, STELLA, ESIS)	If model use requires manipulation of parameters, then students need to be familiar with the components and output of the modeling tool.
Familiarity with task type	Familiarity with materials, protocols, and expectations	Familiarity with the types of tasks and expectations is important for students working on any assessment task.

VARIABLE FEATURES

Assessment designers need to decide whether to scaffold or support these Additional KSAs that may be required for students to demonstrate evidence of the Focal KSAs. The decisions about task design are identified as Variable Features. Variable Features are ways of describing how adjustments can be made to the design of a task to shift difficulty or focus of the task. Table 7 shows the relationship between Additional KSAs and Variable Features. For example, consider the set of Additional KSAs associated with domain area knowledge (e.g., Knowledge of what a population is, Knowledge of what a community is, Knowledge of what an ecosystem is). To influence the extent to which this knowledge is required in a model use task, a task designer may consider whether it is appropriate to present background information about the ecological model, provide definitions of key terms, describe entities represented in the model, and include other visual and linguistic supports to assist students in recall and recognition of elements of the model.

In considering how to implement these Variable Features, it is important to note that several Variable Features potentially relate to multiple Additional KSAs. For instance, visual and linguistic supports, depending on how they are used, can influence task demands associated with domain area knowledge, familiarity with the real-world situation, knowledge of the model of issue, and familiarity with symbolic representations and procedures. These relationships among Additional KSAs and Variable Features illustrate the complexities involved with designing appropriately challenging tasks to assess students' ability to use ecological models.

Table 7: Variable Features to Adjust the Demands of Additional KSAs

	Variable Features
Ways to Vary Task Design to Influence the Demand of **Domain Area Knowledge**	Presentation of background about the ecological model
	Provision of definitions of terminology relevant to ecological model
	Provision of descriptions of entities in an ecological model
	Use of visual and linguistic supports in model
Ways to Vary Task Design to Influence the Demand of **Familiarity with the Real-World Situation**	Problem context/Type of ecological model
	Use of visual and linguistic supports in model
	Degree of scaffolding provided
	Presentation of background about the ecological model
	Provision of definitions of terminology relevant to ecological model
	Provision of descriptions of entities in an ecological model
Ways to Vary Task Design to Influence the Demand of **Knowledge of the Model at Issue**	Problem context/Type of ecological model
	Use of visual and linguistic supports in model
	Degree of scaffolding provided
	Presentation of background about the ecological model
	Provision of definitions of terminology relevant to ecological model
	Provision of descriptions of entities in an ecological model
	Complexity of reasoning required
	Complexity of model
	Relative stability of ecological model
Ways to Vary Task Design to Influence the Demand of **Familiarity with Symbolic Representations and Procedures**	Problem context/Type of ecological model
	Use of visual and linguistic supports in model
	Degree of scaffolding provided
	Complexity of reasoning required
Ways to Vary Task Design to Influence the Demand of **Familiarity with Modeling Tool(s)**	Problem context/Type of ecological model
	Model provided to or generated by student
	Data provided to or generated by student
	Degree of scaffolding provided
Ways to Vary Task Design to Influence the Demand of **Familiarity with Task Type Requirements**	Problem context/Type of ecological model
	Complexity of model
	Degree of scaffolding provided
	Complexity of reasoning required

NARRATIVE STRUCTURES

Narrative Structures are:

> . . . general patterns or preconceived frameworks that serve as reusable plotlines for storyboards. They may be used as a type of advance organizer, aiding storyboard writers in the collection and organization of ideas and information prior to and during the storyboard writing process. They are potentially useful in the construction of a storyboard outline. (Fulkerson, Nichols, Haynie, Mislevy, 2009).

Narrative Structures are one of the explicit links between the *design pattern* and the storyboard. Benchmarks assigned to a storyboard writer cue the writer into which Focal KSAs may be appropriate. Given the underlying processes required by the Focal KSAs (e.g., model use), particular Narrative Structures may be most fruitful in supporting writers in eliciting these understandings implied in the benchmarks and Focal KSAs. In the case of the Model Use in Interdependence among Living Systems *design pattern*, there are four relevant narrative structures: Cause and Effect; Change over Time, General to Specific or Whole to Parts, and Specific to General or Parts to Whole. These Narrative Structures are described in Table 8 with examples related to ecological Model Use.

Table 8: Narrative Structures for Model Use Design Patterns

Narrative Structure	Description	Example
Cause and Effect	An event, phenomenon, or system is altered by internal or external factors.	Changing environmental pressures influence adaptations of organisms
Change over Time	A sequence of events is presented to highlight sequential or cyclical change in a system.	Influence of global warming on glacial melt, rising oceans, species adaptation/extinction
General to Specific or Whole to Parts	A general topic is initially presented followed by the presentation of specific aspects of the general topic.	Introduce the process of photosynthesis, followed by a more specific discussion of the role of chlorophyll in the process
Specific to General or Parts to Whole	Specific characteristics of a system or phenomenon are presented, culminating in a description of the system or phenomenon as a whole.	Introduce the different parts of a flower (e.g., stigma, style), followed by a more general discussion of their roles in the process of reproduction

APPLICATIONS AND CONCLUSIONS

The Model Use in Interdependence among Living Systems *design pattern* is a specific type of *design pattern* that articulates an assessment argument integrating scientific reasoning with content knowledge—in this case, Model Use with ecological systems. We see this content-inquiry hybrid *design pattern* as a critical addition to the family of *design patterns* for this project.

GENERALIZABILITY OF HYBRID DESIGN PATTERNS

One concern about content-inquiry hybrid *design patterns* may be the extent to which they are broad enough to facilitate the generation of families of tasks, as compared with inquiry-only *design patterns*. We believe that hybrid *design patterns* can be quite generative in the task design process. However, the range of tasks resulting from these *design patterns* will be situated in a particular domain or subset of a domain. As illustrated with the Model Use in Interdependence among Living Systems *design pattern*, there is still a range of Focal KSAs to assess and Potential Observations to observe as evidence and Potential Work Products to gather from students depending on the purpose of the assessment (e.g., formative or summative). Moreover, the *design pattern* illustrates multiple ways to vary task difficulty and supports with Variable Features, further illustrating the many different kinds of tasks that can result from this *design pattern*.

HYBRID DESIGN PATTERNS TO STREAMLINE LARGE-SCALE ASSESSMENT DEVELOPMENT

Content-inquiry hybrid *design patterns* may have some advantages in large-scale assessment development as compared with *design patterns* that focus only on content knowledge or scientific inquiry/reasoning skills. Content-only *design patterns* may be too narrowly focused on particular knowledge in a domain, without supporting storyboard designers in conceptualizing content benchmarks in ways that incorporate hard-to-assess and higher-order thinking skills. Inquiry-only *design patterns* are advantageous because they articulate ways to design tasks and find evidence for difficult to assess skills in science. However, that which give them an advantage (i.e., broadly-defined ways to assess important scientific inquiry and reasoning skills) may turn out to be a limitation for storyboard and item writers who need to use specific benchmarks to generate multiple storyboards and items rapidly in the large-scale state test development cycle.

To use “inquiry-only” *design patterns*, storyboard and item writers need to be sophisticated users of both benchmarks and design patterns. Writers need to be able to unpack content-focused benchmarks in terms of scientific inquiry/reasoning processes as well as analyze the components of the different inquiry *design patterns* to find one that is a good match for a benchmark or set of benchmarks. Even if this “matching” is done ahead of time for writers by the design management team, and writers are assigned benchmarks along with specific inquiry-only *design patterns*, there is still much work to be done to determine how to translate inquiry-only Focal KSAs, Additional KSAs, Potential Observations, Potential Work Products, Characteristic Features, and Variable Features in ways that make sense for the assigned set of benchmarks. In addition, the writer must select among and apply one or more Narrative Structures and identify a storyboard context that allows for the assessment of multiple benchmarks.

Advanced storyboard and item writers may be ready for the additional layer of reasoning about assessment design required in using *design patterns*. Intermediate and novice storyboard writers, who need the most support in this process, may be overwhelmed. As a result, less sophisticated writers may only attend minimally to *design patterns*, thus reducing the likelihood of *design patterns* having much influence, if any, in the storyboard and item development process. If training permitted, we may be able to help newer writers think about ways to apply inquiry *design patterns* in storyboard and item writing. Unfortunately, training time for writers in most state assessment development situations is limited.

To streamline the storyboard and item development process for large-scale assessment, it is important to consider how we support all writers (novice, intermediate and advanced) in using *design patterns*. Hybrid *design patterns* may serve almost as a scaffold for writers new to storyboard/item writing and *design patterns*. In hybrid *design patterns*, the link between benchmarks and *design patterns* can be quite explicit, as illustrated in the Model Use in Interdependence among Living Systems *design pattern*. For example, we clearly indicate how benchmarks align to Focal KSAs. Because Focal KSAs are linked to Potential Observations, it is clear to writers the kinds of behaviors that would provide evidence of the Focal KSAs. In addition, this *design pattern* articulates characteristic features of tasks on ecological Model Use and ways to vary tasks to adjust the difficulty and focus. While feel strongly that inquiry-only *design patterns* are an important starting point, we also believe that the burden of linking these *design patterns* with content benchmarks may need to be placed on the design management team in the form of hybrid *design patterns*.

One of the challenges in using hybrid *design patterns* is that it may require the design management team to create more *design patterns* than otherwise might have been created if an inquiry-only *design pattern* approach were used. The trade-offs of producing more content-inquiry integrated *design patterns* by the design management team as compared with requiring writers to use fewer inquiry-only *design patterns* may be weighed differently based on the goals of a particular assessment context, the expertise of the design management team, the expertise of the storyboard and item writing team and resources available for training in the use of *design patterns* for storyboard and item writing.

SUMMARY

As this project moves further in the use of content-only, inquiry-only, and hybrid *design patterns*, it will be important to capture from all perspectives (i.e., the design management team, storyboard writers, and item writers) the value-added of different types of *design patterns* in large-scale assessment design. Regardless, the process of developing these different types of *design patterns* continues to illustrate the flexibility of PADI *design patterns* to address any and all types of assessment arguments. Thus, the burden remains on the design team for determining which type or types of *design pattern* is appropriate given the constraints of the assessment design situation.

REFERENCES

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

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

Fulkerson, D., Nichols, P., Haynie, K., Mislevy, R. (2009). *Narrative Structures in the Development of Scenario-Based Science Assessments (Large-Scale Assessment Technical Report 3).* Menlo Park, CA: SRI International.

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

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

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

Mislevy, R. J., Riconscente, M. M., & Rutstein, D. W. (2009). *Design Patterns for Assessing Model-Based Reasoning (Large-Scale Assessment Technical Report 6).* Menlo Park, CA: SRI International.

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

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

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

Tillers, P., & Schum, D. A. (1991). A theory of preliminary fact investigation*. U.C. Davis Law Review, 24*, 907-966.

!

Sponsor
The National Science Foundation, Grant DRL - 0733172

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

Subgrantees
University of Maryland
Pearson