

**Design Patterns to Support Storyboards  
and  
Scenario-Based, Innovative Item Types**

Prepared for:

2011 Annual Meeting of the American Educational Research Association (AERA)  
April 9<sup>th</sup> – 12<sup>th</sup>, New Orleans, LA

Prepared by:

Eric Snow, SRI International  
Dennis Fulkerson, Pearson  
Mingyu Feng, SRI International  
Paul Nichols, Center for Assessment

---

*Acknowledgments*

This material is based on work supported by the National Science Foundation under grant DRL-0733172 (ECD Large-scale Grant).

*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.

## Abstract

Tasks that measure inquiry-based constructs, such as model-based reasoning and systems thinking, often require students to complete multiple, cognitively demanding steps, frequently in the context of a larger scenario meant to reflect a “real-world” situation. Designing these tasks for use in large-scale assessment requires task writers to integrate multiple content and inquiry targets across a series of scenes, each containing multiple embedded items, and to do so in a way that the resulting scores contain little measurement error and support valid inferences about student performance. *Design patterns* are research-based, representational forms that can help assessment developers organize and integrate information about tasks in terms the focal skills to be measured, the behaviors that reveal the focal skills, and the task features that will elicit the desired behaviors. In this paper we report on current research in which we are examining the extension and application of inquiry-based design patterns to guide the development of complex science tasks for use in a statewide science assessment. As an example, we describe the use of a design pattern representing the practice of model revision to develop a task asking students to consider the introduction of an invasive species into a new ecosystem and to revise a food web to accommodate for the introduction of the new species.

## Introduction

The development and application of the evidence-centered assessment design (ECD) framework (Mislevy, Steinberg, & Almond, 2003) is a research program that aims to make the work of assessment design and development, particularly for large-scale tests, more efficient and potentially more valid than current practices. ECD has been applied variously at Educational Testing Service (ETS; Pearlman, 2001), Cisco Systems (Behrens, Mislevy, Bauer, Williamson, and Levy, 2004), and the IMS Global Learning Consortium (2000). Additionally, ECD was used as the foundation for the Principled Assessment Design for Inquiry project (PADI; Mislevy & Haertel, 2006; Baxter and Mislevy, 2004) and, more recently, to guide the development and revision of Advanced Placement Exams at the College Board (Huff & Plake, 2009). *Design patterns* are ECD-based knowledge representations for organizing assessment arguments at a narrative level (e.g., Kane, 1992), and 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 (i.e., statewide/summative, or classroom/formative).

This paper extends earlier work that focused on “leveraging” evidence-centered design (ECD), particularly *design patterns*, in large-scale test design, development and delivery processes (e.g., Snow, Haertel, Fullkerson, Feng, & Nichols, 2010; Haertel & Mislevy, 2008a; Mislevy 2007; Mislevy & Haertel, 2006; Mislevy, Steinberg, Almond, Haertel, & Penuel, 2003). The next section provides background information on the Minnesota Comprehensive Assessment (MCA) framework, the MN statewide science assessment and the NSF-DRK-12 project from which examples reported on in this paper are drawn. The third section describes design patterns in more detail, particularly the roles they can play in supporting operationalization of the assessment argument during item design and development. The fourth section describes an example of using design patterns to guide the development of complex science assessment tasks in the context of the MN statewide science assessment. The closing discussion will summarize the value of design patterns for supporting the development of scenario-based assessment tasks, and highlight a specific extension to standard design patterns that has proved to be particularly useful in supporting the design and development of storyboards for scenario-based assessment tasks.

## Background

### **MCA Framework & Minnesota Statewide Science Assessment**

The Minnesota Department of Education annually administers the Minnesota Comprehensive Assessments (MCA) for the purpose of monitoring student growth toward achieving the Minnesota Academic Standards. The MCA Science test is administered to students at grades 5, 8, and once in high school.

The MCA Science test is scenario-based and computer-delivered, presenting students

with realistic representations of classroom investigations and real-world phenomena in an online environment. Each test scenario is divided into three to six short segments, called scenes, with each scene including embedded graphic, audio, and/or video media and associated items. Items within a scenario are related by the common scenario context and give students the opportunity to apply their content knowledge and process skills in a simulated situation. Item types include multiple-choice, constructed response, and figural response. Figural response items require students to manipulate graphic elements of an item, to select a point on a graphic, or to complete a graph or diagram. The three basic types of figural response items include hot spot, drag and drop, and graphing. Each item is aligned to one MCA Science benchmark and one Depth of Knowledge level (Webb 1999).

The MCA Science tests at grades 5 and 8 cover four content strands: Nature of Science and Engineering, Physical Science, Earth and Space Science, and Life Science. The high school MCA Science test includes two content strands: Nature of Science and Engineering and Life Science. Scenarios may be interdisciplinary in nature, consisting of content drawn from multiple science strands. A typical MCA Science operational test form consists of seven to nine scenarios with 40-60 items, depending on grade level. Field test scenarios and items are embedded in the operational forms.

Each MCA Science test scenario originates as a storyboard. Storyboards are authored by practicing Minnesota science teachers who have been trained in storyboard writing. Each storyboard includes scripts for each scene, written descriptions of each scene, sample artwork for each scene, and lists of supported benchmarks that may be assessed with each scene. Once a storyboard has been written, approved, and selected for field testing, items and electronic media are developed and integrated, resulting in a complete scenario.

## **DRK-12**

Many states and their contractors have turned to computer-supported assessment task development and delivery to help them meet the challenges posed by large-scale assessment. It is widely accepted that more complex, multipart assessment tasks are better suited to measuring higher-level skills, but cost considerations and incompatibility with conventional test development and implementation practices stand in the way of large-scale use. However, technology-based assessment tasks, such as simulations and investigations, have proved difficult and costly to develop under procedures that evolved for conventional multiple-choice item development (Riconscente et al., 2005). The NSF-funded project, “Application of Evidence-Centered-Design to a State’s Large-Scale Science Assessment” (Haertel & Mislevy, 2008b) is supported within the DR K-12 initiative. The project is designed to explore opportunities to leverage principles and structures from ECD in the context of the *Minnesota Comprehensive Assessment, Science Assessment (MCA-II)*, in the area of middle-school science.

The targeted types of assessment tasks we design address highly valued aspects of science education that are difficult to create and implement at scale (Means & Haertel, 2002). The conceptual approach and technical infrastructure builds on the NSF-supported Principled Assessment Designs for Inquiry (PADI) project. The PADI project developed broadly applicable

data structures, representational forms, and software tools to provide conceptual support to organize and coordinate the work of different experts at different stages of the process. In particular, PADI assessment design patterns (analogous to those in architecture and software engineering) capture design rationale in a reusable and generative form in the Domain Modeling layer of assessment. They help designers think through substantive aspects of an assessment argument in a structure that spans specific domains, forms, grades, and purposes (Mislevy, Hamel, et al., 2003). Design patterns, which we will discuss in more detail in the next section, center on the “big ideas” and pervasive skills in science, such as inquiry cycles and model-based reasoning, and serve as a foundation for templates for authoring more specific families of tasks.

The state of Minnesota’s web-based assessment system currently uses some tasks with features such as simulations to assess students in science at grades 5, 8, and high school. Students view simulations of experiments and phenomena and manipulate visual representations of relevant materials. These simulations and animations are embedded in scenarios that consist of multiple scenes that provide a context for the students when answering the test questions. The tasks developed in the project maintain the look, feel, and standards alignment of the operational MCA framework. The difference is not the tasks themselves but the conceptualization and processes by which they are produced and the design objects through which they are developed. These processes and objects facilitate developing new tasks. To establish the feasibility and effectiveness of ECD as instantiated in PADI as a practical solution, PEM, in conjunction with Minnesota’s Department of Education will provide expertise in assessment design and psychometrics. They will make accessible their resources to support assessment design and implementation and an opportunity to test the products (assessment tasks) of the proposed project within Minnesota’s existing testing system.

## Design Patterns

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 (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.

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 and testing, and, often, technologies such as automated scoring and computer-based simulation environments. Design patterns provide assessment designers with an high-level approach to tackle challenging issues by scaffolding the thinking that must proceed 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. Design patterns organize information about the targeted proficiencies, performance, and use situations in terms of the structure of assessment arguments (e.g., Kane, 1992). 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 complex of knowledge or skill one wants to address (Focal and Additional KSAs), the kinds of behaviors and products that can provide evidence about acquisition of this knowledge or skill (Potential Observations and Work Products), and the features of task situations that allow students to provide this evidence (Characteristic and Variable Task Features). 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 an assessment argument based on Messick's (1994) components (see Table 1, adapted from Mislevy and Haertel, 2006).

**Table 1: Design Pattern Attributes, Definitions & Corresponding Messick Argument Components**

<b>Design Pattern Attribute</b>	<b>Attribute Definition</b>	<b>Messick Assessment Argument Component</b>
Rationale	The connection between the focal KSAs and what people do in what kinds of circumstances.	<b>Student Model/Claim</b> What construct (complex of student attributes) should be assessed?
Focal Knowledge, Skills & Abilities (KSAs)	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	Some possible things one could see students say, do, or make that would provide evidence about the KSAs.	<b>Evidence Model/Actions</b> 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.	<b>Task Model/Situation</b> What tasks should elicit those behaviors?
Variable Task Features	Aspects of assessment situations that can be varied in order to shift difficulty or focus.	

## Using Design Patterns to Design Complex Science Assessment Tasks

In the DRK-12 project, we have developed design patterns for hard to assess science topics (e.g., conducting observational or experimental investigations; see Table 2), as well as a suite of design patterns organized around the concept of model-based reasoning (see Table 3).

**Table 2. Design Patterns for Hard to Assess Science Topics**

Title	Overview
Observational Investigation	This design pattern supports the writing of storyboards and items that address scientific reasoning and process skills in observational (non-experimental) investigations. In observational investigations, variables typically cannot be altered at all (e.g., objects in space) or in a short time frame (e.g., a lake ecosystem). This design pattern may be used to generate groups of tasks for any science content strand.
Experimental Investigation	This design pattern supports the writing of storyboards and items that address scientific reasoning and process skills in experimental investigations. In experimental investigations, it is necessary to manipulate one or more of the variables of interest and to control others while testing a prediction or hypothesis. This design pattern may be used to generate groups of tasks for science content strands amenable to experimentation.
Reasoning about Systems and Complexity	This design pattern supports the writing of storyboards and items that address reasoning within the context of complex systems. Complex systems are characterized interactions among components of the system and, typically, outcomes that emerge without any explicit driving force. The prevalence of complex systems across scientific content domains suggests the development of a design pattern that enables design of tasks that target students' reasoning about complex systems, across domains and, as a result, grade levels.
Genetics Learning Progression for Grades 5-10	This design pattern describes students' evolving knowledge of the characteristics and functions of genes. Its contents are based on a published journal article by Duncan, Rogat, and Yarden (2009) in which the authors posit a learning progression for deepening students' understandings of modern genetics across grades 5-10. This understanding of modern genetics is identified in the paper as consisting of understanding of the genetic model, the molecular model, and the meiotic model. The paper posits eight main ideas about the three models followed by the characteristics of what students in grade bands 5-6, 7-8, and 9-10 are capable of understanding about the main ideas respectively.
Model Use in Interdependence	This design pattern supports developing tasks that require students to

Among Living Systems	reason through the structures, relationships, and processes of ecological models. Use of ecological models is often combined with the formation of ecological models in tasks. Many tasks that address evaluation and revision of ecological models also involve the use of these models.
----------------------	---

**Table 3. Suite of Design Patterns for Model-Based Reasoning**

<b>Title</b>	<b>Overview</b>
Model Articulation	Tasks supported by this design pattern assess student's 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.
Model Elaboration	This design pattern 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 design pattern can considered a special case of model formation in that the aim is to develop a modeled conception of a situation, but 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.
Model Evaluation	This design pattern 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 design pattern is tied closely with model use, and is also associated with model revision and model elaboration.
Model Formation	This design pattern 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 design pattern 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 design pattern also overlaps with those for Model Elaboration and Model Revision.
Model Revision	This design pattern supports developing tasks in which students revise a model in situations where a given model does not adequately fit the

	<p>situation or is not sufficient to solve the problems at hand.</p> <p>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.</p>
Model Use	<p>This design pattern 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.</p>
Model-Based Inquiry	<p>This design pattern supports developing tasks in which students work interactively between physical realities and models, using principles, knowledge and strategies that span all aspects and variations of model-based reasoning.</p>

We have identified storyboarding and item design and development to be leverage points in the assessment development processes that we can impact with the DRK-12 design patterns (Snow, Haertel, Fullkerson, Feng, & Nichols, 2010). Specifically, we found that the newly developed design patterns support the *MCA-II* by helping improve the efficiency of storyboard and item development, assisting writing teams address hard-to-assess benchmarks, and by making explicit the validity argument.

### **Invasive Species, Model Revisions & Burmese Pythons**

Below, as an example, we describe the use of a design pattern representing the practice of model revision (see Table 3) to develop a storyboard for a complex science assessment task that requires students to consider the introduction of an invasive species into a new ecosystem (content target) and to revise a food web to accommodate for the introduction of the new species (inquiry target).

In the *MCA-II* context, storyboard writing typically starts with an assignment consisting of benchmarks from the science framework. For example, consider the following benchmarks embedded in the MN statewide science standards: 1) Explain how ecosystems can change as a result of the introduction of one or more new species (e.g., the effect of migration, localized evolution or disease organisms), and 2) Explain how scientific and technological innovations—as well as new evidence—can challenge portions of, or entire accepted theories and models. Note that, by design, the assignment contains both Biology (e.g., interdependence among living systems), and Nature of Science and Engineering (e.g., use of evidence to challenge accepted theories/models) benchmarks.

The somewhat broad nature of this assignment suggested several possible design patterns for supporting the storyboard writing process. After reviewing the Overview and Characteristic Features of design patterns in the PADI design pattern library, *model revision in model based-*

*reasoning*<sup>1</sup> (see Table 4) seemed to be most strongly related to the Nature of Science benchmarks (i.e., use of evidence to challenge accepted theories/models) assigned to this storyboard.

**Table 4. Selected Attributes of Design Pattern from Model Revision in Model-Based Reasoning**

Title	Overview
Use	Model-based reasoning concerns making inferences about real-world situations through the entities and structures of a model. When the model is not appropriate for the job at hand, either because it does not fit or it does not adequately capture the salient aspects of the situation, it is necessary to be able to revise the model.
Focal knowledge, skills and abilities	<p>Fk1. Ability, in a given situation, to modify a given model so that its features better match the features of that situation for the purpose at hand.</p> <p>Fk2. More specifically: Recognizing the need to revise a provisional model.</p> <p>Fk3. Modifying the provisional model appropriately and efficiently.</p> <p>Fk4. Justifying the revisions in terms of the inadequacies of the provisional model.</p>
Potential observations	<p>Po1. Quality and appropriateness of model revisions in order to address inadequacies of provisional model.</p> <p>Po2. Degree of and appropriateness of general and/or domain-specific heuristics students use to revise their models.</p> <p>Po3. Quality of the basis on which students decide that a revised model is adequate.</p> <p>Po4. Quality of explanation of the basis on which students decide that a revised model is adequate</p> <p>Po5. Efficiency of the process by which students evaluate existing models as adequate ,including use of optimal strategies, sequence, and when to be applied when model revision is part of a larger investigation.</p> <p>Po6. Extent to which students extract true results from a set of false models and recognize them as independent of the specific assumptions that vary across the (false) models.</p> <p>Fk2. More specifically: Recognizing the need to revise a provisional model.</p> <p>Fk3. Modifying the provisional model appropriately and efficiently.</p> <p>Fk4. Justifying the revisions in terms of the inadequacies of the provisional model.</p>
Potential work products	Pw1. Identification of reasons why the provisional model needs revision

<sup>1</sup> The full design pattern is available for open-access from [http://design-drk.padi.sri.com/padi/AddNodeAction.do?NODE\\_ID=2222&state=viewNode](http://design-drk.padi.sri.com/padi/AddNodeAction.do?NODE_ID=2222&state=viewNode)

- 
- Pw2. Select new elements (via drag-and-drop) to add to provisional model to be consistent with logic of needed model revision
  - Pw3. Choice or production of revised model
  - Pw4. Explanation of reasoning for revised model
  - Pw5. Trace of models as constructed/revised (e.g., sequence of Genetics Construction Kit (GCK) models)
  - Pw6. Recordings or transcripts of what students said as they "thought aloud" while revising model
  - Pw7. Computer-kept records of inquiry steps in which model revision steps are embedded
  - Pw8. Notes written by students during model revision
- 

With the assignment and relevant design pattern in hand we turned to brainstorming and selecting, via use of a suitability checklist (see Table 5), a broad storyboard topic: impact of invasive species on local ecology. In order to narrow our topic and locate a specific example of invasive species we scanned external resources such as web sites, text books, and instructional materials, and came across a good deal of relevant information on the introduction of Burmese Pythons into Florida Everglades National Park and their role as an invasive species in the Everglades ecosystem. We then started to identify supporting background information and relevant knowledge representations that could be used to enrich the storyboard scenes and to develop artwork for the scenes.

**Table 5. Suitability Checklist for a Specific Storyboard Topic**

---

<b>Topic: Invasion of Burmese Python in Florida Everglades</b>
<ul style="list-style-type: none"> <li>• Does the storyboard topic align with content and nature of science benchmarks?               <ul style="list-style-type: none"> <li>○ Is topic grade-level appropriate?</li> <li>○ Is the topic accessible (i.e., opportunity to learn) to the target respondents?</li> </ul> </li> <li>• Does the topic address the Design Pattern Characteristic Features and the selected FKSA's?</li> <li>• Is the topic compatible with the selected Narrative Structure?</li> <li>• Is the topic engaging to the target respondents?</li> <li>• Is the topic rich enough to support multiple embedded items?</li> </ul>

---

The next task, following selection of the design pattern and specification of storyboard topic, was to begin using the model-based revision design pattern to help create storyboard scenes. Focal knowledge, skills and other attributes (FKSA's) are among the most important attributes of design patterns to support the development of scenes, as they help make sure the storyboard is aligned with the assigned standards and benchmarks, and more importantly, hard-to-access topics that are applicable across content areas such as the Nature of Science standards

and benchmarks. Since PADI design patterns were developed with the National Science Education Standard’s (NSES) unifying themes in mind, aligning Nature of Science standards and benchmarks with Focal KSAs in the design patterns helps storyboard writers focus on these challenging, but important NSES themes. As shown in Table 6, we examined and linked the assigned benchmarks to the Focal KSAs of the model-based revision design pattern and determined FKSA #2 and #3 should be used in the storyboard.

**Table 6. Relationships between Assigned Benchmarks and Focal KSAs in the activated design pattern**

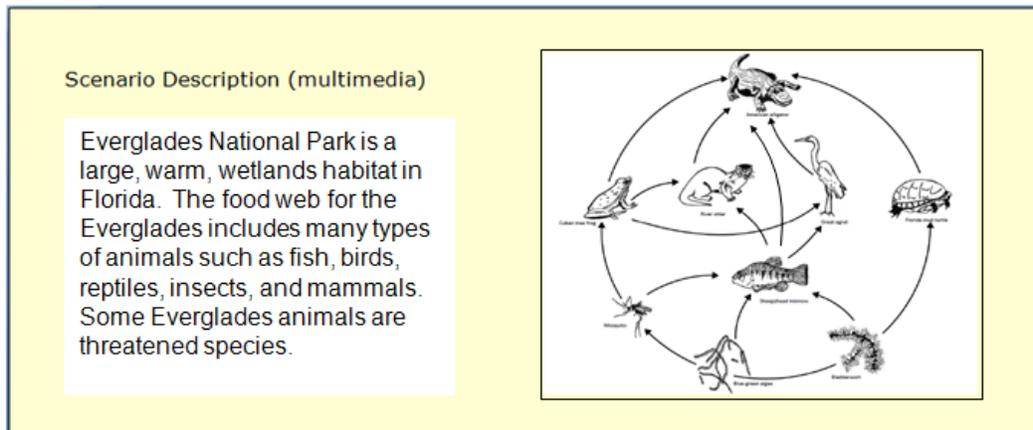
MN Nature of Science and Content Benchmarks	Model-Based Revision Focal KSAs
(Nature of Science) 9.1.1.2.2 Evaluate the explanations proposed by others by examining and comparing evidence, <b>identifying faulty reasoning</b> , pointing out statements that go beyond the scientifically acceptable evidence, and <b>suggesting alternative scientific explanations</b> .	#2. Recognizing the <b>need to revise a provisional model</b> .
(Content) 9.4.2.1.2 <b>Explain how ecosystems can change</b> as a result of the introduction of one or more new species	#3. <b>Modifying the provisional model</b> appropriately and efficiently.

The next step in using the model-based revision design pattern was to decide on the storyboard arc, or the overarching structure of the scenario needed to tell a story, or recreate an authentic context. For this purpose, we reviewed and evaluated six *narrative structures* (Fulkerson, Nichols, Haynie, & Mislevy, 2009) associated with the model revision design pattern. A storyboard using an *Investigation* narrative structure, for example, typically would describe the processes used in an experimental or observational study to examine a variable or a set of variables. It was possible to study a set of variables as a part of the ecosystem, but such a study would be very complex and maybe too difficult for students at this grade level (grade 8-9). Additionally, investigating any specific element of the ecosystem would be tangential to the Design Pattern requirement for “revising a model.” Among the plausible narrative structures, we finally selected *Change over Time*, as storyboards using this narrative structure typically would present a sequence of events that highlight sequential changes in the system. Correspondingly, in our case of invasive species, the story line will be that as the prevalence of the new invasive species increases, it results in sequential changes in the ecosystem over time.

Based on the assigned Nature of Science and content benchmarks, and the selected Focal KSAs and narrative structure, we drew implications for what will be assessed in the storyboard: a) test taker recognizes the original model is inappropriate for the current ecosystem and needs to be revised to reflect change over time; b) test taker modifies the original model to include changes to the ecosystem resulting from invasive species. Overall, the storyboard and its items

had to provide a model of the Everglades food web and provide an opportunity for students to revise the food web model to accommodate for the influence of the Burmese python.

Once we selected an appropriate narrative structure for the storyboard, we moved on to draft storyboard scenes that captured the trajectory of the narrative structure given the selected topic and Focal KSAs. We began this process by conceptualizing five story scenes to describe the invasion of Burmese Python into Florida Everglades National Park and its impact on the local ecosystem. As examples, we present three of the five scenes (#s 1, 2 and 4) below. The first scene (see Figure 1) shows an original food web of the Florida Everglades with alligators as the top-level predator, states that the species in the food web include birds, reptiles, small mammal, rodents, fish, and states some species in the food web are threatened.



**Figure 1. Scene 1 of the storyboard on invasion of Burmese Python into Florida Everglades**

The second scene (see Figure 2) describes the natural habitat of the Burmese python (Southeast Asia) and that the python was a top level predator in that ecosystem. In the 1990s Burmese pythons gained popularity as pets in Florida. Some pet owners released the pythons into the Everglades when they became too difficult to keep as pets. An increase in the python population would have a large effect on the Everglades food web because pythons eat birds, reptiles, mammals as indicated by an associated table of the stomach contents of pythons.

Scenario Description (multimedia)

In the last decade a new snake, the Burmese python, has invaded the Everglades ecosystem. Starting in the 1990s, some pet owners released pythons into the Everglades. This large snake thrives in warm and wet habitats and is the top predator in its native southeast Asia habitat. Scientists in Florida have recorded the content of pythons' stomachs to include many types of mammals and birds.

Narration Control Panel

Stomach Content of Burmese Python in Florida		
<p><b>Mammals</b></p> <ul style="list-style-type: none"> <li>• Rabbit</li> <li>• River Otter</li> <li>• Hispid Cotton Rat</li> <li>• Grey Squirrel</li> <li>• Fox Squirrel</li> <li>• Cotton Mouse</li> <li>• Old World Rats</li> </ul>	<ul style="list-style-type: none"> <li>• Virginia Opossum</li> <li>• Round-tailed Muskrat</li> <li>• Raccoon</li> <li>• Bobcat</li> <li>• Rice Rat</li> <li>• White-tailed Deer</li> </ul>	<p><b>Birds</b></p> <ul style="list-style-type: none"> <li>• Pied-build Grebe</li> <li>• Limpkin</li> <li>• White Ibis</li> <li>• American Coot</li> <li>• House Wren</li> </ul>

**Figure 2. Scene 2 of the storyboard on invasion of Burmese Python into Florida Everglades**

In the fourth scene (see Figure 3), the original model is reviewed in light of the background knowledge about the Burmese Python and the Everglades, and is revised to account for the presence of two top predators (i.e., alligator and, now, the Burmese Python) in the Florida Everglades.

Scenario Description (multimedia)

The Burmese python is now present in the Florida Everglades ecosystem. These pythons have a diet that is very similar to that of the American alligator.

**Figure 3. Scene 4 of the storyboard on invasion of Burmese Python into Florida Everglades**

The process of model revision can be hard to assess, especially when it is combined with disciplinary content in a scenario-based task. Several scenes are needed to create an authentic context for assessing the process of model revision in the context of invasive species. As a result, our storyboard creation process started with simple scenes that supported more basic embedded items and progressed to more complex scenes that supported more difficult embedded items. In our example, the first two scenes were created to support more basic embedded items. The first scene set up the context for us to ask questions about student's understanding of the original model (i.e. the food web for Florida Everglades National Park), while the second scene introduced the Burmese Python into the ecosystem. Because the Burmese Python consumed a

diet that includes many types of organisms in the ecosystem, its proliferation could potentially have a big impact on the food chain, which in turn brought up the need for revising the original food web to reflect this change. Scene four acknowledged the presence of the Burmese Python in the ecosystem and set up the stage for items that would ask students to revise the original model to incorporate the python and justify their reasoning (see Figure 4).

1. Is the Everglades food web pictured above accurate? (MC)

- No, because the food web doesn't contain any invasive species
- Yes, because the food web shows animals that all live in the Everglades
- No, because more types of tropical snakes should be added
- No, because the American alligator should be removed

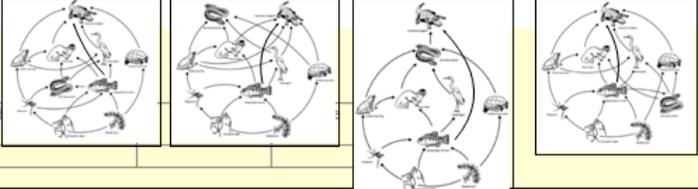
2. How should the original Everglades food web be updated? (MC)

- Updating is not necessary
- Add Burmese python as a secondary consumer
- Add Burmese python as a primary consumer
- Remove the alligator

3. Which food web best represents the current Everglades ecosystem? (MC)

- Explain your choice (CR)

Scenario ID_scene	Story



**Figure 4. Item ideas for scene 4 of the storyboard on invasion of Burmese Python into Florida Everglades**

We have described how we used a design pattern focused on model revision to brainstorm and specify a storyboard topic, and to write storyboard scenes. Another use of design patterns that we have explored in the DRK-12 project is their use in helping sketch item ideas for the storyboard scenes. The “secret weapons” from a design pattern that can inform item ideas are Potential Work Products and Potential Observations. Potential Observations represent the evidence that can be used to draw a conclusion about student understanding of the focal KSAs; Potential work products, like a written product or a spoken answer, are instantiations of evidence that students might produce that will provide evidence KSAs; while item ideas have to do with how to elicit evidence of student’s knowledge about KSAs. Therefore, we reviewed the Potential Observation and Potential Work Products of the “model revision” design pattern in order to generate item ideas for our storyboard on invasive species. For instance, one potential observation stated, “Quality of the basis on which students decide that a revised model is adequate”, for which the associated potential work product is “Explanation of reasoning for revised model”. This inspired us to sketch an item idea as “Please justify your updates to the food web.” Another potential observation, “Quality or appropriateness of model revisions in order to address inadequacies of provisional model” and the associated work product “choice or production of revised model” have also informed us of other item ideas such as asking students the question, “Is the Everglades food web pictures accurate?”. Figure 4 presents the three item

ideas that we created for this storyboard where students would first need to identify the necessity of revising the current model, then choose a correct revised model, and finally justify their selection as constructed response.

Additionally, the Additional KSAs and Variable Features of the activated design pattern can be reviewed and applied if the authors opt to further fine-tune the created scenes and item ideas. Reviewing the Additional KSAs can help determine supports that should be provided to test takers so that student performance responses to items provide evidence of the Focal KSAs, but not other supporting knowledge, presentations, or tools, etc. For example, storyboard writers could ask the following question about the constructed scenes: “Are the scenes focused too heavily on eliciting performances and behaviors consistent with the Additional KSAs (*not* the Focal KSAs)?” Variable Features can help item writers when there is a need to vary features of the scenes or item ideas that, for example, make the task more or less difficult. For instance, in our example presented above, we may choose to make the task more difficult by asking students to develop a new model during an investigation, rather than asking them to revise a given model.

## Summary

Designing tasks that measure inquiry-based constructs, such as model-based reasoning and systems thinking, for use in large-scale assessment often require task writers to integrate multiple content and inquiry targets across a series of scenes, each containing multiple embedded items, and to do so in a way that the resulting scores contain little measurement error and support valid inferences about student performance. Storyboarding is one approach item writers use to deal with this complexity. We have described, by way of an example, several areas where design patterns can help support the storyboard writing process (see Table 7).

**Table 7. How Design Patterns can Support the Storyboard Writing Process**

<b>Design Pattern Attributes</b>	<b>Helps with...</b>
Overview & Characteristic Feature(s)	Identifying key features of the storyboards, particularly the topic, that must be included so that they are aligned with standards, benchmarks and Focal KSAs
Focal KSAs	Supporting development of scenes and storyboards that are aligned with hard-to-assess topics that are applicable across content areas (e.g., observational investigation)
Narrative Structures	Considering the storyboard arc, or the overarching structure of the scenario needed to tell a story, or recreate an authentic context.
Potential Work Products & Observations	Indicating scene contexts that are appropriate for eliciting desired evidence; sketching item ideas for storyboard scenes
Additional KSAs	Ensuring that scenes and storyboard do not stress content or

---

	behaviors that are that are not relevant to target of the assessment
Variable Features	Identifying features (e.g., narratives knowledge representations) of storyboards that can be varied to adjust task difficulty

---

The design pattern *overview and characteristic features* can help writers identify key features of the storyboard, especially a topic, which must be included so that they are aligned with standards, benchmarks and FKSAs. *Focal KSAs* help writers develop storyboard scenes that are aligned with hard-to-assess topics (e.g., model-based revision) that are applicable across content areas. *Narrative structures* help writers consider the “arc” of the storyboard, or the overarching structure of the scenario needed to tell a story, or recreate an authentic context. *Potential observations and work products* help writers develop scene contexts and embedded items that elicit student performances that count as evidence toward the targeted FKSAs. *Additional KSAs* help writers ensure that storyboard scenes do not stress content or behaviors that are that are not relevant to target of the assessment (i.e., the FKSAs). Finally, *Variable Features* help writers identify scene or item features (e.g., knowledge representations used) that can be varied to adjust difficulty or emphasis.

Narrative Structures have proven particularly useful for supporting the design of complex, scenario-based tasks that contain multiple scenes. Narrative Structures were initially identified as undergirding components of storyboards during the development of design patterns for the *MCA-II*. The *MCA-II* test is scenario-based, and the development process begins with the writing of storyboards. Storyboards are precursors to scenarios and items, serving as the context to which standards-aligned items will be associated. Storyboards describe series of events or natural phenomena, thereby creating real-world contexts for assessment tasks. Prior to recognizing that Narrative Structures are inherent components of the storyboard development process, storyboard writers were asked to rely on their own methods and tools to organize thoughts, develop themes, and construct outlines. These tools may include trial-and-error, concept mapping, outlining, graphic representations, and other methods of organization. This condition often resulted in ambiguity, frustration, and inefficiency on the part of the storyboard writers. The situation faced by storyboard writers in the *MCA-II* context isn’t so different than that faced by classroom or research-focused assessment developers wanting to write complex, scenario-based tasks that integrate multiple content and inquiry targets across a series of scenes. In each case there is a need to create a coherent series of scenes, rather than a single scene or disjointed scenes, to contextualize the questions (items) asked. By recognizing, explicating, and distributing Narrative Structures to storyboard writers and other assessment developers for use as advance organizers for designing scenario-based tasks, efficiencies can be gained and frustration can be reduced.

## References

- Baxter, G., & Mislevy, R. J. (2004). *The case for an integrated design framework for assessing science inquiry* (CSE Technical Report 638). Los Angeles: The National Center for Research on Evaluation, Standards, Student Testing (CRESST), Center for Studies in Education, UCLA.
- Behrens, J. T., Mislevy, R. J., Bauer, M., Williamson, D. M., & Levy, R. (2004). Introduction to evidence centered design and lessons learned from its application in a global e-learning program. *The International Journal of Testing*, 4, 295–301.
- Duncan, R.G., Rogat, A.D., Yarden, A. (2009) A Learning Progression for Deepening Students' Understandings of Modern Genetics Across the 5th–10th Grades. *Journal of Research in Science Teaching*, 46(6), 655-674.
- 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.
- Haertel, G. D., & Mislevy, R. J. (2008a). Year 1 annual report for Application of Evidence-centered Design to a State, Large-Scale Science Assessment. (NSF DRL-0733172). Menlo Park, CA: SRI International.
- Haertel, G., & Mislevy, R. (2008b). Application of evidence-centered design in large-scale assessment (NSF DRL-0733172).
- Huff, K., & Plake, B. (2009). Evidence-centered design in practice. Coordinated session presented at the annual meeting of the National Council on Measurement in Education (NCME), San Diego, CA.
- IMS Global Learning Consortium (2000). IMS question and test interoperability specification: A review (White Paper IMSWP-1 Version A). Burlington, MA: IMS Global Learning Consortium.
- Kane, M. T. (1992). An argument-based approach to validity. *Psychological Bulletin*, 112, 527-535.
- Means, B., & Haertel, G. (2001). *Technology supports for assessing science inquiry*. Paper commissioned by the National Academies, Board on Testing and Assessment.
- Messick, S. (1994). The interplay of evidence and consequences in the validation of performance assessments. *Educational Researcher*, 23(2), 13-23.
- Mislevy R. (2007). Validity by design. *Educational Researcher*, 36(8), 463-469.
- Mislevy, R. & Haertel, G. (2006). Implications of evidence-centered design for educational testing. *Educational Measurement: Issues and Practice*, 25(4), pp. 6-20. Malden, MA: Blackwell Publishing.
- Mislevy, R., Steinberg, L., Almond, R. G., Haertel, G. D., & Penuel, R. (2003). *Leverage points*

- for improving educational assessment (PADI Technical Report 2)*. 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.
- Pearlman, M. (2001). *Performance assessments for adult education: How to design performance tasks*. Presented at the workshop “Performance Assessments for Adult Education: Exploring the Measurement Issues,” hosted by the National Research Council’s Board on Testing and Assessment, December 12–13, Washington, DC.
- Riconscente, M., Mislevy, R., Hamel, L., and PADI Research Group. (2005). *An Introduction to PADI Task Templates*, PADI Technical Report No. 3. Menlo Park, CA: SRI International.
- Snow, E., Haertel, G., Fullkerson, D., & Feng, M, Nichols, P. (2010). *Leveraging evidence-centered assessment design in large-scale and formative assessment practices*. Paper presented at the annual meeting of the American Educational Research Association, April 9-12, New Orleans, LA.
- Webb, N.L. (1999) *Alignment Between Standards and Assessment*, University of Wisconsin Center for Educational Research.