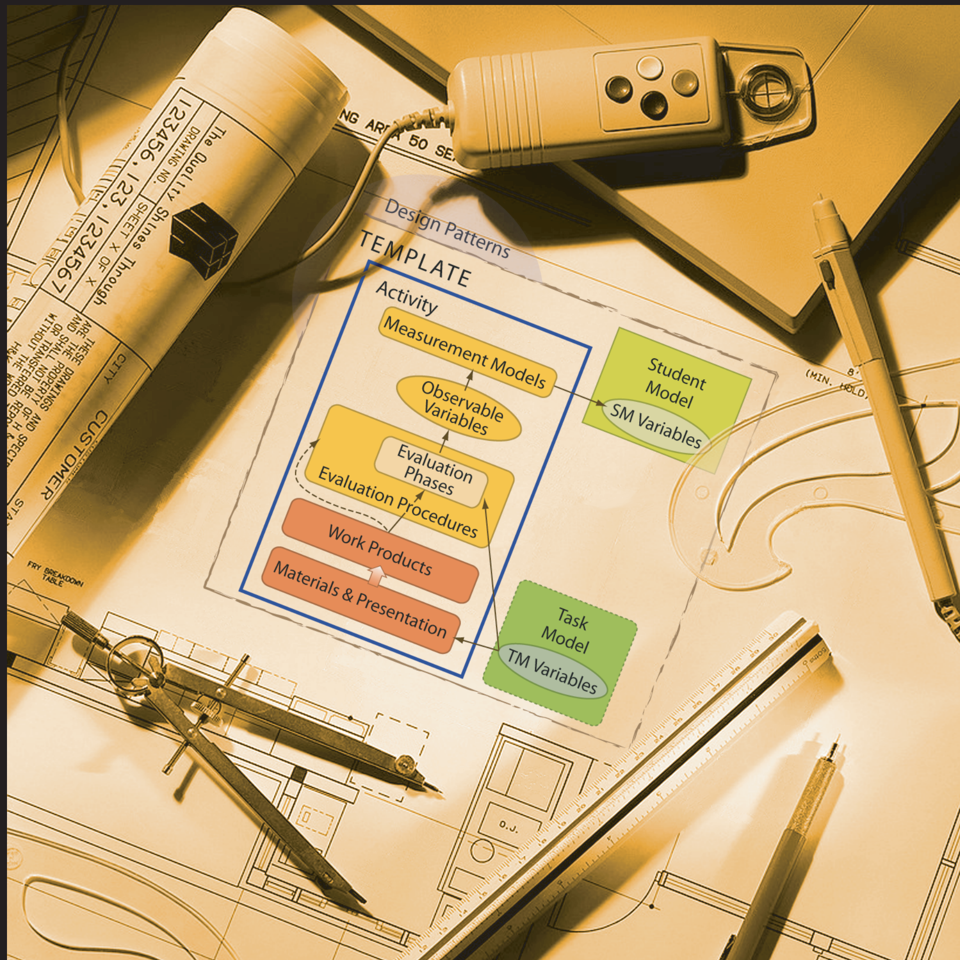


PADI Technical Report 1 | January 2003

Design Patterns for Assessing Science Inquiry

PADI | Principled Assessment Designs for Inquiry

Robert J. Mislevy, University of Maryland
Larry Hamel, CodeGuild, Inc.
Ron Fried, **Thomas Gaffney**, **Geneva Haertel**, **Amy Hafter**, **Robert Murphy**, **Edys Quellmalz**, **Anders Rosenquist**, **Patricia Schank**, SRI International
Karen Draney, **Cathleen Kennedy**, **Kathy Long**, **Mark Wilson**, University of California, Berkeley
Naomi Chudowsky, **Alissa L. Morrison**, **Patricia Pena**, University of Maryland
Nancy Butler Songer, **Amelia Wenk**, University of Michigan

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://padi.sri.com

PRINCIPLED ASSESSMENT DESIGNS FOR INQUIRY (PADI)

TECHNICAL REPORT 1

Design Patterns for Assessing Science Inquiry

Prepared by:
Robert J. Mislevy, University of Maryland
Naomi Chudowsky, University of Maryland
Karen Draney, University of California, Berkeley
Ron Fried, SRI International
Thomas Gaffney, SRI International
Geneva Haertel, SRI International
Amy Hafter, SRI International
Larry Hamel, CodeGuild, Inc.
Cathleen Kennedy, University of California, Berkeley
Kathy Long, University of California, Berkeley
Alissa L. Morrison, University of Maryland
Robert Murphy, SRI International
Patricia Pena, University of Maryland
Edys Quellmalz, SRI International
Anders Rosenquist, SRI International
Nancy Butler Songer, University of Michigan
Patricia Schank, SRI International
Amelia Wenk, University of Michigan
Mark Wilson, University of California, Berkeley

Acknowledgments
This material is based on work supported by the National Science Foundation under grant REC-0129331 (PADI Implementation Grant). We are grateful for discussions on design patterns with the PADI Advisory Committee: Gail Baxter, Audrey Champagne, John Frederiksen, Edward Haertel, James Pellegrino, and James Stewart.

CONTENTS

LIST OF FIGURES

List of Tables

ABSTRACT

Designing systems for assessing inquiry in science requires expertise across domains that rarely resides in a single individual: science content and learning, assessment design, task authoring, psychometrics, delivery technologies, and systems engineering. The goal of the Principled Assessment Designs for Inquiry (PADI) project is to provide a conceptual framework for designing inquiry tasks that coordinates such efforts and provides supporting tools to facilitate them. This paper reports progress on one facet of PADI: *design patterns* for assessing science inquiry. Design patterns bridge knowledge about aspects of science inquiry that one would want to assess and the structures of a coherent assessment argument, in a format that guides task creation and assessment implementation. The focus at the design pattern level is on the substance of the assessment argument rather than on the technical details of operational elements and delivery systems, which will be considered within the PADI system, but at a later stage of the process. We discuss the nature and role of design patterns in assessment design, suggest contents and structures for creating and working with them, and illustrate the ideas with a small start-up set of design patterns.

Introduction

Designing high-quality assessments of science inquiry, especially ones that use advanced technology, is a difficult task, largely because it requires the coordination of expertise in different domains, from science education and cognitive psychology to psychometrics and interface design. Our project, Principled Assessment Designs for Inquiry (PADI), has been supported by the Interagency Educational Research Initiative (IERI) to create a conceptual framework and supporting software to help people design inquiry assessments. This report describes structures we call *design patterns* for assessing science inquiry and specifies their roles and content with some initial examples.

The following section begins with a brief overview of PADI. We then present a rationale for design patterns as organizing schemas built on the principles of assessment design. Design patterns link science inquiry and content with the more technical specifications for an operational assessment. We then mention analogous design objects in other fields, noting parallels to the planned use of design patterns in assessment design. The content and structure of design patterns are described and illustrated with an initial set of examples and applications. We then describe the software design process, including an object model (which lays out the components of a system and how they interrelate) for design patterns within the more encompassing PADI design support system. We close by outlining next steps for the project.

Brief Overview of PADI

The goal of IERI, broadly speaking, is to promote educationally useful research that supports the learning of increasingly complex science content. A major barrier to accomplishing this goal is the scarcity of high-quality, deeply revealing measures of science understanding. Familiar standardized assessments have difficulty capturing the components of scientific inquiry called for in the national standards and in curriculum reform projects. Measures of learning embedded in technology-based learning environments for supporting scientific inquiry reflect the richness and complexity of the enterprise, but they are generally so intertwined with the learning system within which they are embedded as to be impractical for broad administration. Moreover, the production of technology-based learning assessment measures is a resource-intensive process. Research groups and educators find themselves devoting scarce resources to developing inquiry assessments in different content areas from the ground up without benefit of a guiding framework. Few of these measures offer an underlying cognitive or psychometric model that would support their use in broader research contexts or permit meaningful comparisons across contexts (Means & Haertel, 2002).

The Principled Assessment Designs for Inquiry project aims to provide a practical, theory-based approach to developing high-quality assessments of science inquiry by combining developments in cognitive psychology and research on science inquiry with advances in measurement theory and technology. The center of attention is a rigorous design framework for assessing inquiry skills in science, which are highlighted in standards but difficult to assess. The long-range goals of PADI, therefore, are as follows:

- Articulate a conceptual framework for designing, delivering, and scoring complex assessment tasks that can be used to assess inquiry skills in science.
- Provide support in the form of resources and task schemas or templates for others to develop tasks in the same conceptual framework.
- Explicate the requirements of delivery systems that would be needed to present such tasks and evaluate performances.
- Provide a digital library of working exemplars of assessment tasks and accompanying scoring systems developed within the PADI conceptual framework.

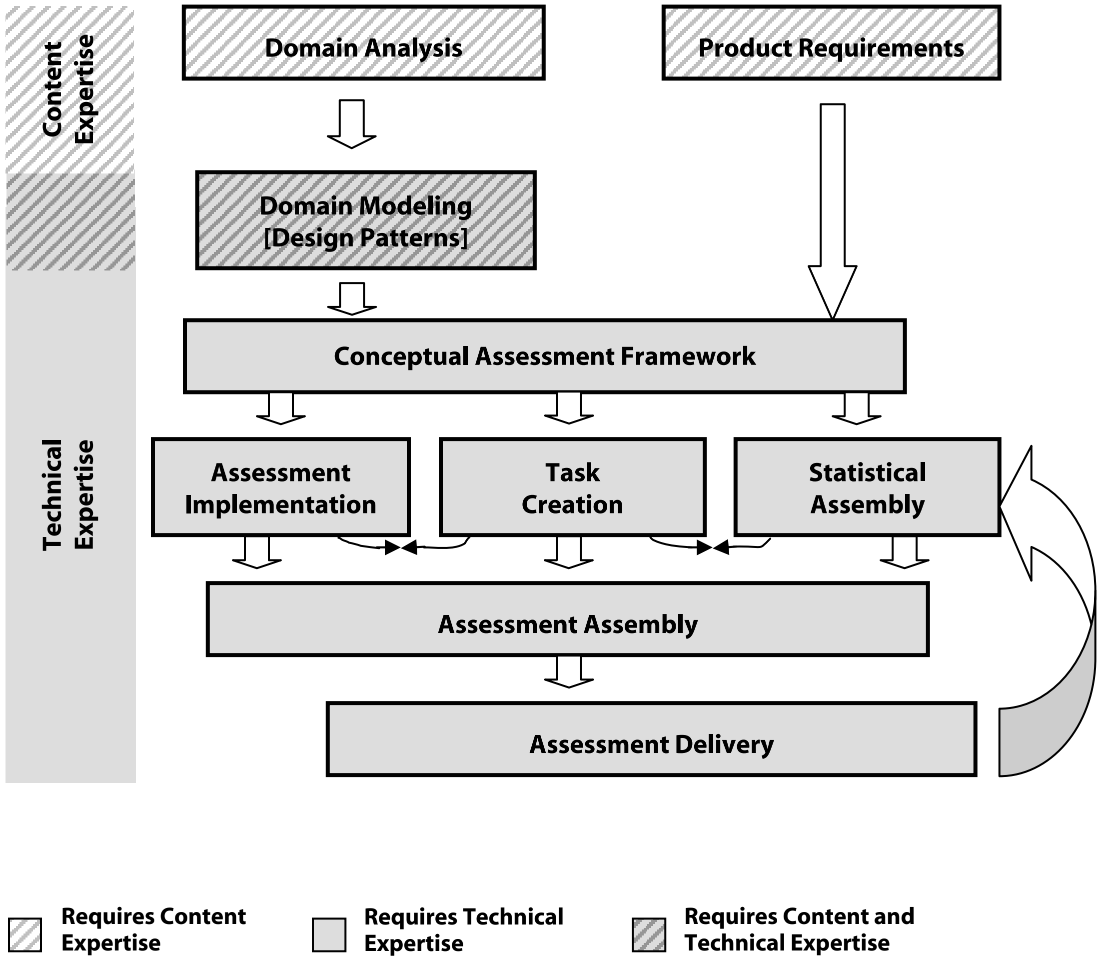
The PADI approach to standards-based assessment moves from statements of standards, through claims about the capabilities of students that the standards imply, to the kinds of evidence one would need to justify those claims. These steps require working from the perspectives of not only researchers and experts in the content area but experts in teaching and learning in that area. In this way, the central concepts in the field and how students come to know them can be taken into account. Moreover, we incorporate the insights of master teachers into the nature of the understanding they want their students to achieve, and how they know that understanding when they see it.

The goals of replicability and scalability require this effort up front, to work through the connections from claims about students' capabilities to classes of evidence in situations with certain properties. We need to go beyond thinking about individual assessment tasks, to seeing instances of prototypical ways of getting evidence about the acquisition of various aspects of knowledge. This approach increases the likelihood that we will identify aspects of knowledge that are similar across content areas or skill levels, and similarly identify reusable schemas for obtaining evidence about such knowledge.

To this end, we are developing in PADI a focused, special-case, implementation of the evidence-centered assessment design (ECD) framework developed at Educational Testing Service by Mislevy, Steinberg, and Almond (2002). The ECD framework explicates the interrelationships among substantive arguments, assessment designs, and operational processes.

Figure 1 shows the major phases in the design and delivery of an assessment system. The bar on the left side of Figure 1 and the shading denote the types of expertise needed in different parts of the assessment system. Science educators who may not be familiar with the technical aspects of creating complex assessments work at the domain analysis level. Their work focuses on specifying the knowledge about which students are assessed in a particular domain. In contrast, technical experts in the areas of psychometrics, Internet-based delivery systems, database structures, and so on, must produce the technical infrastructure to create and deliver the assessments, even though they may lack expertise in the particular science domain being assessed, or knowledge about how students learn. The work of the technical experts takes place at the level of the Conceptual Assessment Framework, and the operational processes below it.

Figure 1. Relationship between components of an assessment system and developer expertise



Design patterns lie in the layer in the assessment system called Domain Modeling. Domain analysis is the activity of identifying the knowledge and skills in a particular subject area to be assessed. Domain modeling specifies the relationships among the knowledge and skills in the area to be assessed. Design patterns are one example of a domain modeling tool. In the case of PADI, the domains of interest are a mix of science content and inquiry processes. The design pattern specifies, in non technical terms, the evidence-centered assessment argument and bridges the content expertise and measurement expertise needed to create an operational assessment.

The technical layers of the assessment system are where the details of psychometric models, scoring rubrics or algorithms, presentation of materials, interactivity requirements, and so on, are specified. This technical work can be carried out in accordance with one or more design patterns that lay out the substantive argument of the planned assessment in a way that coordinates the technical details.

Rationale for an In-Between Layer Connecting the Substance of Inquiry with Assessment Structures

The design patterns that are being developed as part of the PADI system are intended to serve as a bridge or in-between layer for translating educational goals (e.g., in the form of standards or objectives for a particular curriculum) into an operational assessment.

In many ways, design patterns serve as the cornerstone for the PADI system—the place that a PADI user would start when beginning an assessment design project. More specific than content standards but less detailed than technical specifications for particular assessment tasks, design patterns are intended to communicate with educators and assessment designers in a non technical way about meaningful aspects of inquiry around which assessment tasks can be built. In particular, each design pattern sketches what amounts to a narrative structure concerning the knowledge or skill one wants to address (in PADI, aspects of science inquiry), kinds of observations that can provide evidence about acquisition of this knowledge or skill, and features of task situations that allow the student to provide this evidence (Messick, 1994).

Design patterns take a key step from the world of science inquiry into the world of assessment design: beyond simply identifying important aspects of inquiry that should be assessed, they also make explicit the kinds of things one would want to see students doing to demonstrate their understanding and characteristics of assessment tasks that would elicit those kinds of evidence. Design patterns lie in the layer in the ECD framework called Domain Modeling, in which the structure of an assessment argument is explicated. The subsequent layer, in which the argument is incorporated in the specific and technical elements of the design for a particular assessment, will be implemented in the more specialized form of task templates. At that level, specifications for details of psychometric models, scoring rubrics or algorithms, presentation of materials, interactivity requirements, and so on, are specified. The intention, however, is that this work can be carried out in accordance with one or more design patterns that lay out the substantive argument of the planned assessment in a way that coordinates the technical details.

We should emphasize that the primary goal of PADI is to develop an assessment design framework, not to develop full sets of filled-in design patterns, task templates, tasks, or assessment systems per se. A framework cannot be developed, however, without actually putting the ideas to the test—seeing what works and what doesn't, where to extend and how to revise, in real assessment applications. Thus, involved in the PADI project are three different science inquiry curriculum projects, representing intended users of the system, that are serving to try out and refine the PADI processes.

- *Global Learning and Observations to Benefit the Environment (GLOBE)* is a worldwide, hands-on science education program that focuses on the collection, reporting, and studying of environmental data. Before the PADI project, the SRI developers had created a series of integrated investigation tasks to assess students' ability to investigate real-world problems. We will describe how, working backward from those tasks that are already being used successfully in classrooms, a set of start-up design patterns was created.

- The *BioKIDS: Kids' Inquiry of Diverse Species* project offers students in grades 5-8 opportunities to explore biodiversity both locally and worldwide. Instructional activities revolve around the collection of animal diversity data using simple, powerful technologies such as personal digital assistants (PDAs) for tracking animals in students' own schoolyards. The programs are targeted at high-poverty, urban students—groups not often fluent with inquiry science approaches or emerging technologies that support inquiry thinking. We will describe how the BioKIDS project is using design patterns to refine existing formative and summative assessment tasks and create new ones that help exemplify the PADI framework.

- The *Full Option Science System (FOSS)* is a K-8 project focusing on core science curriculum as described in the National Science Education Standards (NSES) (NRC, 1996) and the American Association for the Advancement of Science (AAAS) Benchmarks (AAAS, 1993). FOSS developers are developing a system of formative and summative assessments to aid teachers in making decisions about their instruction. FOSS currently focuses on three progress variables: science content, conducting investigations, and building explanations. The developers' work with PADI is focused on strengthening their understanding of how these progress variables can best be assessed. They are working backward from tasks they have already published to develop design patterns, as well as working forward by developing design patterns that will lead to the creation of new assessments.

Design Schemas from Other Fields

Similar tools or schemas have been generated in other disciplines that provide useful analogies for explaining the role of design patterns in assessment design. The following sections discuss in turn Levi-Strauss's analysis of the structure of myths, Georges Polti's 36 narrative themes in literature, and design patterns in architecture and computer programming.

The Structure of Myths

The French anthropologist Claude Levi-Strauss studied complex social phenomena in terms of recurring and universal patterns. He argued in *The Structure of Myths* (Levi-Strauss, 1958) that while the content, specific characters, and events of myths may differ widely, there are pervasive similarities based on recurring relationships among their elements. He established a structure for myths in terms of arrangements of elements he called "mythemes." Mythemes concern relations that can be abstracted from a particular myth, be rearranged, and reappear in other myths. A mytheme is a basic story element, such as the slaying of monsters that appears in Beowulf, the Odyssey, and repeatedly in the Oedipus myth. Such a structure allows myths to vary in composition and details while maintaining their overall importance as myths. Like assessment design patterns, myths relate the same human themes again and again with surface-level transformations of the elements that make up each particular story.

Polti's 36 Plot Themes in Literature

In 1868, Georges Polti laid out *The Thirty-Six Dramatic Situations* that he claimed all literary works are based on and can be categorized by (Ray translation, 1977). Examples of his plot themes include "Falling Prey to Cruelty or Misfortune" and "Self-Sacrifice for Kindred. The latter appears in plays and novels such as Shakespeare's *Measure for Measure*, Rostand's *Cyrano de Bergerac*, Dickens' *Great Expectations*, and Edith Wharton's *Ethan Frome*. What is common to all these works is a critical combination of elements: the hero, a kinsman, the "creditor" and the person or thing to be sacrificed. Much can be varied within this structure, such as what is sacrificed and why, and the relationships among the hero, the kinsman, and the creditor.

Polti did not intend these classifications to limit or constrain writers' creativity, but rather to provide a springboard for original plotting directions, to which authors would add their imagination, skill, and inventiveness. Polti's dramatic situations are still considered by many today a valuable resource of plots to spark the imagination and inventiveness of writers, and are used in many writing courses. Whether there are 36 or some other number of dramatic situations is immaterial; to be sure, some would categorize the universe of plot themes differently. The point is that discernible themes do recur in literature, and that such structures can serve as a useful tool for writers, either for analyzing existing literary works or for helping writers generate new ones. Likewise, we intend for assessment design patterns to be useful for analyzing the structure of already-existing assessment tasks or generating new ones.

At the same time, one does not walk away from Polti with clear direction on how to write a story, construct a plot, or even develop a meaningful dramatic situation.[1] The same can be said of assessment design patterns. Without some training and practice in assessment theory and design and without strong, explicated examples, design patterns alone cannot ensure that people will create good assessments.

Design Patterns in Architecture

Architect Christopher Alexander (1977) coined the term *design pattern* in the mid-70s when he abstracted common design patterns in architecture and formalized a way of describing the patterns in a "pattern language." A design pattern concerns a problem that occurs repeatedly in our environment, and the core of the solution to that problem—but at a level of generality that the solution can be applied many times without ever being the same in its particulars. The same perspective can be applied to the structure of a city, a building, or a single room. Patterns for communities include Health Centers, Accessible Greens, and Networks for Paths and Cars. Alexander stressed the importance of having overall designs emerge naturally from communities as they grew, with design patterns a useful aid to discussion and planning—as opposed to a top-down overall design enforced from above, the approach behind Brasilia that is now widely seen as fundamentally flawed. The lesson we take for PADI is the importance of providing an open system, not a straitjacket for assessment designers, but a resource that captures some hard-won lessons from assessment and science as a jump start for their own insights and experiences, to serve their own students and purposes.

Design Patterns in Computer Programming

Years later, computer scientists picked up on Alexander's work when they noticed patterns recurring in their designs. The seminal book is *Design Patterns* (Gamma et al., 1994). Many observers in the software industry acclaim design patterns as one of the most important software concepts of the 1990s. They provide developers a high level of reuse of both experience and software structures. There are many common software design patterns in use today, such as Model View Controller (MVC), "Proxy/Delegation," and "Object Factory." Although there are different types of design patterns in the software industry, each pattern has four essential elements:

1. *Pattern name (a word or two).* For communication and documentation.
2. *Problem/Context.* When to apply the pattern; explains the problem and context. May include list of conditions that must be met before it makes sense to apply the pattern.
3. *Solution.* Elements that make up the design, relationships, responsibilities, and collaborations. Not a concrete design or implementation, because a pattern is like a template that can be applied in many situations. An abstract description of how a general arrangement of elements solves a problem.

[1] See www.wordplayer.com/columns/wp12.Been.Done.html

4. *Consequences.* Results and tradeoffs of applying the pattern. The discussion in this section helps the programmer to evaluate alternatives and tradeoffs of alternative solutions addressed in a design pattern.

Table 1 further details the attributes of a software design pattern as they are laid out by Gamma et al. Many of both the generally stated components of design patterns listed above and the details of the particular style illustrated in the table have analogues in our assessment design patterns.

Table 1. Elements of a software design pattern (based on Gamma et al., 1994)

Attribute	Comments
Pattern name and classification	
Intent	What does it do/address?
Also Known As	Other names, if any.
Motivation	Scenario that illustrates problem and solution.
Applicability	What are the situations in which it can be applied? What are examples of poor designs the pattern can address? How can you recognize these situations?
Structure	Graphical representation to illustrate sequence and collaborations between solution components.
Participants	Components participating in the pattern.
Collaborations	How participants carry out their responsibilities.
Consequences	How does the pattern support its objectives? Tradeoffs and results—what can vary?
Implementation	Pitfalls, hints, techniques when implemented.
Sample Code	Optional.
Known Uses	Examples of pattern found in real systems, at least two from different domains.
Related Patterns	Other similar patterns and differences, or patterns it can be used in conjunction with.

Design Patterns, Assessment Design, and Science Inquiry

As part of the standards-based reform movement over the last two decades, states and national organizations have developed content standards outlining what all students should know and be able to do in core subjects, including science (e.g., NRC, 1996). These efforts are an important step toward furthering professional consensus about the kinds of knowledge and skills that are important for students to learn at various stages of their education. However, standards in their current form are not specifically geared toward guiding assessment design. A single standard for science inquiry will often encompass a broad domain of knowledge and skill, such as "develop descriptions, explanations, predictions, and models using evidence" (NRC, 1996, p. 145) or "communicate and defend a scientific argument" (p. 176). They usually stop short of laying out the interconnected elements that one must think through to develop a coherent assessment: the specific competencies that one is interested in assessing, what one would want to see students doing to provide evidence that they had attained those competencies, and the kinds of assessment situations that would elicit those kinds of evidence.

Interest in complex and innovative assessment is increasing these days for a number of reasons. For one, we have opportunities to capitalize on recent advances in the cognitive sciences about how people learn, organize knowledge, and put it to use (Greeno et al., 1997; Pellegrino, Chudowski, & Glaser, 2001). These advances broaden the range of what we want to know about students and what we might look for to give us evidence. We also have opportunities to put new technologies to use in assessment, to create new kinds of tasks and bring them to life, and interact with examinees (Bennett, 1999; Board on Testing and Assessment, 2002). In the design of complex assessments, design patterns help organize the assessment designers' thinking in ways that lead to a coherent assessment argument.

Design patterns lay out the chain of reasoning, from evidence to inference. Complex assessments must be designed from the very start with an explicit understanding of the inferences one wants to make, the observations needed to ground them, and the situations that will evoke those observations. The focus at the design pattern level is on the substance of the assessment argument rather than the technical details. The design pattern structure helps to prepare for the more technical details of operational elements and delivery systems, which will also appear in the PADI system, but at a later stage of the process.

In this paper, we will discuss the components of design patterns in detail and present several examples (see Tables 2-6 below) to illustrate these ideas. Development of an initial set of design patterns focused on the middle school level and drew on existing science content standards, while keeping the following principles in mind:

- Design patterns may, but do not have to, correspond to standards.
- Design patterns may be at coarser or finer grain size than a standard. One standard may link to numerous design patterns and vice versa.
- Multiple assessment tasks can be generated from a single design pattern. One assessment task may link several design patterns sequentially.
- Design patterns can be hierarchically organized.

Table 2. Sample design pattern "Viewing real-world situations from a scientific perspective"

Attribute	Value(s)	Comments
Title	1. Viewing real-world situations from a scientific perspective.	
Summary	In this design pattern, a student encounters a real-world situation that lends itself to being framed from a scientific perspective. Does the student act in a way consistent with having done so?	Viewing a situation from a scientific perspective can be contrasted with, for example, personal, political, social, or magical perspectives. This is a design pattern that is clearly appropriate for younger students. It is also appropriate for adults, once they are outside their areas of expertise.
Rationale	A scientific perspective says that there are principles and structures for understanding real-world phenomena, which are valid in all times and places, and through which we can understand, explain, and predict the world around us. There are systematic ways for proposing explanations, checking them, and communicating the results to others.	
Focal KSAs	Knowledge and understanding of how to view real-world phenomena from a scientific perspective.	
Additional KSAs	Ability to structure setting so that knowledge of particular scientific content or models is required or is minimized.	
Potential observations	Posing a scientifically answerable question.	Question should be relevant, realistic, and potentially addressable in light of the situation.
	Explaining how to get started investigating the situation.	
	Identifying reasonable scientific next steps.	
	Critiquing responses offered by other students, either predetermined or as they arise naturally.	
Potential work products	Verbal (oral or written) question, explanation of how to get started investigating the problem, etc.	
	Diagram of the situation.	Looking for relevant features, especially if there are particular substance or knowledge representations the student should be employing.
	Identification, from given possibilities, of those that reflect a scientific perspective.	
Potential rubrics		
Characteristic features	Motivating question or problem.	
	Background information provided so student can provide a meaningful question and answer.	

Attribute	Value(s)	Comments
Variable features	Amount of prompting/cueing.	Less cueing gives better evidence about whether student is internally inclined to see situations from a scientific perspective; more cueing gives better evidence about whether student is able to proceed knowing that it is appropriate to think from a scientific perspective.
	Degree of substantive knowledge involved.	"Content lean" vs. "content rich" in Baxter and Glaser's terms. Light content focuses evidence on inquiry perspective. Heavier content puts stress on knowledge of that content and calls for seeing situation in terms of models/principles. This confounds the inquiry and content KSAs, but makes it possible to get evidence about whether the student sees situations scientifically with respect to given content. [Note: see diSessa research below.]
	Amount of substantive knowledge provided.	When substantive knowledge, such as models, formulas, knowledge representation tools, or terminology, is required for an appropriate response, to what degree is it provided? Providing them reduces the load on the substantive KSAs. Not providing them means the response requires, conjunctively, the substantive KSA and the focal inquiry KSA.
I am a kind of	Scientific reasoning.	This design pattern is a part of a more encompassing pattern of assessing students' articulating between specific real-world situations and representations of those situations in terms of scientific concepts, models, and principles.
These are kinds of me	Planning solution strategies.	
I am a part of	Conducting investigations.	Viewing a real-world problem and situation can be a first phase of an investigation.
Educational standards		
Unifying concepts	*Evidence, models, and explanations.*	
Science as inquiry standards	*Abilities necessary to do scientific inquiry* ▪ *Identify questions that can be answered through scientific investigations.*	
Templates (task/evidence shells)	GLOBE generic template.	Posing a question, one of the kinds of observations that bear on the focal KSA, is the first step in a GLOBE investigation.
Exemplar tasks	[Various GLOBE tasks.]	
Online resources	www.globe.gov	
References	diSessa, A. (1982). Unlearning Aristotelian physics: A study of knowledge-based learning. *Cognitive Science*, *5*, 37-75.	Harvard physics students solve complicated mechanics problems in the classroom, but fall back on naïve explanations when asked what will happen next with kids on playground equipment—even though exactly the same models apply.
Miscellaneous associations		

Table 3. Sample design pattern "Re-expressing data"

Attribute	Value(s)	Comments
Title	4. Re-expressing data.	
Summary	In this design pattern, a student encounters data organized in one or more representational forms (RFs) and must re-express it in terms of a different RF. Can the student convert the data from one representational form to another?	RFs can include both general representations, such as charts, graphs, and tables, and specialized representations. An RF is a schema for organizing information; it has conventions such that spatial or relational relationships of elements in the RF correspond to relationships among entities, processes, or events. Re-expressing data involves recognizing the elements being addressed in an RF, understanding the relationships among them as expressed through that RF, then producing/identifying/ critiquing the mapping of those relationships into a different RF.
Rationale	Scientific data are measurements, observations, counts, or classifications of real-world phenomena, organized in terms of some scientific RF. They may be organized in a standard way, or in a way connected by a particular scientific understanding of the situation at hand.	
Focal KSAs	Knowledge of how to re-express data.	
	Ability to interpret data in RFs.	
Additional KSAs	Knowledge of particular RFs may be required.	
	Knowledge of appropriate RFs.	
	Content knowledge may be required.	
	Verbal abilities, if response mode is verbal.	
	Some knowledge of mathematics may be required.	
Potential observations	Identifying appropriate RFs for given data.	
	Putting data into new representation correctly.	
	Combine data from multiple RFs into a new RF.	
	Constructing new representation with appropriate layout.	Correct axis labels, units, etc.
	Identification of correct/incorrect representations from given ones.	
	Explanation of rationale for student's own re-expression.	
	Critique of or rationale for other students' re-expressions.	
Potential work products	Written explanation of appropriate/inappropriate RFs for given data.	
	Oral explanation of appropriate/inappropriate RFs.	
	Construction of new RF.	Draw, create on computer, etc.
	Identification, from given possibilities, of most appropriate knowledge representation or rationale for using particular RF.	

Attribute	Value(s)	Comments
Potential rubrics		
Characteristic features	One or more RFs are required for original presentation of data; one or more *different* RFs are involved for the re-expression.	It must be possible for salient relationships among the entities addressed in the RFs to be expressible in both stimulus and response RFs.
Variable features	Familiarity of RFs.	Are these RFs the student is known to have experience with? If so, then the stress on knowledge of the RFs is lessened.
	Number of RFs.	Combining information from multiple representations into a single new one is more difficult than straight one-to-one re-expression.
	Complexity of the RFs.	The more complicated the relationships, numbers of variables, etc., the more difficult the task will generally be.
	Directness of translation.	Re-expressions that involve computation or transform information to a different form (e.g., numerical to visual) are more difficult than ones that don't.
	Using representational forms.	
I am a kind of		
These are kinds of me	Interpreting data.	
I am part of	Analyzing data relationships	
These are parts of me		
Educational standards		
Unifying concepts	*Evidence, models, and explanations*	
Science as inquiry standards	*Abilities necessary to do scientific inquiry* ▪ *Use appropriate tools and techniques to gather, analyze, and interpret data.* ▪ *Think critically and logically to make the relationships between evidence and explanations.* ▪ *Use mathematics in all aspects of scientific inquiry.*	
Templates (task/evidence shells)	GLOBE generic template.	Re-expressing data for errors is an optional step that may be required in a GLOBE investigation
Exemplar tasks	Various GLOBE tasks.	
Online resources	www.globe.gov	
References		
Miscellaneous associations		

Table 4. Sample design pattern "Designing and conducting a scientific investigation"

Attribute	Value(s)	Comments
Title	7. Designing and conducting a scientific investigation.	
Summary	In this design pattern, students are presented with a scientific problem to solve or investigate. Do they effectively plan a solution strategy, carry out that strategy, monitor their own performance, and provide coherent explanations?	This broad design pattern spans all phases of a scientific investigation. Phases are examined more closely as their own design patterns, "parts of" this one. Anyone planning an investigation should consult both this overall design pattern and the more focused parts of it.
Rationale	Cognitive studies of expertise show that these are components of reasoning that differentiate more competent from less competent problem solvers in a domain.	
Focal KSAs	Ability to carry out scientific investigations.	This is an overarching design pattern on scientific investigations, which pertains when considering a student organizing and managing the iterative steps in an investigation. See subpatterns for further discussion of KSAs involved in various aspects of an investigation.
Additional KSAs	Metacognitive skills.	
Potential observations	Self assessment of where one is in the investigation. Self assessment of whether investigation is proceeding appropriately or needs to be refocused.	Sample rubrics: John Frederiksen's, on self-assessment ratings for use during the course of investigation.
	Quiz on process used in investigation.	
	Pose steps of scientific investigation.	
	See subpatterns for observations that can be associated with different aspects of investigation.	
Potential work products	See subpatterns.	
Potential rubrics		
Characteristic features	Motivating question or problem to be solved.	
	Open-ended; little/no cueing.	To enable students to come up with own solution strategy.
Variable features	Holistic vs. discrete task.	The task might require students to develop and carry out solutions from start to finish, or the task might address only a part (or a few parts) of the solution process (e.g., have students come up with a plan for solving problem, but not actually carry steps out).
	Complexity of inquiry activity.	There is a broad range of inquiry tasks that students might be asked to perform.
	Extent of substantive knowledge required.	Prior knowledge: tapping into what students already know. Provided information: asking students to use what you have taught them.

Attribute	Value(s)	Comments
Variable features (continued)	Focus on domain-specific vs. general knowledge.	Specific: knowledge specific to domain (e.g., conservation of energy). General: principles that cut across scientific domains (e.g., control of variables).
	Focus on process vs. content.	Process: emphasis on how students approach the problem. Content: how students bring to bear their content knowledge in coming up with a plan.
	Authenticity.	E.g., simulations vs. hands-on investigation.
	Viewing real-world situation from scientific perspective.	
I am a kind of	Model-based reasoning.	[Doesn't exist yet.]
These are kinds of me		
I am part of		
These are parts of me	Planning solution strategies.	
	Implementing solution strategies.	
	Monitoring strategies.	
	Generating explanations based on underlying principles.	
Educational standards	NSES: relates to all of the *Science as Inquiry* standards	
Unifying concepts	*Systems, order, and organization.* *Evidence, models, and explanations.* *Change, constancy, and measurement.*	
Science as inquiry standards	*Abilities necessary to do scientific inquiry* ▪ *Identify questions that can be answered through scientific investigations.* ▪ *Design and conduct a scientific investigation.* ▪ *Use appropriate tools and techniques to gather, analyze, and interpret data.* ▪ *Develop descriptions, explanations, predictions, and models for using evidence.* ▪ *Communicate scientific procedures and explanations.* ▪ *Use mathematics in all aspects of scientific inquiry.*	
Templates (task/evidence shells)		
Exemplar tasks	Mystery Powders (Baxter, Glaser & Elder, 1996).	In this performance assessment students are asked to investigate which of three white powders (salt, baking soda, and cornstarch)—individually or in combination—are contained in each of six bags.
Online resources		
References	Baxter, G. P., Elder, A. D., & Glaser, R. (1996). Knowledge-based cognition and performance assessment in the science classroom. *Educational Psychologist, 31*(2), 133-140.	
	John Frederiksen's work on self-assessment ratings for use during the course of investigation	
Miscellaneous associations		

Table 5. Sample design pattern "Participating in collaborative scientific inquiry"

Attribute	Value(s)	Comments
Title	12. Participating in collaborative scientific inquiry.	
Summary	In this design pattern, a student collaborates with one or more peers on an inquiry-based activity. For instance, a group might be presented with a situation that requires them to jointly generate a hypothesis to explain some data, plan and conduct an investigation, or create and test a model. Do students demonstrate effective collaborative skills?	This design pattern will usually be coupled with other, more substantive inquiry design patterns that tend to focus on students' working on their own. A team of students might be trying to tackle the same issues, but additional KSAs come into play when students work together.
Rationale	Some of the most important real-world science involves social activity. E.g., scientists frequently think through ideas in conversations with others, work in teams to conduct experiments, and coauthor reports of their findings and conclusions.	*Situative* learning theories emphasize that much of what we know is acquired through discourse and interaction with others.
Focal KSAs	Abilities to communicate, work cooperatively, and build on ideas of others.	Each individual needs to possess these skills to function effectively in the group.
Additional KSAs	Inquiry skills specific to the task at hand.	Required of the group as a whole (rather than each individual).
	Knowledge of particular content may be required.	Again, of the group as a whole.
Potential observations	Constructing shared understandings through discussion and clarification of ideas.	Some observations might be made at the group level.
	Developing criteria for evaluating own and peers' work.	See if values of the scientific community show up in students' criteria.
	Giving effective help.	This (and observations that follow) could be made at the individual level. Webb and colleagues (2001) describe effective help as (1) relevant to the target students' need for help, (2) timely, (3) correct, and (4) sufficiently elaborated (i.e., explanations, not just the answer).
	Receiving help.	
	Initiating topics.	
	Presenting substantive assertions, explanations, or hypotheses.	.
	Adapting communication to the needs/ abilities/ understandings of other group members.	
	Clarifying questions and ideas.	
	Creating opportunities for others to participate.	
	Recognizing and resolving contradictions between one's own and peers' perspectives.	
	Proposing resolutions to conflicts.	
	Producing coherent work product.	

Attribute	Value(s)	Comments
Potential work products	Group interactions directly observed and recorded by teacher.	
	Student-produced rubrics for self and peer evaluations.	
	Written report of solution, findings, model etc.	
	Oral presentation.	
	Something like computer-based *Knowledge Map* (see References below) to track construction of communal knowledge.	
Potential rubrics		
Characteristic features	Significant, socially shaped activity.	*Significant* implies work that is meaningful and authentic to the discipline. From a situative perspective, a multiple-choice test does not meet this criterion. Performance on traditional tests is viewed as performance on the situation that the test presents (e.g., responding to a series of questions with four options, under timed conditions, with no access to resources). According to this view, such a test can produce reliable observations, but those observations tell one about something that is relatively trivial.
	Activity structured so that several participants can/ must contribute to the group's accomplishments.	
Variable features	Structured vs. open task.	Are groups given step-by-step instructions for working through the activity, or is that left for them to figure out? Open tasks tend to require more collaboration than constrained ones.
	Assigned vs. open roles.	Are roles assigned to students or must they divide up the work themselves?
	Complexity of inquiry activity.	There are a broad range of inquiry tasks that students might collaborate on.
	Extent of substantive knowledge required.	Situations that require a lot of complex prior knowledge will place higher demands on sharing of knowledge.
	Group composition: ▪ Number of people in group ▪ Familiarity among group members ▪ Homogeneous vs. heterogeneous in ability.	Will affect the types of interactive KSAs required.
	Setting.	Students might interact in person, via e-mail, etc.
I am a kind of		
These are kinds of me		
I am part of		
These are parts of me	Using the tools of science.	May or may not be used in the design pattern.
	Using the representational forms of science.	May or may not be used in the design pattern.
	Using resources.	May or may not be used in the design pattern.

Attribute	Value(s)	Comments
Educational standards		
Unifying concepts	*Evidence, models, and explanations.*	
Science as inquiry standards	*Relates to all the NSES Science as Inquiry standards.*	
Templates (task/evidence shells)		
Exemplar tasks	Thinkertools Inquiry Project: http://thinkertools.berkeley.edu:7019/ curric/TchGd-Mod1-1Dim-1994_2.pdf	Frederiksen and White at UC Berkeley have developed instructional units with embedded assessments that require collaboration. Scoring rubrics for group projects.
	Middle School Math through Applications Project (MMAP): http://mmap.wested.org/pathways/ comp_soft/index.html#Habitech	Instruction and assessment activities from a situative perspective. E.g., for the Antarctica task, students work in groups and role play architects designing a research station.
Online resources	CRESST publications: http://cresst96.cse.ucla.edu/index.htm	Includes several reports related to assessment of student collaboration.
References	Greeno, Pearson, and Schoenfeld (1996). Implications for NAEP of research on learning and cognition. National Academy of Education.	
	Webb, Farivar and Mastergeorge (2001). Productive helping in cooperative groups. CRESST report.	
	Hewitt, Scardamalia, and Webb (2002). Situative design issues for interactive learning environments http://csile.oise.utoronto.ca/ abstracts/situ_design/	Describes use of *The Knowledge Map*, a computerized utility for recording and tracking communal (e.g., class) work on a shared problem.
Miscellaneous associations		

Table 6. Sample design pattern "Evaluating the quality of scientific data"

Attribute	Value(s)	Comments
Title	5. Evaluating the quality of scientific data.	
Summary	In this design pattern, a student encounters data that may or may not contain anomalies. Can the student recognize and/or offer potential explanations for data anomalies?	
Rationale	Scientific data are measurements, observations, counts, or classifications of real-world phenomena, organized in terms of some scientific representational form (RF). A student should realize that data cannot be taken at face value; there are one or more phases in which one cycles between what one knows already about the instruments, the procedures, and the context of data gathering, and using the data for further investigation.	
Focal KSAs	Ability to evaluate data quality.	
	Knowledge of kinds of errors that can cause anomalies in general.	
	Knowledge of particular content.	
Additional KSAs	Knowledge of measurement devices/conventions may be required for particular kinds of anomalies and their causes.	
	Knowledge of particular RFs.	
	Verbal abilities, if response mode is verbal.	
	Some knowledge of mathematics may be required.	
Potential observations	Identifying outliers.	
	Explaining error checking.	Whether or not there are errors, the student can indicate what kinds of things he/she is looking for and why.
	Proposing explanations for outliers.	
	Identifying inconsistencies across RFs.	
	Proposing explanations for inconsistencies.	
	Re-expressing data into a different RF to reveal anomalies.	
Potential work products	Written identification and/or explanation of outliers, errors, inconsistencies.	
	Oral identification and/or explanations.	
	Creation of new RF to reveal errors.	
	Selection—from given possibilities—of anomalies, inconsistencies, etc.	
Potential rubrics		
Characteristic features	Data presented to or generated by student, with or without embedded anomalies.	Data may be presented to the student, be preexisting and sought by the student, be generated by the student, or be generated by the student and peers.

Attribute	Value(s)	Comments
Variable features	Amount of data and number of RFs.	The greater the mass and heterogeneity of data, the harder it is to detect anomalies.
	Subtlety.	Stark anomalies are easier; subtle ones are harder.
	Change of representation required.	Having to re-express data to find anomalies adds difficulty; requires additional knowledge about RFs.
	Extent of substantive knowledge required.	The more identifying an anomaly depends on understanding the measurement process or the underlying phenomenon, the more the evidence depends on the KSAs involved.
	Familiarity.	Data from kinds of measurements students have had experience with will (1) tend to make the task easier and (2) make it more likely the student has the required substantive KSAs, so there is less confounding of evidence about the focal inquiry KSAs.
	Data source.	Data might be "dropped in from the sky," preexisting but sought and acquired by students in the course of an investigation, or gathered by the students themselves.
	Outlier vs. inconsistency.	An outlier is an anomaly that is identifiable in the context of its own kind—e.g., a negative number when all the data should be positive, or a value 5 standard deviations from the mean. An inconsistency is the co-occurrence of data that are not anomalies individually, but their joint appearance is. E.g., a temperature of 70-90 degrees on a given day along with 2 inches of snow is inconsistent, even though both numbers on their own are plausible.
	Interpreting data.	Error checking is a necessary part of interpreting data—one should be aware that data can contain errors and be alert to signs of anomalies.
	Re-expressing data	Can be a feature, if re-expression is involved.
I am a kind of …		
These are kinds of me		Checking data for errors is an early step in a GLOBE investigation.
I am part of…		
These are parts of me		
Educational standards		
Unifying concepts	▪ *Evidence, models, and explanation.* ▪ *Constancy, change, and measurement.*	

Attribute	Value(s)	Comments
Science as inquiry standards	*Abilities necessary to do scientific inquiry* ▪ *Use appropriate tools and techniques to gather, analyze, and interpret data.* ▪ *Develop descriptions, explanations, predictions, and models for using evidence.* ▪ *Think critically and logically to make the relationships between evidence and explanations.* ▪ *Recognize and analyze alternative explanations and predictions.* ▪ *Use mathematics in all aspects of scientific inquiry.* *Understandings about scientific inquiry* ▪ *Central role of mathematics.* ▪ *Scientific explanations.* ▪ *Role of critical evaluation.*	
Templates (task/evidence shells)	GLOBE generic template.	
Eexemplar tasks	[Various GLOBE tasks.]	
Online resources	GLOBE home page.	
References		
Miscellaneous associations		

Assessment as a Case of Reasoning from Complex Data

Educational assessment requires making sense of complex data to draw inferences or conclusions about what students know and can do. In thinking about how to make sense of complex data from assessments, we can begin by asking how people make sense of complex data more generally. How do people reason from masses of data of different kinds, fraught with dependencies and hidden redundancies, each addressing a different strand of a tangled web of interrelationships? Put simply, humans interpret complex data in terms of some underlying "story." It might be a narrative, an organizing theory, a statistical model, or some combination of these. This is how we reason in law, in medicine, in weather forecasting, in everyday life (Schum, 1994). The story addresses what we really care about, at a higher level of generality and a more basic level of concern than any of the particulars, building on what we believe to be the fundamental principles and patterns of the domain.

For instance, in law, every case is unique, but the principles of reasoning and story building are common. Legal experts use statutes, precedents, and recurring themes from the human experience as building blocks to understand each new case. Kadane and Schum (1996) present a fascinating example based on the famous Sacco and Vanzetti murder trial of the 1920s, which resulted in the execution of the two defendants, who many believe were innocent. More than 70 years after the trial, the researchers used mathematical models to establish the relevance, credibility, and probative or inferential credentials of the hundreds of pieces of evidence that were presented during and after the trial. The information that they analyzed initially existed in narrative form, often as testimonies during the trial. Kadane and Schum's sense-making process made use of two related frameworks: diagrams for the structure of arguments, and probability-based reasoning models to express directions and weights of evidence for those arguments. Alternative models based on the views of different experts allowed the authors to analyze the strength of evidence for various propositions and to compare the robustness of conclusions under different assumptions. Probability models allowed them to combine these numbers in consistent ways to provide probabilistic statements about possible endings to the stories, such as conclusions about the defendants' probable guilt or innocence, which could then be translated back into terms that make sense to people who may not be conversant with the technical methods.

In much the same way, the PADI system will develop frameworks for representing assessment arguments or chains of reasoning at two levels—one narrative (design patterns) and the other mathematical and technical (measurement models, delivery systems, human or automated scoring routines, etc.). The focus at the design pattern level will be on the substantive layer of reasoning underlying assessment tasks, to be expressed in words; at the deeper, more technical layer, the story will be constructed by using mathematical models and technological processes and data structures. But the results of these machinations will need to be translated back into words so that non technical users can understand the assessment results without having to understand the underlying technical machinery. Indeed, one of the main motivations for PADI is the need for heavy technical machinery, such as multivariate psychometric models, to meet the objectives of some assessment applications envisioned by science educators.

Design Patterns as Assessment Stories

Like Polti's narrative themes, assessment design patterns provide the story lines for assessment tasks. In the PADI system, a design pattern helps the assessment designer structure a coherent assessment story line by making explicit each of the three building blocks for an assessment argument to which we referred earlier:

1. The knowledge, skills, and abilities (which we abbreviate as KSAs for now, without making any commitment to their nature) that are related to inquiry that one wants to know about.
2. The kinds of observations that would provide evidence about those KSAs.
3. Characteristic features of tasks describing the types of situations that could help evoke that evidence.

It can be argued that all assessments are composed of these three elements, whether they are explicit or implicit in the assessment designer's mind. One purpose of the PADI system, and of design patterns in particular, is to help the designer think through these building blocks explicitly, from the very beginning, so that they guide the entire assessment design process.

KSAs (knowledge, skills, and abilities[2]) are the terms in which we want to talk about students to determine evaluations, make decisions, or plan instruction. The central set of KSAs for a design pattern can include any inquiry competencies that the assessment designer views as a meaningful unit or target for assessment, presumably because they are valued educational goals, or aspects of inquiry that research on learning suggests are important for developing scientific competence. The KSAs in the design pattern examples are expressed as general inquiry competencies that cut across science content areas (e.g., assessing the quality of scientific data, planning solution strategies, making arguments based on scientific data). These competencies may look somewhat different when instantiated in different scientific domains (e.g., biology versus chemistry). Furthermore, students who have demonstrated the competency in one domain or context will not necessarily be able to transfer the KSAs to other domains (Bransford & Schwartz, 1999). But for the purposes of laying out story lines for assessment tasks, focusing on KSAs that are important across domains of science has proven a useful starting place.

Potential observations include the variety of things that one could see students do that would give evidence that they have attained the target KSAs. Since we cannot directly see inside students' minds, we must rely on things that students say, do, or create in the task situation as evidence. Usually, there will be a variety of potential observations that would constitute evidence for a given set of KSAs. For instance, for a design pattern focused on students' abilities to evaluate the quality of scientific data, the potential observations might range from seeing students identify outliers or inconsistencies in the data, explain strategies they use for error checking, propose explanations for anomalies, or re-express data in a different representational form to reveal anomalies. And there are a variety of response modes or work products in which students could produce such evidence. They might write down an explanation in their own words, talk through their thinking with a teacher or peer, draw a new representation of the data that reveals the errors, circle anomalies, or select the error from a given set of possibilities.

Characteristic features of tasks describe the kinds of situations that can be set up to evoke the types of evidence one is looking for. Features of tasks might include characteristics of stimulus materials, instructions, tools, help, and so on. One might create a variety of types of situations to assess any given set of KSAs, but the proposal is that at some level they have something in common that provides an opportunity to get evidence about the targeted KSAs. Continuing with the example about students' abilities to evaluate the quality of scientific data, it seems that a necessary feature of the tasks would be to present students with—or have them generate their own—data with or without embedded anomalies. There are also features in the situation that can be varied to shift its difficulty or focus. For example, one could control the amount and complexity of the data that students are presented, the subtlety of the errors, and the degree of prior knowledge required about the particular measurement method used to collect the data. Clearly, from a single design pattern, a broad range of assessment tasks can be created. In fact, one purpose of design patterns is to suggest a variety of possible ways to assess the same KSAs, rather than dictating a single approach.

[2] Industrial psychologists use the phrase "knowledge, skills, or abilities", or KSAs, to refer to the targets of the inferences they draw. We borrow the term and apply it more broadly with the understanding that for assessments cast from different psychological perspectives and serving varied purposes, the nature of the targets of inference and the kinds of information that will inform them may vary widely in their particulars.

In addition to laying out these three essential elements of an assessment argument, design patterns include other information intended to be helpful to the assessment designer, including links to content standards, exemplar tasks, scoring rubrics, and other design patterns.

An assessment task could correspond to a single design pattern or a sequence or assemblage of more than one design pattern. For instance, the design pattern about evaluating the quality of scientific data could be linked with ones that require students to design their own investigation and collect their own data. Assessing the quality of the data collected could be a later stage of the task.

What Is in Design Patterns

Persistent elements and relationships. All coherent assessment arguments will include elements and relationships (i.e., the three essential elements described above), across assessments of different kinds meant for different purposes. However, the structure of design patterns is neutral with respect to the particular content, purposes, and psychological perspective that goes into them. Design patterns can be used, for example, to generate diagnostic or large-scale assessment tasks. With PADI, the focus is on science inquiry, but design patterns could as easily be created for assessing literacy or history. Within the domain of inquiry, the same structure can be used to create design patterns focused on individual cognition or social aspects of learning, factual recall, or more complex abilities. Assessment designers can be coming from a behaviorist, cognitive, or situative perspective, and still use the same design pattern *structure* for laying out their assessment arguments.

On the other hand, science educators' needs are not neutral. We will be looking to the National Science Education Standards, cognitive research on how students learn, and existing examples of good inquiry curricula and assessments as inspiration for design patterns.

For instance, the Standards describe abilities necessary to do scientific inquiry and important understandings about scientific inquiry. Those standards cut across specific content areas (physical, life, and earth science) and focus on things like identifying questions that can be answered through scientific investigations, designing and conducting investigations, and using appropriate tools and techniques to gather, analyze, interpret data, etc., that are clearly relevant to the development of inquiry design patterns.

Also particularly informative for the development of design patterns is the first category of Standards called "unifying concepts and processes." This category includes five areas that transcend grade and disciplinary boundaries:

- Systems, order, and organization
- Evidence, models, and explanation
- Change, constancy, and measurement
- Evolution and equilibrium
- Form and function

One of these areas—evidence, models, and explanation—seems to us to be at the heart of inquiry. The other four concern key relationships and structures in science, but the area of evidence, models, and explanation concerns the act of reasoning through all of these, as well as more content specific structures. What we want students to be doing, fundamentally, is carrying out the interactive process of building scientific model-based understanding of situations in the real world. This process encompasses many of the activities that people (from students to scientists) carry out in inquiry: recognizing possible models that might apply in a situation, matching up elements and processes from models with aspects of the situation, proposing model-based explanations, checking for the fit of the model, determining what else one needs to know to reason through the model, reasoning through the model to make predictions or fill in gaps, recognizing anomalies, revising a provisional model in light of new information, etc. The models will differ in their nature and complexity, and the kinds of things people must do to carry out these activities will vary in their specifics—maybe by branches of science or discipline.

We have developed some initial design patterns that, like the "unifying concepts and processes" category, cut across domains of science, but it may also be productive to have specialized design patterns that are more powerful, but limited to certain content areas. It is important to emphasize that by having design patterns that are applicable across different content areas, we are not implying that inquiry should be considered a set of generalized skills that can be assessed in the absence of science content. Instead, the goal is to create design patterns that can be instantiated in a wide variety of science disciplines. To be sure, an assessment of how well students can analyze data relationships in biology will look different from one in physics, but the same general design pattern should be useful for thinking through the basic assessment argument in both contexts.

Design patterns also emerge from analyzing exemplary inquiry curricula and assessments. This is an important component of the PADI project, which includes curriculum developers from GLOBE, FOSS, and BioKIDS. Starting with GLOBE, we developed an initial set of design patterns by working backward from a set of GLOBE assessment tasks that are already being used successfully in classrooms. We have broken the tasks down into the essential building blocks to reveal the underlying structure so that it can be made explicit (through design patterns and task templates) and eventually reused to develop more good tasks. Work is also in progress with FOSS and BioKIDS curriculum developers to develop design patterns (and then more detailed task templates) that are tied to their inquiry goals and that will be useful for developing assessment applications in those contexts.

Finally, we are looking to research on how students learn science inquiry. There is a rich body of empirical research that points to important aspects of developing competence in science, including the quality of students' explanations and problem representations, students' abilities to monitor their own problem solving, and students' ability to function in varying social and situative contexts of learning science (e.g., Baxter, Elder, & Glaser, 1996; Chi, Feltovich, & Glaser, 1981; Greeno et al., 1997; White & Frederiksen, 2000). These findings suggest targets for assessment that are largely untapped by current measures. By providing starting points for guiding the assessment of these kinds of skills, PADI design patterns broaden educators' conceptions of the competencies in inquiry that one might want to assess.

The goal of the PADI project is not to create "right" or "complete" sets of inquiry design patterns. Although we aim to start with some that are both useful and defensible to science educators, we are building an open system that users can pick and choose from, as well as add to. The system can handle design patterns aligned with very different beliefs about science inquiry, representing different theories of learning and instruction, and very different purposes for assessment. Design patterns from these different perspectives can coexist in PADI without a problem.

What Isn't in Design Patterns

Design patterns concern aspects of science inquiry that are meant to apply across levels of study (maybe different ranges for different design patterns) and across content domains (maybe more broadly for some design patterns, more narrowly for others). They do say something about how one might learn about students' inquiry capabilities in some area, but they specifically do not lay out what that area might be—that is, domain or domains, principles or themes, which models or techniques are involved. In this paper, we will show, primarily with the subsequent GLOBE examples, how the articulation from design patterns to tasks or task templates can be negotiated. More generally, however, the NSES guidelines on science standards offer good advice on writing inquiry tasks, which we may now view as instantiating science inquiry design patterns.

We have just stated that design patterns don't include particular content to create tasks and families of tasks as assessment designers necessarily must incorporate content, going beyond the design patterns proper. The NSES Standards do provide guidance to the assessment designer. In particular, Standards talk about inquiry as the skills and abilities to carry out investigations, and give lots of examples. The examples are contextualized in particular content domains, and on page 109 of the Standards we see a list of features of what makes for "fundamental content." These are requirements for every substantive standard in the book which is a broad and generative base for thinking about the content of any particular science task. However, we would want to be able to think about design patterns so that they relate to more than one of the fundamental content bullets in all of the sections in Chapter 6 that deal with physical science, life science, and earth and space science. The first four bullets, in particular, are features that an assessment designer can use to think about substantive bases of inquiry assessment when building tasks according to a design pattern, especially in relation to the unifying concepts of evidence, models, and explanations:

- *Represents a central event or phenomenon in the natural world.* This gets at the possibility of an important model or set of relationships that is relevant to many kinds of real-world situations—ones that presumably have some characteristic features at some appropriate level of generality. Task Model features (see Appendix A) would be where we lay out critical features for activating such a model, and describe variations such models can take which will help us in assessing different aspects of a student's inquiry KSAs in the context of this model.
- *Represents a central scientific idea and organizing principle.* This is the underlying model or script, which is presumably the basis of the reasoning back and forth

between the scientific substance and the real-world situation that underlie inquiry activities.

- *Has rich explanatory power*. This means that there are many issues concerning evidence and explanation that can be explored in different ways (e.g., what kind of further evidence would a student you need to various what kinds of predictions?). This further suggests there can be explanations and predictions in situations that are more familiar or less familiar to the student, distances, with more or fewer links of reasoning between them, or more or fewer steps in investigation required.

- *Guides fruitful investigations*. This point connects with the preceding one, but goes farther by saying that there are nontrivial things students can actually do, live or simulated, in instruction and assessment.

The Structure of Design Patterns, Part 1: Attributes

Design patterns are created in matrix form; the cells are filled with text and links to other objects in the PADI system, including other design patterns, as well as Web pages or resources outside the system. Table 7 provides a definition of each attribute of the design pattern structure. To illustrate each attribute, we use the design pattern introduced earlier in Table 6, for assessing students evaluating the quality of scientific data.

Table 7. Attributes of a PADI assessment design pattern

Attribute	Definition
Title	A short name for referring to the design pattern.
Summary	Overview of the kinds of assessment situations students encounter in this design pattern and what one wants to know if they can do about their knowledge, skills, and abilities.
Rationale	Why the topic of the design pattern is an important aspect of scientific inquiry
Focal KSAs	Primary knowledge/skills/abilities of students that one wants to know about.
Additional KSAs	Other knowledge/skills/abilities that may be required.
Potential observations	Some possible things one could see students doing that would give evidence about the KSAs.
Potential work products	Different modes or formats in which students might produce the evidence.
Potential rubrics	Links to scoring rubrics that might be useful.
Characteristic features	Kinds of situations that are likely to evoke the desired evidence.
Variable features	Kinds of features that can be varied in order to shift the difficulty or focus of tasks.
I am a kind of	Links to other design patterns that this one is a special case of
These are kinds of me	Links to other design patterns that are special cases of this one
I am part of	Links to other design patterns that this one is a component or step of.
These are parts of me	Links to other design patterns that are components or steps of this one.
Educational standards	Links to the most closely related NSES *Science as Inquiry* standards.
Templates (task/evidence shells)	Links to templates, at the more technical level of the PADI system, that use this design pattern.
Exemplar tasks	Links to sample assessment tasks that are instances of this design pattern.
Online resources	Links to online materials that illustrate or give backing for this design pattern.
References	Pointers to research and other literature that illustrate or give backing for this design pattern.
Miscellaneous associations	Other relevant information.

The *Title* is a short name for referring to the design pattern. The title of the example is "Evaluating the quality of scientific data."

The *Summary* is an overview of the kinds of assessment situations students would encounter in tasks that are instantiations of this design pattern and what one wants to know about students' knowledge, skills, and abilities. In the example, a student encounters data that may or may not contain anomalies. Can the student recognize and/or offer potential explanations for data anomalies?

The *Rationale* is a brief discussion of why the topic of design pattern is an important aspect of scientific inquiry. In the example, we see this topic is relevant to inquiry because a student should realize that data cannot be taken at face value; there are more phases in which one cycles between what one knows already about the instruments, the procedures, and the context of data gathering and what one knows about using the data for further investigation.

Focal KSAs are the primary knowledge/skills/abilities of students that are addressed. Here we are concerned first with capabilities in evaluating data quality and understanding the kinds of errors that can cause anomalies. This need not be construed as a global ability, we should point out; a student's propensity to check for data quality and competence for doing so might vary considerably from one domain to another and one situation to another.

Additional KSAs are other knowledge/skills/abilities that may be required in a task developed from the design pattern. Here we must be aware of the role of knowledge about the specific content, measurement devices and conventions; or particular types of representational forms. An assessment designer must consider, for example, whether a task's content should be familiar to students, so it is not a nuisance factor in assessing data-quality procedures, or whether familiarity is unnecessary. If the latter, it may be that a multivariate model will be appropriate, for use with tasks that address different content areas and different aspects of inquiry.

Potential observations are what students say, do, or make that give evidence about the focal KSAs. Here, for example, we can consider identifying anomalies like outliers or inconsistencies in the data, proposing explanations for them, reexpressing data into a different representational form to reveal anomalies, and explaining error-checking procedures. Related to both potential observations and potential work products, discussed next, are links to potential rubrics for evaluating what one observes.

Potential work products are various modes or formats in which students might produce the evidence relevant to the focal KSAs. Here we consider written identification and/or explanation of anomalies, oral identification or explanation, creation of a new representational form to reveal errors; and selection of anomalies from given possibilities.

Potential rubrics are links to scoring rubrics, perhaps in a PADI library or elsewhere on the Internet, which might be useful in evaluating student work in situations that correspond to this design pattern. There is a generic rubric from the GLOBE project that is helpful for our example.

As discussed earlier, *Characteristic features* play a central role in design patterns. They concern features that must be present in a situation in order to evoke the desired evidence. In this example, obviously there must be some data, either presented to or generated by students. It may or may not have anomalies.

Variable features of tasks can be varied to shift the difficulty or focus of tasks. In this example, one can vary the amount and complexity of data; whether an anomaly is outlier vs. inconsistency; subtlety of anomalies; and the familiarity of students with types of measurements presented. All the while, the argument structure for assessing students with respect to evaluating data quality remains the same.

The *I am a kind of* attribute is a list of links to other design patterns that this one is a special case of. "Evaluating the quality of scientific data" could be a special case of a more genera design pattern called "Analyzing data."

The *These are kinds of me* attribute is a list of links to other design patterns that are special cases of this one. "Evaluating the quality of scientific data" could encompass special cases such as "Evaluating the quality of data collected by self," in contrast to data simply presented to students.

The *I am part of* attribute is a list of links to other design patterns of which this design pattern is a component or step. "Evaluating the quality of scientific data" can be viewed as a distinguishable aspect of "Interpreting data", which may require quality checking in addition to model fitting or transformation.

The *These are parts of me* attribute is a list of links to other design patterns that are components or steps of this one. "Reexpressing data" can be a part of "Evaluating the quality of scientific data," so it will appear on the list, but it may or may not play a role in particular tasks generated from this design pattern, at the discretion of the assessment designer.

The *Educational standards* attribute is a list of links to the most closely related National Science Education Standards. For example, the design pattern "Evaluating the quality of scientific data" has links to the following national standards for science inquiry:

Abilities necessary to do scientific inquiry

- Use appropriate tools and techniques to gather, analyze, and interpret data
- Develop descriptions, explanations, predictions, and models using evidence
- Think critically and logically to make the relationships between evidence and explanations
- Recognize and analyze alternative explanations and predictions
- Use mathematics in all aspects of scientific inquiry

Understandings about scientific inquiry

- Central role of mathematics
- Scientific explanations
- Role of critical evaluation

The *Templates* (or task/evidence shells) attribute is a list of links to templates, at the more technical level of the PADI system, that use this design pattern. Our example design pattern is used in the GLOBE task template.

The *Exemplar tasks* attribute is a list of links to sample assessment tasks that are instances of this design pattern. For example, we can find example links to various GLOBE tasks that ask students to evaluate the quality of scientific data.

The *Online resources* attribute is a list of links to online materials that illustrate or provide background or support for the design pattern. The GLOBE Web site appears in our example.

References are pointers to research and other literature that illustrate or provide background for this design pattern. Studies in the cognitive literature could appear here.

Miscellaneous associations offers free-text fields for any other relevant information that the creator or enabled users of the design pattern may wish to add, including links to Web sites.

Content and Structure of Design Patterns: Several Examples

The following section further illustrates the structure, content, and features of design patterns.

Example: A Design Pattern Concerning Investigations

This example was inspired by research on cognition and performance assessment in the science classroom by Baxter, Elder, and Glaser (1996). They discuss the difficulty of assessing complex scientific reasoning through performance assessment. In an attempt to make explicit the mental processes of students' scientific reasoning, a set of tasks tapping both declarative and procedural knowledge was devised to explore the relationship between content domain knowledge and science inquiry. Their Mystery Box tasks required students to infer what configuration of wires, bulbs, and/or batteries was inside a closed box, by attaching wires, bulbs, and/or batteries to two clips leading into the box, then reasoning from the results.

Students who demonstrated proficiency in these problem-solving tasks demonstrated a clear and concise plan, a strategy for implementing that plan, and the ability to explain their plan and strategy. Further, they were able to monitor the implementation of their plan and strategy and use feedback appropriately. We worked backwards from the Mystery Box tasks to abstract the KSAs that Baxter et al. wanted to learn about, the features they built into the task situations, and what they looked for in students' solutions. Using this information, we created a design pattern, "Designing and conducting a scientific investigation" shown as Table 4. In a task that instantiates this design pattern, one would look for students to generate a plan for solution that is guided by an adequate representation of the problem situation and possible procedures and outcomes; to implement solution strategies that reflect goals and subgoals; to monitor their actions and flexibly adjust their approach on the basis of performance feedback; and to provide coherent explanations based on underlying principles rather than descriptions of superficial features or single statements of fact.

Within "Designing and conducting an investigation," there are distinguishable aspects of activity that can be described in terms of subpatterns in their own right, whether or not these activities were demarcated in the flow of students' activities. They could be used to address tasks dealing with only one particular aspect of an investigation. The overarching design pattern "Designing and conducting an investigation" is made up of four smaller design patterns, each reflecting the most crucial aspects of the overarching ability of interest: "Planning solution strategies," "Implementing solution strategies," "Monitoring strategies," and "Generating explanations based on underlying principles." For example, in a task generated from the subpattern "Planning solution strategies," students would be presented with an open-ended problem to investigate and must generate a plan for solving the problem. Such subpatterns are compatible with scaffolded instruction and assessment. In unscaffolded investigations, students have the responsibility to devise their own ways to carry out an investigation. These same design patterns would help the assessment designer think through task requirements, and develop scoring rubrics that ensure evidence would be obtained about the different aspects of designing and conducting an investigation.

"Designing and conducting an investigation" addresses planning, strategy, explanation, and monitoring. These would be central to any task that instantiates the design pattern. However, it allows for flexibility across content and process. While the research that sparked our creation of the design pattern involved electrical circuits and middle school students, the same design pattern could be used by, say, a college professor to create a template for a final project in a class on experimental psychology. The extent of substantive knowledge required is another variable feature of this design pattern. This framework allows for flexibility in terms of how much the knowledge required to perform the task is known to be familiar to students, is provided with the task, or is itself a target of assessment.

Example: Design Patterns for Situative Aspects of Inquiry

Some education practitioners and researchers have focused on the situative (also sometimes referred to as the sociocultural) aspects of learning (e.g., Greeno, Pearson, & Schoenfeld, 1996). This perspective emphasizes the importance of practical activity, context, and social interactions in the learning process. From the situative standpoint, assessment means observing and analyzing how students use their knowledge and skills to participate in an authentic community of practice (e.g., biologists, engineers). Most typical testing practices are not a good match with the situative perspective. It has been argued that traditional testing presents students with abstract situations, removed from the actual context in which people typically use the knowledge being tested.

As discussed earlier, the *structure* of design patterns can accommodate any perspective of learning and any approach to assessment, although, of course, the targeted KSAs, observations, work products, and task features within those structures could turn out to be very different. Some design patterns that we have created with an eye toward the situative perspective on learning include students' abilities and tendencies to use the tools, representational forms, and resources of scientific communities of practice. Here we describe an example called "Participating in collaborative scientific inquiry" (Table 5). This design pattern encompasses situations in which a student collaborates with one or more peers on a scientific inquiry activity, such as designing and conducting an investigation, interpreting data, or creating and testing a model. The design pattern focuses on the assessment of students' collaborative KSAs—are they able to communicate, work cooperatively, and build on the ideas of others? This design pattern will typically be coupled with other, more substantive inquiry design patterns that focus on students working individually, such as "Designing and conducting an investigation." A team of students might be trying to tackle the same issues, but additional KSAs come into play when students work together to construct shared understandings. The rationale, or reason that this design pattern is important, is that some of the most important real-world science involves social interaction. Scientists frequently think through ideas in conversations with others, work in teams to conduct experiments, and coauthor reports of their findings and conclusions.

Whereas the design pattern in Table 4 lists many potential types of observations and work products that could provide evidence of students' abilities to participate in collaborative inquiry, there are a few essential features of tasks that belong to this design pattern. First, the situation presented to students should be a significant, socially shaped activity. That is, in keeping with a situative approach, the activity should be meaningful and authentic to the discipline. A second essential feature is that tasks be structured so that several participants can (or must) contribute to completion of the work. The focus or difficulty of tasks within this design pattern can be varied by controlling features like the nature of the inquiry skills or specific content knowledge required. Tasks that require complex inquiry skills and/or sophisticated content knowledge will place higher demands on sharing of knowledge. Other variable features include group composition and whether roles are assigned or students are left to themselves to determine who does what work.

Examples from the GLOBE Assessment

Now we turn to our work applying the ideas about assessment design patterns to a real project—in this case, reverse-engineering a successful framework for authoring inquiry tasks. We began the project by thinking of the GLOBE assessment framework as reflecting a combination of inquiry design patterns, creating such design patterns, and further identifying other tasks that are very different as to surface features, yet can still be seen as other instantiations of the same design patterns.

Researchers in the Center for Technology in Learning at SRI International have explored the use of assessment templates in designing classroom assessment tools for the Global Learning and Observations to Benefit the Environment (GLOBE) Program. GLOBE is a worldwide, hands-on primary- and secondary-school-based science and education program that focuses on the collection, reporting, and studying of environmental data.[3] An explicit goal of GLOBE is to improve students' understanding of science by involving them in performing real science—taking measurements, analyzing data, and participating in research in collaboration with scientists. SRI researchers have developed a series of integrated investigation problems to assess GLOBE students' ability to investigate real-world problems by analyzing and interpreting GLOBE data sets, communicating their findings and conclusions, and designing related investigations. These problems are designed according to templates for related, modular sets of tasks that address the GLOBE assessment framework. The templates specify specific investigation phases: planning, conducting, analyzing, comparing, interpreting, and communicating.

As an example of a task created from a GLOBE template, students are presented with a set of climate-related criteria for choosing a site for a Winter Olympics. Given multiple types of climate data for a set of feasible candidate cities, students analyze the data in terms of the climate criteria, decide which city best meets the criteria overall, and prepare a persuasive presentation for the Olympic Committee contrasting the city they have chosen with the default candidate, Salt Lake City. From the performance of this complex task, SRI researchers extract evidence about both specific skills, such as the ability to comprehend quantitative information presented in graphic form, and broader aspects of scientific inquiry, such as the ability to communicate and defend a scientific argument (Coleman & Penuel, 2000).

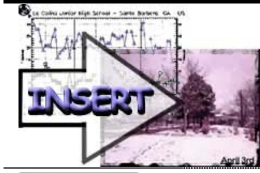
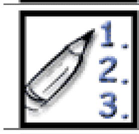
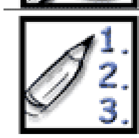
The assessment designers at SRI attempted to link inquiry learning goals with key scientific concepts found in GLOBE. In this way, the designers are able to think about what kind of knowledge and skills they want students to develop and then create tasks that will allow

[3] See www.globe.gov for more information.

observation of that knowledge and skill. One of the tools the assessment designers conceived was integrative investigation problems. The challenge was to develop very general tools that could be used in highly specific situations. That is, they wanted to develop performance assessments for unspecified learning activities, all of which had the same structure. Such assessments could be used “as is” or be modified by teachers to determine how well students interpret and analyze particular GLOBE data.

The resulting GLOBE templates connect a general form of science inquiry developed by researchers, learning specialists, and individual teachers. The templates follow a general framework across all six GLOBE content areas: hydrology, atmosphere, earth systems, soils, land cover, and visualizations. The general form of a template is shown as Figure 2. These templates include 10 specific frames or investigation phases, within which teachers may enter variable features of the task. The ten frames include presenting problems requiring the use of GLOBE data archives, asking students to pose relevant questions, asking questions about data in the table, and so on. Depending on the particular data representation, driving question, etc., most, if not all, of the frames are included in any given assessment task. Templates allow teachers to modify tasks for their particular student and classroom needs. Changing the information within the template changes the focus of the assessment but not the underlying principles of science inquiry. For example, within the same general task, the difficulty of the data presentation can be low, moderate, or high, affecting the overall difficulty of the task without changing the basic intent. Therefore, despite having a fixed template, there is a great deal of flexibility within each task.

Figure 2. Generic GLOBE task template

List school information here	
	[[Insert GLOBE data or graphs for the schools listed above here]]
	**Present problem requiring use of GLOBE data archives.** Present problem situation/driving question with background and role of the student in the investigation.
	**Planning Investigations: Ask students to pose relevant questions.** Ask the student to look at the GLOBE data/graphs provided above and come up with possible questions that she/he might ask regarding the data. Provide a sample question to help guide the student.
	**Analyzing and Interpreting GLOBE Data: Ask questions about data in the table.** Ask the student to find observable trends in the data.
	**Conducting Investigations: Assuring data quality.** Ask the student to look through the data for possible measurement or data entry errors, and suggest ways to avoid these types of errors in the future.
	**Analyzing and Interpreting Data: Ask questions requiring interpretation of data.** Ask the student what the relationship is between the two variables given. Provide the student with an example of a trend.
	**Analyzing and Interpreting Data: Ask to represent data in a graph or table.** Ask the student to use the data provided to generate new data representations to analyze trends.
	**Analyzing and Interpreting Data: Ask for interpretation of data on the graph.** Ask the student about specific features of graphs and what indications there are for various maximums, minimums, etc. Ask the student to explain her/his answer.
	**Planning Investigations: Ask to set up another problem.** Ask the students to choose another school from the GLOBE database that has some related feature as the schools they just analyzed. Have them copy a relevant data set for this new school and to perform a similar analysis on this new data set. Ask them what other variables they would be interested in looking at and why.
	**Summarizing Data: Ask to summarize and report findings.** Ask the student to summarize their analysis of the original schools and to write a short report or to prepare a short presentation of their findings and recommendations, supporting their conclusions with the analysis they have done, and to suggest other data that might be helpful for further study of the situation.

Because many GLOBE assessments have been successfully used by teachers and students, mapping features of the GLOBE assessment framework into design patterns proved to be a valuable, informative process—a source of feedback on the connection between the PADI framework and a set of existing, inquiry-based assessments. Nine design patterns[4] are needed to account for the kinds of knowledge and skills that the various frames, or investigation phases, that a GLOBE integrative investigative problem tap into. Table 8 shows the relationship between these design patterns and the frames of the GLOBE template. Note that in several cases, more than one design pattern is instantiated in a given frame.

Table 8. Design patterns corresponding to phases of a GLOBE investigation

Design Pattern(s)	Investigation Phase
	Problem Presentation (Description of interaction & flow)
▪ Viewing real-world situations from a scientific perspective ▪ Interpreting data or observations ▪ Scientific reasoning (Planning solution strategies)	**Investigation Planning** Activity 1 (Flow & directives)
▪ Interpreting data or observations ▪ Assessing the quality of scientific data ▪ Scientific reasoning (Planning solution strategies)	**Data Quality Assessment** Activity 2 (Flow & directives) Activity 3 (Flow & directives)
▪ Interpreting data or observations ▪ Analyzing data relationships ▪ Re-expressing data ▪ Using the representational forms of science	**Data Analysis** Explain relationship (Flow & directives) Re-expressing data (Flow & directives)
▪ Viewing real-world situations from a scientific perspective ▪ Using resources ▪ Scientific reasoning	**Planning of New, Related Investigation**
▪ Generating explanations based on underlying principles ▪ Forming scientific explanation from evidence	**Presentation of Results** Write report (Directives) Present oral summary (Directives)

The example design patterns introduced in Tables 2 and 3, *Viewing real-world situations from a scientific perspective* and *Re-expressing data*, are two of the design patterns we reverse-engineered from the GLOBE tasks and templates. Here we step through the basic elements of each of these design patterns and their relationship to the GLOBE template.

Example 1: Viewing Real-World Situations from a Scientific Perspective

Each GLOBE assessment starts with a problem presentation that situates a driving question in a real-world context, includes appropriate background information, and provides an initial data representation (e.g., multiple scatter plots, data tables, bar graphs). The goal of this phase of an assessment task is to have students display their ability to think about this driving question from a scientific perspective, as opposed to a personal, political, or naïve perspective.

[4] Design patterns found in the GLOBE assessment template: *Analyzing data relationships, Assessing the quality of scientific data, Generating explanations based on underlying principles, Interpreting data or observations, Re-expressing data, Scientific reasoning, Using resources, Using the representational forms of science, Viewing real-world situations from a scientific perspective*.

As noted in the "Focal KSAs" and "Additional KSAs" sections, this design pattern focuses on *how* students use their knowledge and understanding to apply a scientific perspective. As with all the design patterns we have developed so far (about 15), this design pattern is not content specific but can be adapted to focus on particular scientific content or models by adjusting the structure of the setting. An example of another task that targets thinking about situations from a scientific perspective, but with a different content area and a different form, is Chi, Feltovich, and Glaser's (1981) problem-sorting experiment. Expert physicists were observed to sort problems into categories based on fundamental relationships such as equilibrium, Newton's third law, or conservation of energy; novices sorted the same tasks on the basis of surface features, such as having to do with pulleys, springs, or inclined planes.

Example 2: Re-Expressing Data

The second design pattern, *Re-expressing data* (Table 3), focuses on taking one data representation or representational form and transforming it into another. On the surface, this design pattern centers around data transformation. However, at a deeper level, it involves recognizing the elements being addressed in a representational form, understanding the relationships among these elements as expressed in the representational form, and being able to identify, generate, and critique the mapping of these relationships into a different representational form. For example, in one instance of a GLOBE assessment task, the student is asked to take a table of GLOBE measurements taken over several months, average the data by week, and then plot the averages on an x-y axis. Here the student takes one representational form—a data table—manipulates the data, and transforms it into another representation—a line graph. During the reverse-engineering process, this design pattern was taken almost directly from a specific GLOBE investigation phase that requires students to generate a new representation from a given one. In pulling this design pattern directly from the GLOBE template we felt that it was general enough to use across multiple science content areas, and not just in relationship to the GLOBE content.

Although this design pattern can be considered more specific than *Viewing real-world situations from a scientific perspective*, it is still general enough to be effectively applied across many content areas and at different ability levels. For example, elementary students may be asked to create a bar or a pie chart from a relatively small data set of water temperature data, whereas an introductory college physics student might be asked to take the information from a plot of capacitor charge vs. time and replot it onto a logarithmic scale. In each case it is important not only whether the student can generate the data transformation but also whether the student understands the benefits (or drawbacks) of creating a new representation.

Same GLOBE Design Patterns, Different Tasks

Following are four examples of assessment tasks that can be viewed as instantiations of design patterns that emerged from reverse-engineering assessment tasks originally developed for GLOBE. These four examples are intended to suggest how tasks can look quite different on the surface, yet still provide evidence about the same, general underlying KSAs. They illustrate how features of the situations can be varied within a design pattern to produce tasks that vary in difficulty or focus.

Example 1: Flight of the Maplecopter

Design patterns addressed:

- Viewing real-world situations from a scientific perspective
- Designing and conducting a scientific investigation

Variable features:

- Provides cues to approach from scientific perspective
- Process is open (task does not provide step-by-step instructions)
- Requires prior substantive knowledge

This example from Baxter and Glaser (1998) prompts high school physics students to scientifically explore a real-world phenomenon: the flight pattern of a maple seed as it falls to the ground. More specifically, students are asked to design and carry out experiments with a maple seed to explain its flight to a friend who has not studied physics.

The flight of the maple seed represents a delicate equilibrium between gravity, inertia, and aerodynamic effects. Understanding how and why the maple seed falls as it does has drawn attention from a broad spectrum of scientists because of its complexity and the unresolved controversy over the most appropriate model. Given that the problem does not have a single, simple solution, it is rich with opportunities for scientific reasoning and exploration. Observations that would provide evidence of student competence include adequate representation of the problem, sustained and systematic exploration strategies, monitoring progress toward describing the flight of the maple seed, and explanation of the causal relationships observed and tested.

Example 2: Three Bottles

Design patterns addressed:

- Viewing real-world situations from a scientific perspective
- Designing and conducting a scientific investigation

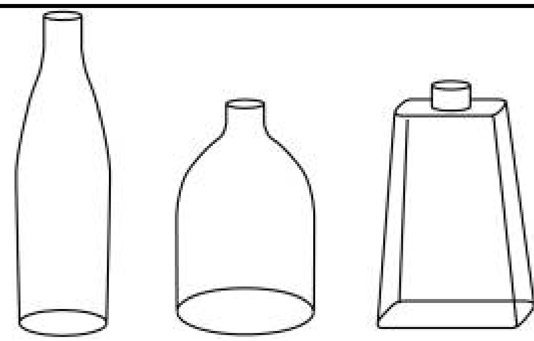
Variable features:

- No cues to approach from scientific perspective
- Process is open
- Requires little prior substantive knowledge

In this grade 4 science task from the National Assessment of Educational Progress (NAEP), students are shown a picture of three bottles of different sizes and shapes and asked to explain how they would figure out which bottle will hold the most water (see Figure 3). In contrast to the previous example, students are not cued to take a scientific perspective to solve the problem (besides knowing that they are taking a science test), so it assesses whether they are inclined to do so. The task requires basic knowledge and skills related to measurement, but no other prior content knowledge. Thus, while Examples 1 and 2 correspond to the same design patterns, Example 2 is geared toward a lower grade level and leaner in content. The two tasks also differ in terms of observations and work products: Example 1 actually requires students to carry out an investigation, whereas Example 2 simply asks students to produce a short, written explanation of how they would investigate the problem.

Figure 3. Three bottles task

You are going to the park on a hot day and need to take some water with you. You have three different bottles, as shown in the picture below. You want to choose the bottle that will hold the most water. Explain how you can find out which bottle holds the most water.



Source: NAEP grade 4 released task

Example 3: Plot of Four Planets

Design pattern addressed:

- Re-expressing data

Variable features:

- Single representational form (RF) to be translated into another RF
- RFs probably familiar to students
- Data not very complex
- Direct translation

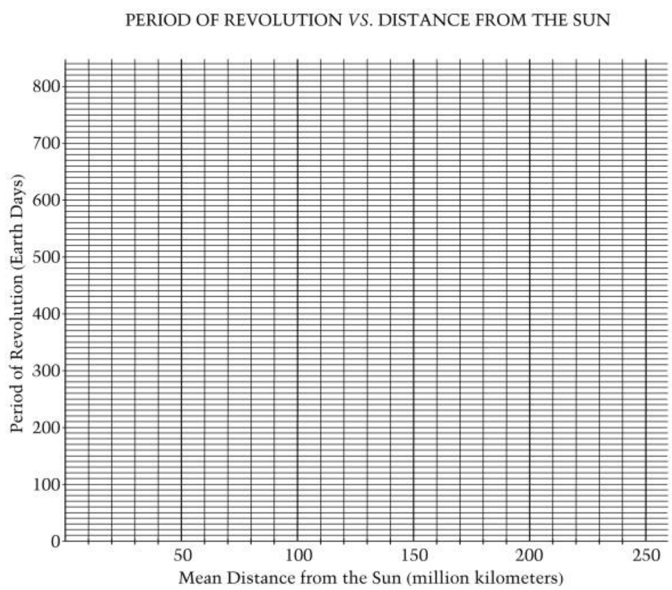
This NAEP grade 12 item (Figure 4) is an example of a relatively simple task that corresponds to the design pattern called re-expressing data. This task provides a lot of cueing for the student: the student is told which kind of RF to translate the data into (a line graph), given a structure for the graph, and told the order of steps to take. The data are also relatively simple, involving the relationship between two variables and just a few data points. There are some extraneous data in the table that students are to ignore, which adds a bit of difficulty to the task.

Figure 4. Plot of four planets task

The table below gives information about the planets: their periods of revolution about the Sun and rotation about their axes.

Planet	Mercury	Venus	Earth	Mars
Mean Distance from the Sun (million kilometers)	58	108	150	228
Period of Revolution (Earth time)	88 days	225 days	365 days	687 days
Period of Rotation (Earth time)	59 days	243 days	23.9 hours	24.6 hours

On the graph below, plot a point for each of the four planets showing the planet's period of revolution and its mean distance from the Sun. Then draw the line or curve that best illustrates the relationship between the period of revolution and the mean distance from the Sun that is suggested by the points.



Source: NAEP grade 12 released task

Example 4: Planetary Patterns

Design patterns addressed:

- Re-expressing data
- Participating in collaborative scientific inquiry

Variable features:

- Unfamiliar, complex representational forms (RFs)
- Two RFs must be re-expressed in a single RF
- Requires transforming data spatially

A released task from the Maryland School Performance Assessment Program (MSPAP) provides a more complex illustration of re-expressing data.[5] Grade 8 students work in groups and must take the information from two representational forms (RFs)—a table and an orbit diagram—and represent the information in a new RF called an "orbit data log." The context (distance of planets from sun) is similar to that of Example 3 at the surface level, but the demands of the task are considerably more difficult. The RFs that students are working with (orbit diagram and orbit data log) are probably unfamiliar to them, increasing the difficulty of the task.

Note that students work together during part of this task. We have developed a design pattern for group work that is applicable to assessing this aspect of inquiry, *Participating in collaborative scientific inquiry*. The contents of that design pattern are useful in setting up collaborative work, even when observations are neither made nor used in this aspect of an investigation, as was the case with this task.

Examples from BioKIDS: An Application of Design Patterns

One of the project goals of BioKIDS[6] is to design high-quality inquiry assessments to evaluate students' developing understandings of scientific inquiry related to the scientific concepts associated with animal biodiversity. In the initial run of the BioKIDS curricular program in schools (spring 2002), student understandings were evaluated in three areas:

- Science inquiry, particularly the inquiry concept of formulating scientific explanations from evidence.
- Science content, particularly the concepts of animal interactions, biodiversity, animal classification, and food webs.
- Fluency with technology, including demonstrated meaningful use of emerging technological tools, which includes the use of tools for inquiry and content understandings.

In spring 2002, the BioKIDS group designed three assessment instruments in three different formats to assess these multiple and interwoven understandings:

- A multiple choice assessment, consisting largely of released national and international items.
- An open-ended assessment, consisting largely of items designed by the BioKIDS project researchers themselves to closely match the work children were doing in classroom activities.
- A practicum exam, designed to run somewhat like a biology lab practicum with stations where students used particular resources, such as personal digital assistant (PDA) computers, for problem-solving and inquiry activities focusing on the use of the data available with that technological resource.

[5] This released task can be found at http://mdk12.org/share/publicrelease/planetary_task.pdf. Scroll about halfway down the document to get to "task #2."
[6] For more information about BioKIDS see http://groundhog.sprl.umich.edu/site/BioKIDS.html.

An ongoing issue with the design of BioKIDS assessment instruments has been both the need to develop comprehensive instruments that could capture the range and development of students' emerging understandings of complex science, and the need to articulate the kinds of levels of understanding that the various assessment instruments measured. Design patterns have helped the BioKIDS researchers to:

- Develop, articulate, and map similarities between test items because of the connection to the same inquiry idea or KSAs (e.g., formulating scientific explanation from evidence).
- Develop a suite of comprehensive items across the various tests through a systematic analysis of the items that represent the same design pattern yet are distinct in terms of content load (e.g., "content rich" versus "content lean").
- Develop a suite of comprehensive items across the various tests through a systematic analysis of the items that represent the same design pattern yet are distinct along a dimension of complexity [e.g., building explanation from evidence item in which students are given a claim and several options of relevant evidence, and students are asked to match the correct evidence, to the claim (low level), versus building explanation from evidence item where students are given a claim and asked to create evidence that supports this claim (higher level)].

In the next year, BioKIDS will use design patterns to help in creating a suite of comprehensive inquiry items that bridges its different curricular programs. This is an important goal for the program's research on the longitudinal development of inquiry thinking in Detroit Public Schools students as they encounter three or more coordinated inquiry units. The goal is to develop comprehensive assessment instruments that provide evidence of the development of inquiry skills over multiple curricular units and years.

Appendix B provides some examples of BioKIDS test questions and how they map to various design patterns in the start-up set. The BioKIDS group is currently developing some design patterns for their own program.

Examples from FOSS: Another Application of Design Patterns

The FOSS (Full Option Science System) *Populations and Ecosystems* course is a science curriculum designed for middle school students.[7] It is part of a larger project that includes 27 modules for elementary school and has been in use in classrooms since the early 1990s. The middle school project will eventually have nine courses completed for grades 6, 7, and 8. At the time that FOSS joined the PADI project, the *Populations and Ecosystems* course had been through its initial creation by the development team, working daily in middle school classrooms near the Lawrence Hall of Science, as well as through national field trials in 30 different classrooms across the country.

Formative and summative assessments are developed for each course, designed within the framework of three progress variables: science content, conducting investigations (inquiry processes), and building explanations. The FOSS goal in developing the assessment system is to provide teachers and students with feedback about student learning which can be

[7] For more information, see http://www.lhs.berkeley.edu/FOSS/FOSS.html.

used immediately to make decisions about further instruction. Part of this process involves not just reporting scores on individual items (which can result in an unwieldy amount of information to manage), but grouping scores by variable and classifying the resulting scores into criterion-based levels that describe a typical or average performance across a set of items. This process allows the teacher to see at a glance what a student's performance tends to be like and what kind of assistance might be needed to move a student to higher levels of performance.

Within each progress variable are several elements, specific to whatever course teachers are working on that, describe the proficiencies they are looking for with regard to student achievement. FOSS looked to design patterns to confirm the design of some of the assessments that had already been created and to help develop some new tasks, especially for the conducting investigations variable, which was an area they wanted to strengthen.

Before beginning to look at design patterns, though, the assessment group within the FOSS development team decided to go back to the investigations revised after national trials to define a clear overview of the curriculum. The course includes 10 investigations. The first seven investigations focus around understanding populations and ecosystems, including topic areas such as population growth, organizational aspects (such as food webs that identify producers, consumers, and decomposers), reproductive potential, limiting factors, and energy transfer through the system. The last three investigations focus around genetics, adaptations, and natural selection. For the purposes of this paper, we address here only the first seven investigations and the national standards that identify what students should understand about populations and ecosystems.

Once this decision was made, the FOSS researchers created a chart listing the standards from the NSES that applied, then the FOSS content objectives from each investigation that matched each standard. They also listed the objectives in three categories.

- Level I is characterized as intuitive knowledge that we expect many of our students to come to class with or very basic knowledge, such as definitions of vocabulary that students need to know how to use in scientific rather than everyday terms in order to begin the larger study of ecosystems.

- Level II includes content and processes that are fairly easy for students to learn, but requires some focused interaction or instruction in order for most students to consider when thinking about the study of ecosystems. This level describes fairly discrete aspects of the topic, but students are beginning to consider more factors and may begin to weave them together to see bigger parts of the system and how they work together. Common misconceptions may also emerge at this level as students begin to consider more things and build relationships among them.

- Level III describes the integration of knowledge we expect middle school students to achieve by the time they finish the course. At this level, students are called on to put many pieces of knowledge together to build multiple relationships and then integrate those relationships into broader understandings.

During this process, it became clear that content and inquiry cannot and should not be separated—the two develop together through an iterative process as the course progresses. One has to know something about the content in order to make decisions about what inquiry needs to be done, and one has to know something about inquiry in order to think about the kinds of data that can be gathered, how to gather that data, how to analyze it, and so forth, to add to and support growing content knowledge. So whereas the assessment team had often thought of content and process separately in the past, they now looked at how the two worked together and included them as interacting on the overview chart.

With a clear picture of the first seven investigations of the course, the FOSS team found it fairly easy to look at the start-up set of design patterns that had already been developed within PADI and decide which could be used in the development of the assessment system for the *Populations and Ecosystems* course. They chose three design patterns they thought would provide important information about students' progress for the conducting investigations variable:

- Viewing real-world situations from a scientific perspective
- Using resources to conduct scientific inquiry
- Interpreting data.

The FOSS researchers also thought they might want to develop one or two other design patterns, including "using mathematics in scientific inquiry" and some version of "different kinds of questions suggest different kinds of investigations." Many students in the field trials showed evidence of thinking that systematic observation, the preferred method of inquiry when studying ecosystems, was simply not as valid or as rigorous as a controlled experiment! For the purposes of this paper, however, we will focus the discussion on describing the application of only one design pattern: *Viewing real-world situations from a scientific perspective*. The following sections address some of the more involved parts of the design pattern. Appendix C provides a summary chart that describes all parts of the design pattern and how it would apply to *Populations and Ecosystems*.

Design Pattern: Viewing Real-World Situations from a Scientific Perspective

Summary and Rationale. In *Populations and Ecosystems*, students learn about the many factors that must be considered to understand the interactions within an ecosystem. Many of students' naïve understandings about organisms and their interactions must be transformed into more objective understandings in order to address some of the issues that face ecosystems around the world with a scientific view. Being able to take a scientific perspective will be important to students when they become adult citizens and need to make decisions, for example, about development that involves the disturbance of particular ecosystems.

Knowledge, Skills, and Abilities. As described generally in the design pattern, the focal inquiry skill addressed in this design pattern is *viewing real-world situations from a scientific perspective*. As discussed above, however, in the FOSS curriculum such skills are not viewed as decontextualized or independent from substantive knowledge; students are always reasoning with some body of knowledge. Here, then, is how this focal KSA becomes instantiated for the tasks FOSS researchers began to envision.

The substantive KSAs crucial to being able to untangle any given ecosystem are presented in several ways in the curriculum. Students gain firsthand experience with population growth by raising milkweed bugs, bred to feed on sunflower seeds in a lab setting. Students begin by constructing a habitat for an adult male and female and then follow the development of a new generation from egg-hatching through several instars (intermediate stages of growth), to mature adults laying eggs of their own. Students study interactions among other organisms using firsthand observations of aquatic and terrestrial mini-ecosystems. These investigations require students to use systematic observation over several weeks to note the interactions among the organisms, as well as the possible effects of abiotic factors within the mini-ecosystems.

Students receive additional information about the study of ecosystems through video (one about Jane Goodall and her long-term observation of chimps; another to introduce the Mono Lake ecosystem with a brief description of the biotic and abiotic factors involved there). Several readings, as well as several multi-media resources, also provide basic information about the study of ecosystems. These include an organism database, food web construction for the Mono Lake ecosystem, a simulation to look at milkweed bug reproduction (limited vs. unlimited), and a program that simulates the interactions of a few organisms in a small community.

Additional KSAs. For the tasks that FOSS researchers began to envision, students would need to bring to the task general knowledge about the practices scientists use to study ecosystems, as well as some of the content in terms of the organization of ecosystems (food webs as models of producer, consumer, decomposer relationships, etc.), other interactions within ecosystems, and energy transfer. Note that these are important knowledge and skills for the inquiry tasks envisioned, and for which students will have been assured adequate preparation through the FOSS program and the known relationship of the inquiry tasks to it. Thus, the structure of the curriculum and the timing of the assessment will have ensured that lack of this necessary background will not be a barrier to students. If the tasks were used in a drop-in-from-the-sky assessment, they certainly would become a primary—and probably unintended—source of difficulty.

The FOSS researchers began to plan a set of assessment tasks that would be given at different points during the course, so that they could observe changes in sophistication of understanding and viewpoint regarding ecosystems. Several small tasks would be developed for midcourse assessments given at the end of Investigations 2, 4, 5, and 6. Then a larger task would be embedded within Investigation 7, in which students work in groups to research real-world ecosystems that have received national attention.

Potential Observations. Next, the design pattern guided the researchers to think about how these tasks might be presented. There were several suggestions in the design pattern, and these seemed like viable options for some of the midcourse tasks they would develop. Adapting these specifically to the *Populations and Ecosystems* course, they might ask students to (1) pose a scientifically answerable question, given a particular ecosystem scenario; (2) have students explain how to get started investigating a particular issue within an ecosystem; (3) identify next steps, given an investigation already under way; and (4) critique responses offered by other students, or possibly the arguments of people involved in an ecosystem controversy. Another potential observation included a task already present in the curriculum. This was the study that students complete in

Investigation 7: Ecoscenarios—finding out about the interactions of a real-world ecosystem and looking at some of the current issues facing that ecosystem. Eight different ecosystems are each studied by a group of four students. The groups then present a description of the ecosystem they studied to the class and take a stand on the issues, providing evidence to support their stand from a scientific view.

Potential Work Products. Those suggested in the PADI design pattern included (1) a question students would ask (oral or written), (2) a diagram of a situation, and (3) an identification, from given possibilities, that reflects a scientific perspective. In Investigation 7, students would spend several days gathering information about their assigned ecosystems. Questions they asked or statements they made to the teacher could be used to assess whether they were in fact taking a scientific perspective. This process would precede the written report and oral presentation students would make to the class. These were also potential work products that could be used to gather evidence about students' viewing the world from a scientific perspective.

Potential Rubrics. The next thing to think about was what the teacher might actually see that would provide evidence about whether students were using a scientific perspective. The FOSS researchers made up a chart (see Table 9) that compared what they were looking for in terms of a scientific view vs. what they termed the "naïve view." The chart was based on what they had actually seen students do or say during the local and national trials.

Table 9. Comparison of scientific and naïve views in FOSS

Scientific View	Naïve View
Different questions suggest different kinds of investigations—sometimes systematic observation is preferable to a controlled experiment.	Systematic observation is not as rigorous or as "scientific" as a controlled experiment.
See organism objectively with certain behaviors but not human attributes.	Anthropomorphic comments made.
Animals are generally found in a certain range.	Animals choose to go places to look for a better environment (similar to people who move to another place for a better climate).
Most animals have more than one food source, but they don't simply chose to eat something else if their usual food sources disappear.	Animals choose to eat something else if their normal food source is not available.
Choices about controversial issues based on objective evidence, not personal or emotional preferences.	Personal/emotional choices (cute vs. ugly; plant eater [gentle] vs. predator [violent]).
Feeding relationships are not restricted to those represented in the food web (food web is only a model, it doesn't necessarily include all organisms).	Feeding relationships are restricted to the food web represented, and that's all that interacts.
Able to see similarities between different ecosystems; can group ecosystems into large categories, e.g., desert, temperate forest, rainforest, coral (marine).	Students don't always relate a particular ecosystem they are studying to a larger group of similar ecosystems (e.g., desert, ecosystems). Transfer of information is minimal.
Scientific use of vocabulary.	Everyday use of vocabulary.

FOSS currently bases most of its scoring guides on the taxonomy presented in Table 10, adapted from the SOLO taxonomy. (This taxonomy is still evolving and changes as more is learned about student learning as each new course is created.) The researchers needed to decide which sorts of scientific thinking vs. naïve thinking fit into which levels of scoring.

Table 10. SOLO taxonomy

Level	Description
4: Rational/Mastery	The student is able to provide more than one piece of content knowledge and is able to put those pieces of content into a relationship. All content knowledge is correct and all connections/relationships/conclusions are made correctly.
3: Relational/Progress	The student is able to provide more than one piece of content knowledge and is able to put those pieces of information into a relationship. The information may have minor errors, or the relationship may have minor errors, but all information is relevant and it is clear that a connection has been attempted between pieces of content knowledge that contribute to the understanding of a larger concept.
2: Multistructural	The student is able to provide more than one piece of content knowledge that is relevant to the task or question asked, but no connections are being made between pieces of knowledge.
1: Unistructural	The student is able to provide one piece of content knowledge related to the question asked.
0: No Attempt/ Irrelevant	The student does not respond to the question or task, or gives and answer that has nothing to do with what was asked.

Characteristic and Variable Features. For the tasks that would be developed for the midcourse assessments, students would be provided the pertinent information to answer a series of questions involving different ecosystems. For the ecoscenarios, the researchers decided that the characteristic features would include giving students specifications about what should be included in their reports. Ideally, they would like to provide just enough information to get the students interested in any given ecoscenario, but practically, they may have to provide more information, or at least make it an option for teachers. In many classrooms, only one computer is available during most sessions, making it impossible for an entire class of students to search the Internet for additional references.

The variable features from the design pattern would again vary with the task at hand, depending on the ecosystem described and the nature of the questions that would follow. For the ecoscenario task, the substantive knowledge about the study of ecosystems would have been developed during the course of the first six investigations—the specific populations and abiotic factors that make up a particular ecosystem would provide the details.

Sample Tasks Developed from the Design Pattern

Sample task #1 (see Figure 5) would be given to students after Investigations 1 and 2. In these investigations, students are introduced to some of the conventions that scientists use to study ecosystems, including the vocabulary to identify various parts of the system, the need to consider both biotic and abiotic factors within the system, and how systematic observation is important when studying ecosystems.

Figure 5. Sample FOSS task #1

Directions: Read the Lab Report on the next page. Then answer the following questions. If you need more room, write on the back of this sheet.

A. What questions could you ask these students to help them improve the setup of their habitat?

B. What could you tell these students to help them write better observations?

Lab Notebook entry:

Cricket Habitat Setup—
We decided we wanted to raise some crickets in the classroom to see how many we could get in two months. We built a habitat for the crickets as follows:

1. We used a ten-gallon glass terrarium for the habitat. On one side of the terrarium we put dirt, since we know they lay their eggs in dirt. On the other side we put dry sand, and in between we placed egg cartons for a climbing structure.
2. We decided to use a light bulb to provide heat. The light always shines down on the middle section where the climbing structure is.
3. We put two crickets in the habitat to see what would happen.

Cricket Observations—
The crickets were sitting next to each other in the middle section under the light. They looked happy together.

One cricket was sitting in the dirt, the other was sitting in the sand. Both of them were making chirping noises. They must have been talking to each other.

Both crickets are dead and there are no baby crickets observable.

Sample task #2 (Figure 6) would be given at the end of Investigation 4, after students have been introduced to the interactions among organisms and their roles as producers, consumers, and decomposers. They have also been introduced to several issues involving the Mono Lake ecosystem.

Figure 6. Sample FOSS task #2

> In a natural area, there existed a community of organisms that included deer and mountain lions. When this natural area became a popular recreation area, people suggested removing the mountain lions for safety reasons.
>
> What would you do to study this ecosystem in order to help people understand why they should or should not remove the mountain lions?

Sample task #3[8] is embedded within Investigation 7. Students are assigned an ecosystem to study with a group of three other students. They are given some basic information about their assigned ecosystem; then they prepare a report and a presentation to the class. In each case, they are required to describe the particular ecosystem, construct a food web to represent the interactions within the ecosystem, and describe a controversial issue that people are debating about that ecosystem. Further, the students need to take a stand in the debate and present evidence from a scientific point of view for one side of the argument.

[8] Sample ecoscenario task will be posted on the FOSS Web site when the task is in final form.

The Structure of Design Patterns, Part 2: Software Design Process

In the preceding sections, we have elaborated a structure for design patterns and given a variety of examples of their applications. The next question is: how will we plan, design, and construct the PADI design system given this structure? We need to choose a notation in which we will capture our knowledge and design decisions about the system, develop a detailed specification of features that the system must support, and define a software architecture for the system. In this section, we discuss our processes for elaborating the functional specifications of our system; present our notation for representing components of the system, including design patterns; and present screen shots from a software prototype that we have developed to support browsing, viewing, editing, relating, and creating design patterns.

Determining Functional Requirements: Use-Cases

One of the most important activities in any software development process is to define the primary use-cases for the system. Use-cases are narrative, short stories, often a paragraph or two, about people who represent each type of user of a system—for example, teachers, curriculum developers, and students—and how they interact with the system.[9] We have developed about a dozen use-cases for the PADI design system (see example in Figure 7 and Appendix D) and after several iterations, we have converged on a set that is fairly comprehensive and captures the primary requirements of PADI.

[9] See Armour and Miller (2000), for a general discussion of use-cases and their role in systems design.

Figure 7. Sample use-case about creating a design pattern

> Anders is a researcher who wants to reverse-engineer some design patterns from the existing GLOBE assessments. Every GLOBE assessment introduces some motivating problem and data, and then asks students to pose a question, along with other activities. Anders intends to capture the essence of these assessment tasks, and he starts by creating a new instance of a design pattern in the PADI system. He thinks about why the student is being asked to come up with a question. He sees that the GLOBE assessments both explain a realistic problem and present data that are relevant to the problem. Anders notices that the GLOBE assessments do not tell the student what kind of model to use, what domain of science to think about, or even to think about science at all. Thus the student is given a chance to bring, or not to bring, a scientific perspective to a real-life problem (though the student knows he/she is are doing a science unit). Anders enters this generalization into the design pattern description under the subtitle "characteristic features." He also indicates that this design pattern is linked with an observable called "Posing a scientifically answerable question." As he thinks further, Anders comes up with a number of different ways he might get clues about the same essential knowledge/skill/ability (KSA). For example, he might have a student draw a diagram of a situation and then explain to a friend how to get started investigating the problem, including a list of possibilities as to what are good next steps. At this point, Anders is not worried about specific tasks, specific delivery mechanisms, or other details; each of these could play out in many ways. As he thinks through similar means to elicit questions, he adds other notes to the PADI design pattern, including the features of the situations that create the right effect. Some of these features are important with some kinds of observations but not others, so he indicates the key relationships and annotates each with a sentence or two of his reasoning. Anders decides to call this design pattern "View real-world situations from a scientific perspective."

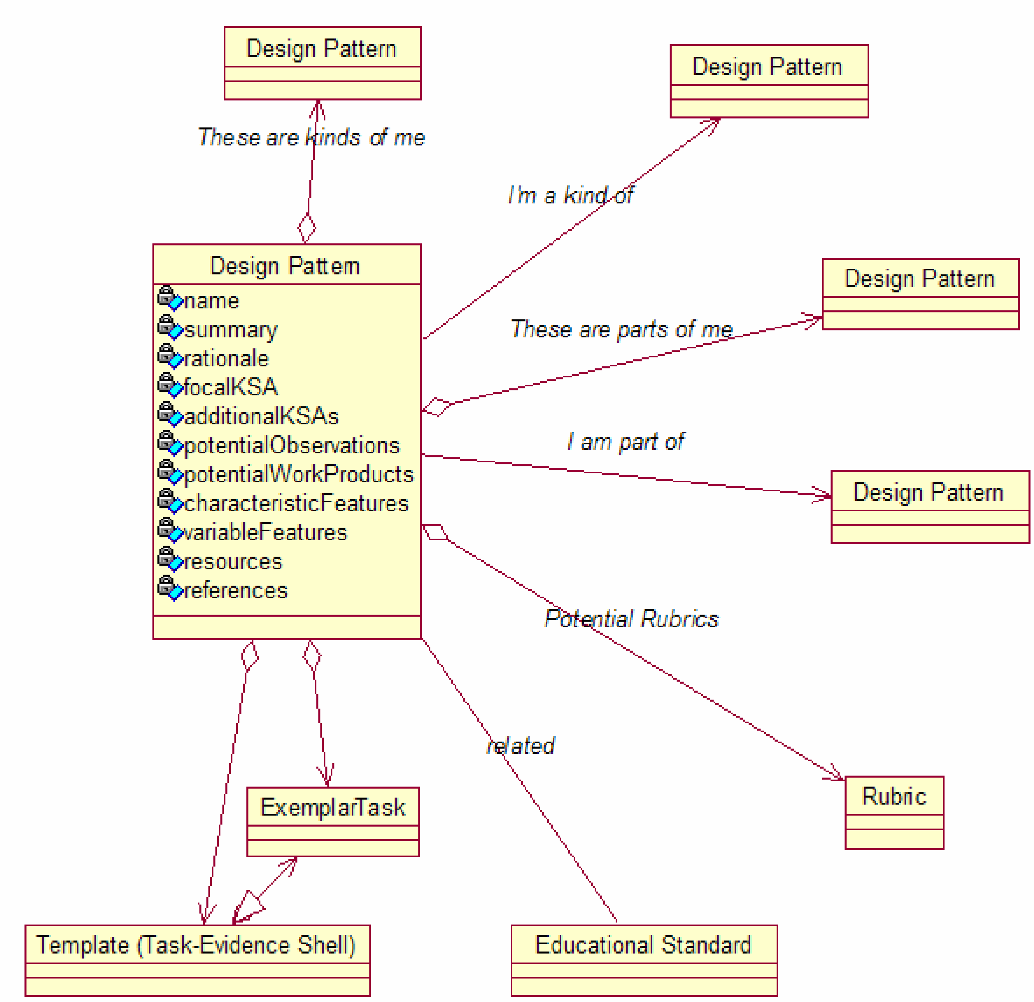
For some of the use-cases, we drew sketches that showed a proposed flow of pages that the person would see when doing the task described in the use-case. These mockups were used later as a basis for the prototype described in more detail below.

Representing Design Pattern Components: Object Modeling

Armed with the knowledge about the structure of design patterns and the features specified in the use-cases, we needed to define the components of the system and how they interrelate. We are using an object model representation for this purpose. An object model is a taxonomy of a system and interrelationships: roughly, a model in which one identifies all the nouns in the system––such as design patterns, tasks, rubrics––and the attributes for each noun and how they interrelate. We are using the Unified Modeling Language (UML) notation to depict the object model.

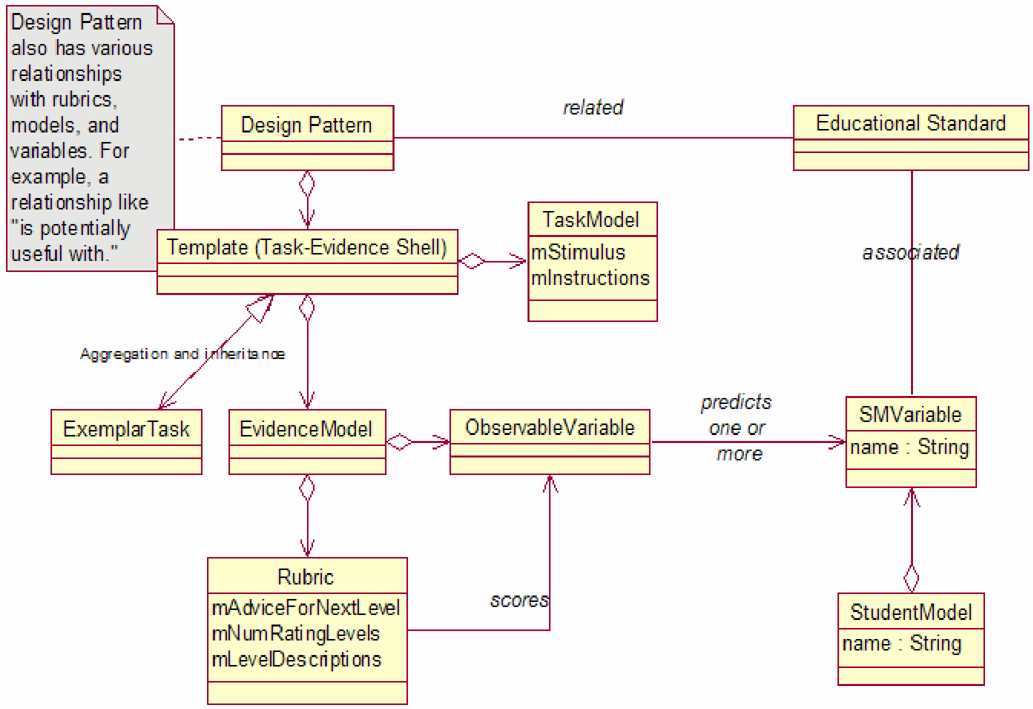
Figure 8 shows the UML diagram for the Design Pattern object, which will be a part of the larger PADI design system. In UML, an object is represented by a rectangle with three compartments. The top compartment shows the object name, the list of attributes appears in the middle compartment, and the list of operations that the object can carry out appears in the bottom compartment (not shown here).

Figure 8. Object model for design pattern in UML notation



The object model for the full PADI design system is in the initial stages of development. To provide a feel for the relationship of design patterns to other, more technical design elements, Figure 9 provides a provisional high-level view of some of the important objects.

Figure 9. Higher-level object model diagram, including design patterns and other components

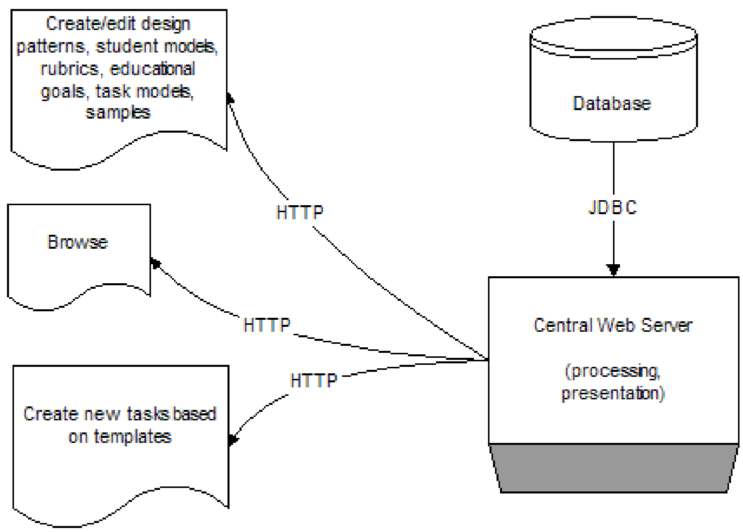


Elaborating the System: Constructing a Prototype

Once we had a better understanding of the components and functionality of our system as defined by our use-cases and object models, we put together a "birds-eye" view of the overall architecture of the system, including the high-level components of the system, communication protocols, and data formats, and developed a prototype to demonstrate and test our design. For implementation, we are using a three-tier architecture, a type of client-server architecture consisting of three well-defined and separate processes: the user interface (the client) that the user interacts with, the functional modules (application tier) that actually process data, and the database management system (backend tier) that stores the data (see Figure 10).[10]

[10] Note that the focus in PADI is on the assessment design system, not systems for authoring, delivering, or using any particular assessment. An application system, such as might be employed by FOSS and BioKIDS, would have its own object model for carrying out these processes. An assessment application system might include modules such a scoring engine, a database to handle student score information, and user interfaces to allow teachers to deliver assessments, input student scores, and view scoring engine output. The relationship between the PADI design system and an application system would be that the PADI system creates a conceptual structure and design specifications for the objects and processes that constitute an assessment application system.

Figure 10. Prototype application system architecture for PADI design system



The PADI architecture is also based on a Model View Controller (MVC) design that separates core data access functionality from the presentation and control logic that uses this functionality. Such separation allows multiple views to share the same enterprise data model, which makes supporting multiple clients easier to implement, test, and maintain.

After a review of several candidate frameworks, the PADI technology team decided to build a prototype on JCorporate's open-source Expresso Framework.[11] Expresso is a three-tier, model view controller architecture implemented in Java, build on top of the popular Jakarta Struts Framework.[12]

At present, we have developed a rudimentary prototype that implements the basic functionality for browsing and editing design patterns. The following screen shots provide a feel for what a PADI user would see upon entering into the PADI prototype, begining with the PADI entry page (Figure 11). A user (with editing privileges) might see a list of design patterns (Figure 12), view the contents of a selected design pattern (Figure 13), and edit a design pattern's attributes (Figure 14). A user who is not granted edit permissions would not see the Edit or Add links, and would only be able to view information about design patterns in the system (Figure 15).

In the future months we will be elaborating and refining this prototype (including functionality and look and feel) and developing use-cases and object models for the remaining components of the system.

[11] http://www.jcorporate.com.
[12] http://jakarta.apache.org/struts.

Figure 11. Entry page.

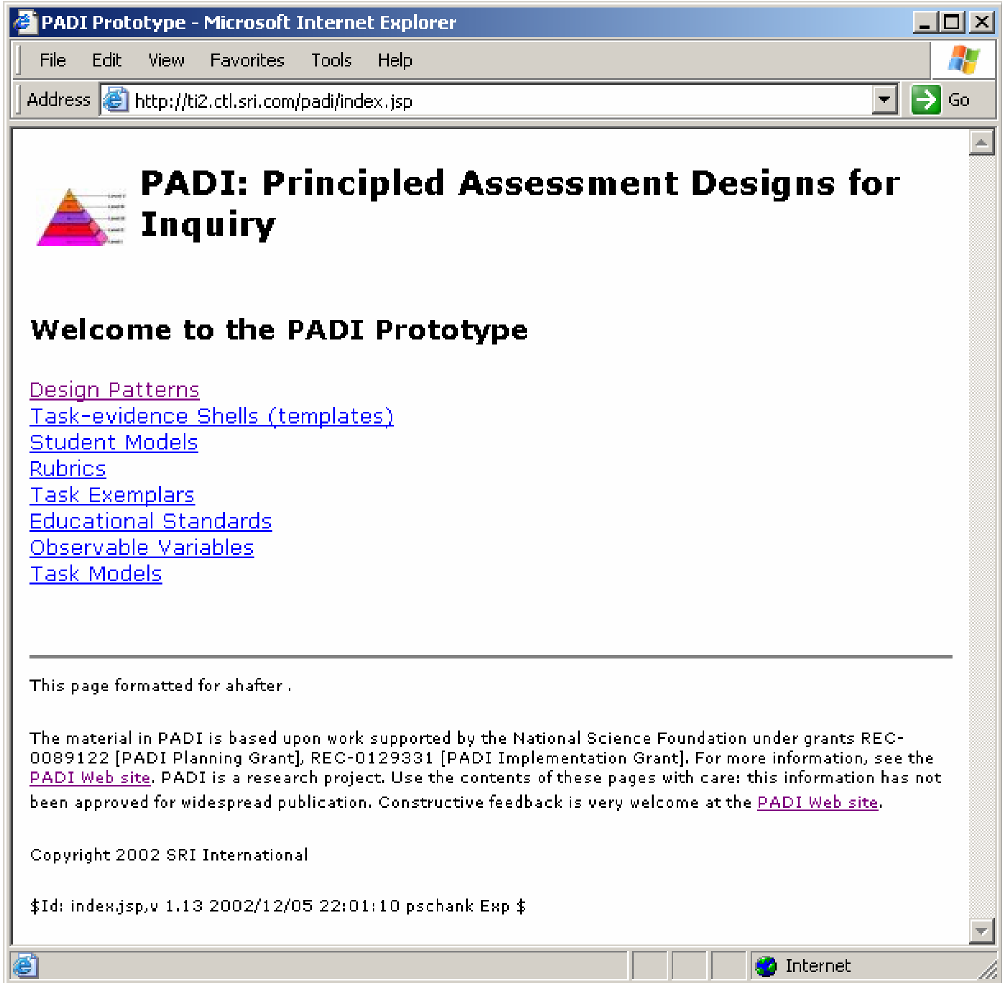


Figure 12. Design pattern list, for user with edit permissions.

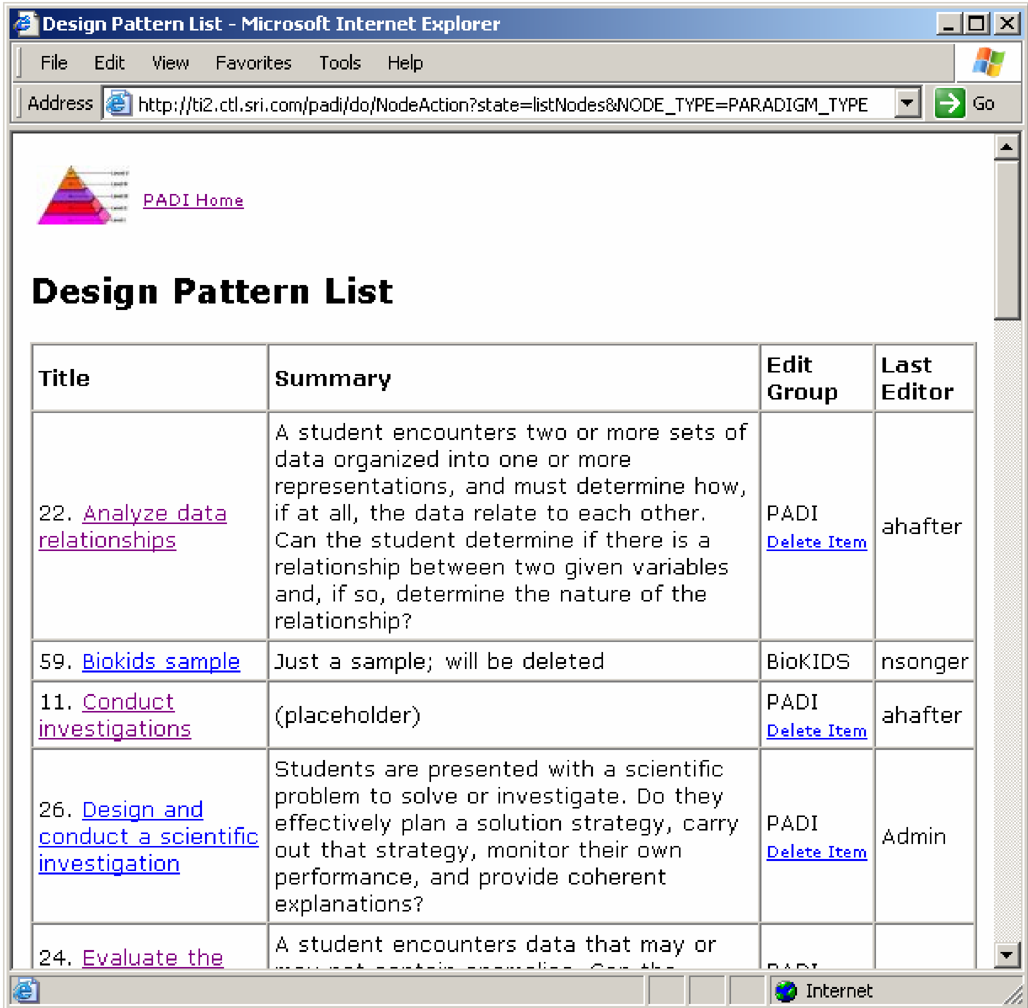


Figure 13. Part of a design pattern information page, with editing privileges.

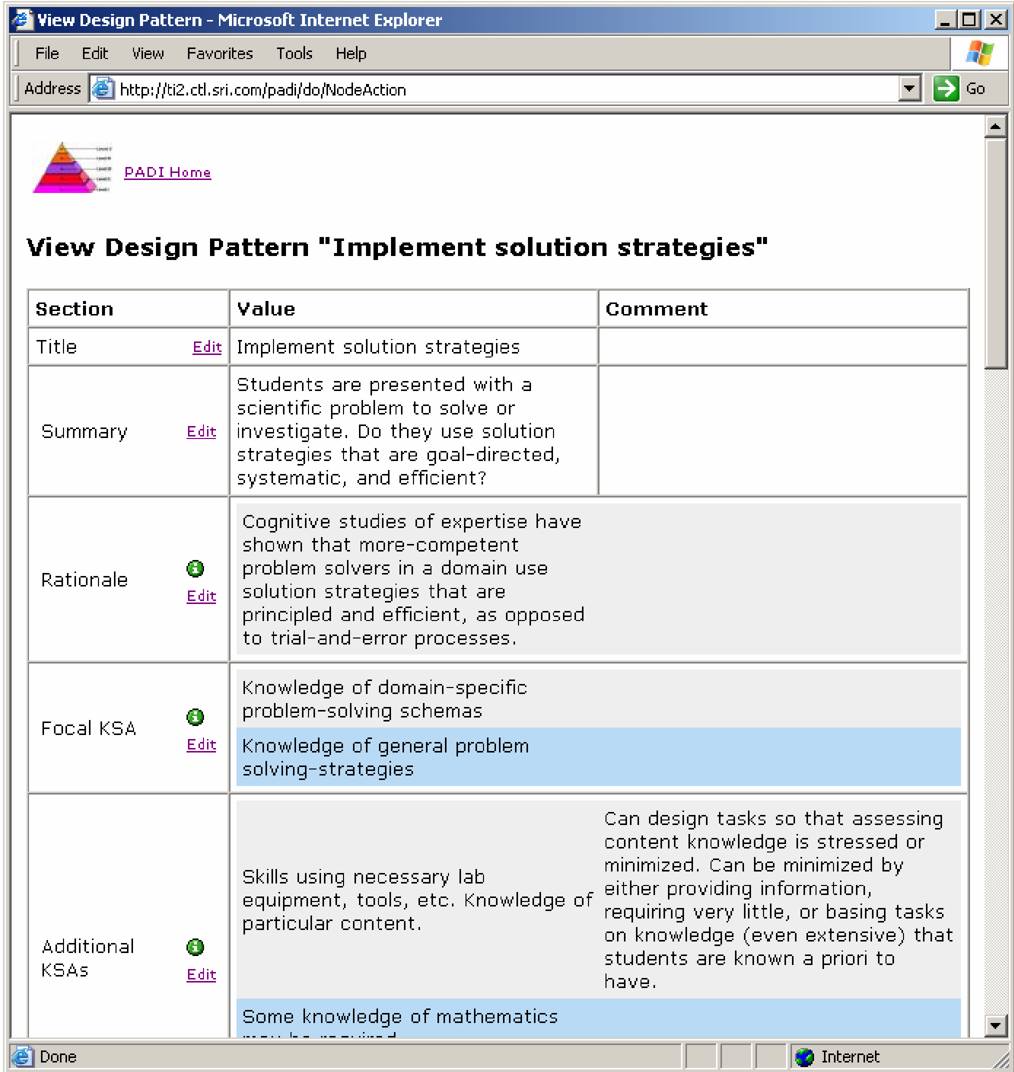


Figure 14. Part of a design pattern attribute's page, with editing privileges.

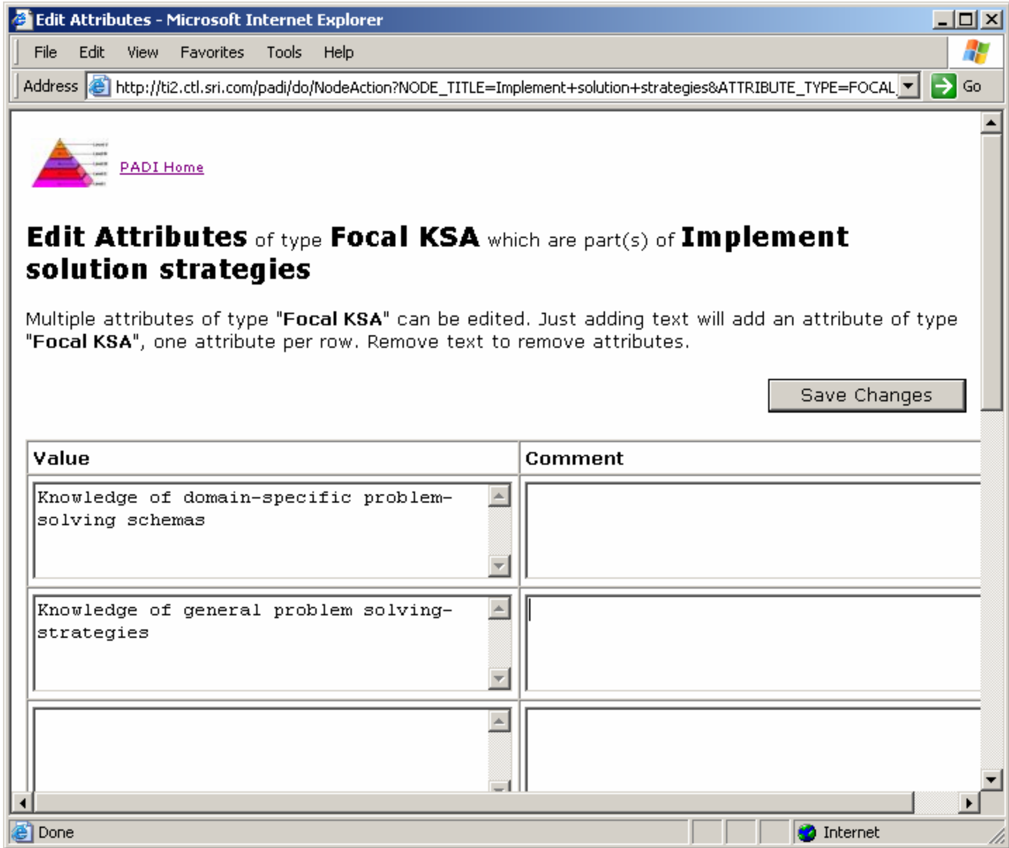
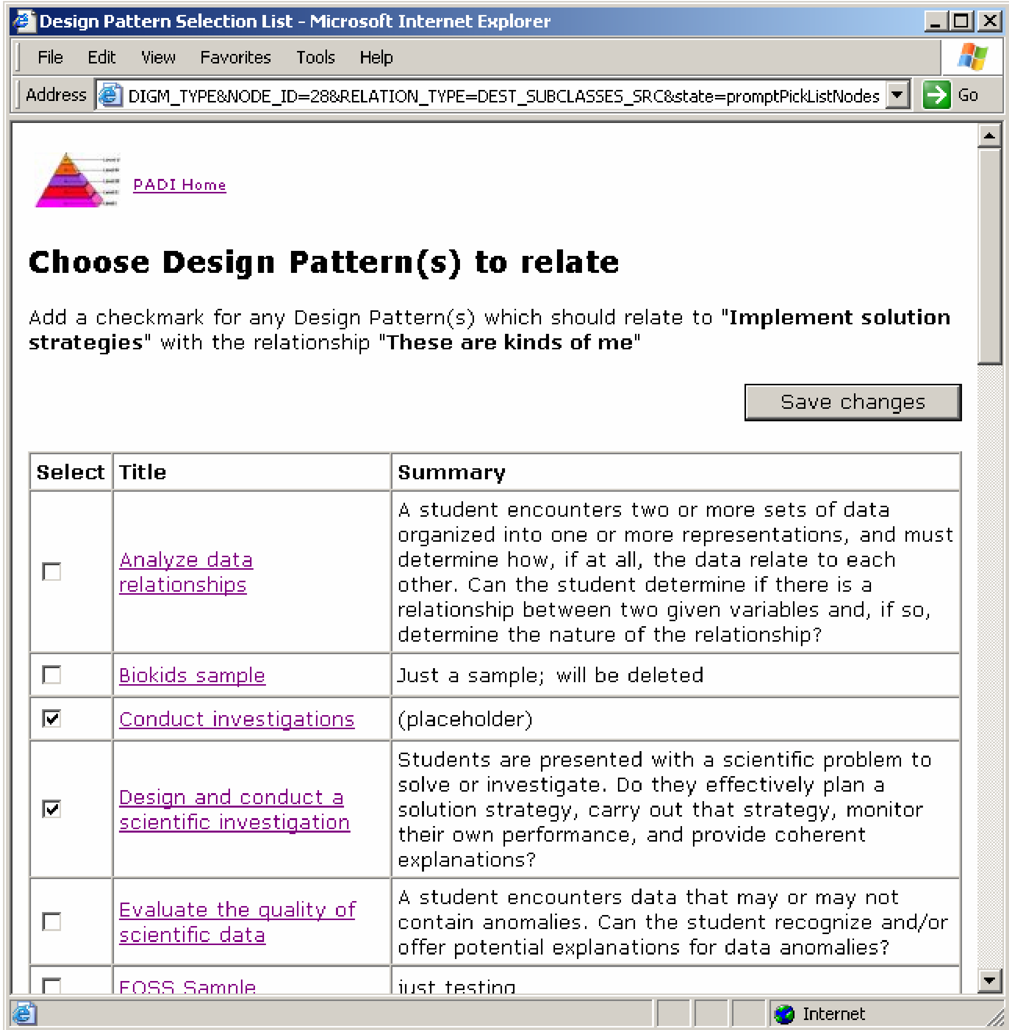


Figure 15. Part of a design pattern's relationship page, with editing privileges.



Looking Ahead to the More Technical Design Elements of PADI

This brief section looks ahead to future work, to give the reader a feel for how design patterns will be connected with more technical design structures in the PADI design framework.

Whereas design patterns define a narrative structure for getting evidence about some aspects of science inquiry, the more detailed, technical specifications for tasks are defined in a later stage of the PADI system, in design objects that we are calling *task-evidence shells*. A filled-in task-evidence shell provides a blueprint or complete set of specifications for creating a family of tasks. At this level of the PADI system, the student, task, and evidence models of the evidence-centered design framework are specified. We return to the GLOBE example to briefly illustrate some of the contents of task-evidence shells.

The student model in a psychometric model specifies the variables in terms of which we wish to characterize students. The nature and number of student model variables express an assumption about the purpose of the assessment. Although GLOBE has no proposed statistical model, it does suggest a space of potential models that are all compatible with the task template but would be suited to different assessment contexts or purposes. One potential model is an overall measure of student proficiency. All of the tasks would combine to provide evidence of one overall proficiency variable.

Another possible student model would be one in which content and inquiry were separated into separate measures of proficiency. In this case, there would be two student model variables, whose skills are separate and independent: domain knowledge and inquiry skills. The assumption of this model is that a student may know all the content necessary to be proficient in a specific domain but be unable to apply that content knowledge to scientific inquiry.

Yet another potential model is a student model that has a very small grain size and assesses student proficiency in several small subcontent and subinquiry skills areas. Separate student model variables are used to manage belief about these skills, which are theoretically discernible even though they may be called on jointly to solve problems. Multiple student model variables are necessary when multiple aspects of knowledge are required in combination to support a claim and when students can possess differing proficiency in those skills.

In the evidence-centered design model, assessment designers develop a set of task model variables by considering concurrently the substantive and procedural points of view. A task model (TM) is a design object that bridges the gap between the area of proficiency we are interested in looking at and the operationalizing of tasks that will help to demonstrate that proficiency. A PADI task model will be a data structure in which specifications for task materials, instructions, presentation requirements, and so on, are delineated by or for task authors. This design work will be guided by the prior selection (or creation) of design patterns, which provide a narrative description and rationale for the more technical specifications of classes of tasks. With regard to science inquiry, the GLOBE task model consists of rather specific variables and instructions. As a whole, the task model incorporates beliefs about the nature and structure of tasks that will allow students to demonstrate their proficiency in both science content areas and inquiry. Note that the task model itself is a blueprint for constructing tasks; it does not just indicate why or how the elements of tasks contribute to the assessment argument. It is the design pattern that lays out the structure of the assessment argument.

Each task represents a formalized notion of features of performance situations (key instances of which are discussed at a more general level in design patterns). The basis of GLOBE assessments is a general template that teachers can use to develop integrated investigation problems. The assessment designers developed these templates, rubrics, and exemplars to guide individual teacher assessments. Teachers may use the general framework to choose the assessment components that are appropriate for their particular classroom goals. Therefore, the variable features of the task and the different levels of each task represent different levels of proficiency within science inquiry. For a particular task, the values of task model variables, such as high, moderate, and low data presentation difficulty, are data for the argument of proficiency.

Creating the task evidence shell structure and producing filled-in examples, by reverse-engineering again from GLOBE, BioKIDS, and FOSS, represents the next major stage in our work. Creating specifications for new families of assessment tasks, then authoring and field testing the resulting tasks, will follow.

References

Adams, R., Wilson, M. R., & Wang, W. C. (1997). The multidimensional random coefficients multinomial logit model. *Applied Psychological Measurement, 21*, 1-23.

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

Almond, R. G., Steinberg, L. S., & Mislevy, R. J. (1999). *A sample assessment using the four process framework*. White paper prepared for the IMS Working Group on Question and Test Inter-Operability. Princeton, NJ: Educational Testing Service.

American Association for the Advancement of Science (AAAS). (1993). *Benchmarks for science literacy: Project 2061*. Oxford University Press, New York.

Armour, F., & Miller, G. (2000). *Advanced use-case modeling*. Boston: Addison-Wesley.

Baxter, G. P., & Glaser, R. (1998). Investigating the cognitive complexity of science assessments. *Educational Measurement: Research and Practice, 17*(3), 37-45.

Baxter, G. P., Elder, A. D., & Glaser, R. (1996). Knowledge-based cognition and performance assessment in the science classroom. *Educational Psychologist, 31*(2), 133-140.

Bennett, R. E. (1999). Using new technology to improve assessment. *Educational Measurement: Issues and Practice, 18*(3), 5-12.

Bransford, J. D., & Schwartz, D. (1999). Rethinking transfer: A simple proposal with multiple implications. In A. Iran-Nejad & P. D. Pearson (Eds.), *Review of research in education* (Vol. 24, pp. 61-100). Washington, DC: American Educational Research Association.

Center for Technology in Learning. (2001). *GLOBE year 5 evaluation*. Menlo Park, CA: SRI International.

Chi, M. T. H., Feltovich, P. J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science, 5*, 121-152.

Coleman, E. B. & Penuel, W. R. (2000). Web-based student assessment for program evaluation. *Journal of Science Education and Technology, 9*(4), 327-342.

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

Greeno, J. G., Collins, A. M., & Resnick, L. B. (1997). Cognition and learning. In D. Berliner & R. Calfee (Eds.), *Handbook of educational psychology* (pp. 14-47). New York: Simon & Schuster Macmillan.

Greeno, J. G., Pearson, P. D., & Schoenfeld, A. H. (1996). *Implications for NAEP of research on learning and cognition*. Report of a study commissioned by the National Academy of Education. Panel on the NAEP Trial State Assessment, conducted by the Institute for Research on Learning. Stanford, CA: National Academy of Education.

Kadane, J. B., & Schum, D. A. (1996). *A probabilistic analysis of the Sacco and Vanzetti evidence*. New York: Wiley.

Kuhn, T. S. (1962). *The structure of scientific revolutions*. Chicago: University of Chicago Press.

Levi-Strauss, C. (1958). *Structural Anthropology.* Allen Lane, The Penguin Press.

Means, B., & Haertel, G (2002). Technology supports for assessing science inquiry. In *Technology and assessment: Thinking ahead* (pp. 12-25). Washington, DC: National Academy Press.

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

Mislevy, R.J., Steinberg, L.S., & Almond, R.G. (2002). On the structure of educational assessments. *Measurement: Interdisciplinary Research and Perspectives*, *1*, 3-67.

National Research Council. (1996). *National science education standards.* Washington, DC: National Academy Press.

Pellegrino, J., Chudowsky, N., & Glaser, R. (Eds.). (2001). *Knowing what students know: The science and design of educational assessment*. Washington, DC: National Academy Press.

Board on Testing and Assessment. (2002). *Technology and assessment: Thinking ahead*. Proceedings of a workshop. Washington, DC: National Academy Press.

Polti, G. P. (1868). *The thirty-six dramatic situations*. L. Ray (Translator). Boston: The Writers, Inc.

Schum, D. A. (1994). *The evidential foundations of probabilistic reasoning*. New York: Wiley.

White, B. Y., & Frederiksen, J. R. (2000). Metacognitive facilitation: An approach to making scientific inquiry accessible to all. In J. Minstrell & E. van Zee (Eds.), *Teaching in the inquiry-based science classroom*. Washington, DC: American Association for the Advancement of Science.

APPENDIX A

A Narrative Description of a Framework for Developing Student, Evidence, and Task Models for Use in Science Inquiry

Appendix A

A Narrative Description of a Framework for Developing Student, Evidence, and Task Models for Use in Science Inquiry

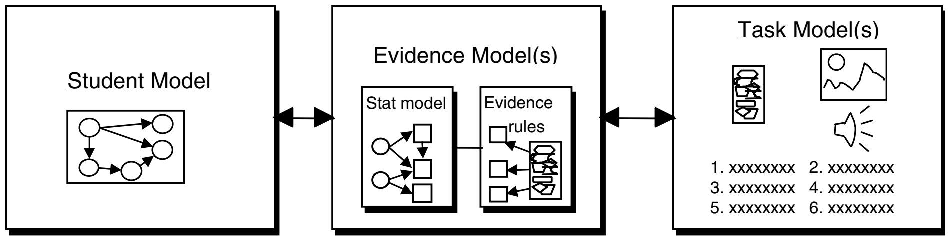
The SRI and University of Maryland research teams initiated discussions concerning ways to elaborate the Conceptual Assessment Framework (CAF) for application to science education (See Figure 1 below). In contrast to applications for training workplace skills, science education is concerned with broader goals, understandings, and skills that will support later learning. It is not possible to pinpoint a small set of specific critical tasks for science inquiry in the way one can for dental hygiene or hydraulic system troubleshooting. Nevertheless, the need for a student model based on an analysis of cognitive requirements remained. Two decades of research on the nature of learning in important content areas, including science inquiry (see Bransford, Brown, & Cocking, 1999), provided a resource for identifying key elements of science inquiry and, in the case of some of those elements, the typical trajectory or stages of development. The purpose of PADI-supported assessments will be to differentiate levels of proficiency on these understandings and skills identified in the standards and through learning research.

The templates to be used for generating assessments were described in terms of models of the Conceptual Assessment Framework (CAF). That is, the structures, relationships, salient features of tasks, and generally stated evaluation rules laid out in terms that were sufficiently general to guide task design in many domains of science.

The Conceptual Assessment Framework (CAF). Evidence-centered design encompasses both assessment design and assessment delivery. The design portion of this approach reflects Messick's (1992) emphasis on the importance of starting the assessment design process with a thoughtful consideration of just what one wishes to assess.

Mislevy and his colleagues have extended Messick's idea into an explicit framework for assessment development, called the Conceptual Assessment Framework, for short. Conceptually, the core of CAF is the evidentiary reasoning that links elements of students' work to scoring and to inferences about students. Figure A-1 is a high-level schematic of the three central models in this framework. The exhibit also shows how these models map onto the guiding questions of assessment design.

Figure A-1. The three central models of the Conceptual Assessment Framework (CAF)



Key Guiding Questions	***What complex of knowledge, skills, or other abilities should be assessed?***	***What behaviors or performances should reveal the relevant knowledge and skills described in the student model?***	***What tasks or situations should elicit the behaviors or performances described in the evidence model?***

Student model. Within this framework, assessment design begins with the simple question "What complex of knowledge, skills, or other abilities should be assessed?" Configurations of values of student-model variables approximate selected aspects of the infinite configurations of skill and knowledge real students have, as seen from some perspective about skill and knowledge in the domain. The number and nature of the student model variables in an assessment also depend on its purpose. A single variable characterizing overall proficiency might suffice for an assessment just meant to support a pass/fail decision, but a larger number of more detailed variables would be needed for a coached practice system designed to help students develop the same proficiency. For research purposes or to support classroom teaching and learning, an intermediate level of detail is likely to be appropriate. Given a student model at the appropriate level of detail for the assessment's purpose, we use a statistical model to manage our knowledge about a given student's (unobservable) values for these variables in terms of a probability distribution that can be updated in light of new evidence

Evidence models. What behaviors or performances should reveal the relevant knowledge and skills described in the student model? An evidence model details how observations for a given task situation constitute evidence about student model variables. Exhibit 1 shows that there are two parts to the evidence model. One part ("evidence rules") is about evaluating the key features of what the student says, does, or creates in the task situation—the "work product." These are the "observable variables," evaluations of whatever the designer has determined are the key aspects of the performance. The other part ("stat model") of the evidence model is about the way that the observable variables depend, in probability, on student model variables. This is how we combine evidence across tasks. Familiar psychometric models, such as item response theory and latent class models, can be seen as special cases of these ideas.

In the proposed, long-term PADI project, two distinct ways of evaluating students' performance will be outlined and built into the prototype student and evidence models. One is accumulating information about overall proficiency, or quality of response, and the second is to gather measures of multiple proficiencies. The most common way of evaluating tasks is the first of these, using a single measure of proficiency. Familiar models from classical test theory and item response theory can be used to accumulate information about tasks in this approach. This approach has the advantages of being familiar and simple, but it has two serious shortcomings. First, variation from one task to another can be high (the infamous "low generalizability" problem of performance assessment) if different mixes of complex skills and domain knowledge are simply collapsed into a single measure of response quality. Second, the opportunity to accumulate evidence about distinct aspects of component knowledge and inquiry skills is lost. For this reason, we will also plan to develop a framework with a student model and evidence models for modeling performance in terms of multiple skill and knowledge variables. This second approach better capitalizes on advances in measurement technology and cognitive psychology.

The aspects of proficiency in the PADI evidence models will be derived from (1) one or more sets of authoritative standards relevant to science education, and (2) understanding of expert and novice patterns of knowledge and behavior, both in specific science content areas and in inquiry skills that cut across domains, such as formulating and testing hypotheses and using and revising models.

Task models. Having thought through the behaviors or performances that reveal the constructs the assessment is targeting, we then ask, "What tasks or situations should elicit those behaviors?" A task model provides a framework for constructing and describing tasks (i.e., the situations in which examinees act). The variables one uses to describe tasks play many roles, such as guiding task construction, focusing the skills a task elicits, and providing an operational definition of the student model variables (Mislevy, Steinberg, & Almond, 2002). A task model includes specifications for the task environment, including, for example, characteristics of stimulus material, instructions, help, and tools. It also includes specifications for the work product, the form in which what the student says, does, or produces will be captured.

Although the spatial layout of Figure 1 suggests a temporal order to the assessment design process, in practice it is more a matter of iterative bootstrapping than one of discrete sequential steps. We begin with a set of learning outcomes we want to include in the student model, but as we become involved in creating tasks that provide a context for eliciting those learning outcomes (and later as we field test the assessment with students), we often develop new insights into their nature and limitations. These insights may modify the student model, the evidence model, the task model, or all three.

APPENDIX B

Mapping Between BioKIDS Assessment Tasks and Design Patterns

Appendix B

Mapping Between BioKIDS Assessment Tasks and Design Patterns

Question	Test Type	Design Pattern and Rationale

A biologist studying birds made the following observations about the birds. She concluded the birds would not compete for food.

	Bird 1	Bird 2	Bird 3
Food	Berries	Berries	Berries
Feeding	Dawn/Dusk	Dawn/Dusk	Dawn/Dusk
Where They Feed	Trees, middle	Trees, lower	Trees, upper

What evidence supports her conclusion?

A. Insects are plentiful.
B. They feed at different times.
C. They feed in different parts of the trees.
D. They lay eggs at different times.

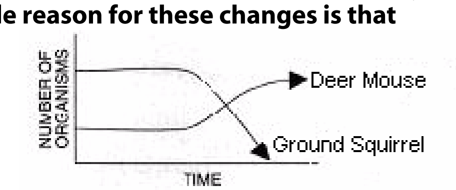
Test Type: Multiple Choice

This question tests the low level of the design pattern **"Formulating scientific explanation from evidence."**

Initially, we were going to create a new DP called "matching evidence to claim." However, we decided that this question gives students the conclusion and some evidence and students are asked to use this information and match the correct evidence to the table.

This question could also be considered **"Interpreting data"** since they have to take the data in the table form and match which statement is correct based on this data.

The graph below shows changes in populations of Ground Squirrels and Deer Mice in a grassy field. A possible reason for these changes is that



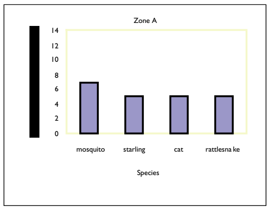
A. All of the plant populations in this habitat decreased.
B. Deer Mice competed more successfully for food than Ground Squirrels did.
C. Ground Squirrels produced more offspring than Deer Mice did.
D. Ground Squirrels consumed the Deer Mice.

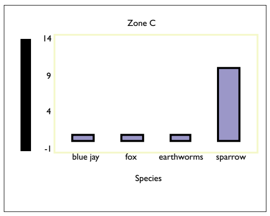
Test Type: Multiple Choice

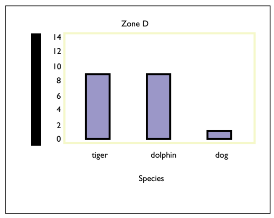
This question tests a medium level of **"Interpreting data."** Students have to use the data and are asked to determine which statement explains the data that the graph shows.

Question	Test Type	Design Pattern and Rationale

Compare the graphs of four zones below. Which zone has the greatest biodiversity?







Multiple Choice

This question tests the design pattern **"Analyzing data relationships."** Here students are asked to compare a set of graphs and determine which one has the highest components of animal abundance and richness and relate this to the concept of biodiversity.

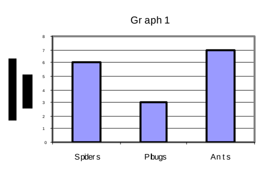
Lisa and Juan observed many animals in different parts of their schoolyard. They recorded their observations in the table below:

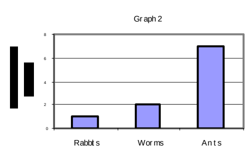
	Zone A	Zone B	Zone C
Abundance of Animals	30	30	10
Richness of Animals	1	7	3

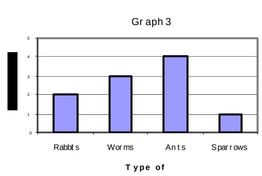
1. **Which zone of the schoolyard has the greatest biodiversity? Explain why you chose this zone.**

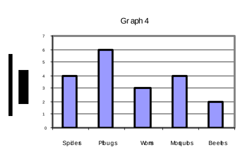
   I think that zone ______ has the greatest biodiversity because …

2. **Circle the graph that best represents Zone C.**









Open Ended

Part 1 of this question addresses the design pattern **"Interpreting data."** Students are asked to look at the table (a representational form, RF) and using the data decide which one has the highest biodiversity and then defend their answer.

Part 2 of this question addresses the design pattern **"Re-expressing data."** This question is a medium level question because students are asked to look at data and choose from a set of four given graphs which one best represents a set of the given data.

Question	Test Type	Design Pattern and Rationale
Biologists measured the biodiversity of animals in a city park in two different years. **# of animals each year / 1995 / 2000** **Starling** 150 / 300 **Mouse** 50 / 0 **Cricket** 35 / 0 **Lady beetle** 100 / 20 **Total** 335 / 320 **1. Using the table above, circle the correct statement about the changes in the park over five years:** A. Species richness decreased. B. Species richness increased. C. Abundance stayed the same. D. Abundance increased. 2. **Some crickets are dark brown or black in color. Based on their color, think of a microhabitat where you might find a cricket.** A microhabitat where I might find a cricket is … **3. One reason that a cricket's color helps them to live in this microhabitat is…**	Open Ended	Part 1 of this question is a multiple choice question that addresses the design pattern **"Interpreting data."** Students are asked to look at the table (a representational form RF) and using the data decide which statement is correct. Part 1 of this question that could also address a low level of the design principle **"Formulating explanations using scientific evidence."** Students are asked to read the table and use this evidence match the evidence to the correct conclusion.
Spreadsheets are used to organize data that scientists collect. **1. Using the spreadsheet at your station, answer the following questions:** **i. In Zone B, what was the animal abundance?** ____________ **ii. In Zone A, what was the animal richness?** ____________ **iii. How many ants were found in the whole schoolyard?** _____________ **2. i. Which zone do you think contains the most microhabitats?** **I think that zone __________ contains the most microhabitats.** **ii. Tell one reason why you made this decision.** **Using the data on the spreadsheet and the graph paper below, create a *bar graph* of *animal abundance* for each part of the schoolyard. Make sure to fill in the title and label the axes.**	Practicum	The first part of this question addresses the design principle **"Interpreting Data."** Students are asked to look at a spreadsheet of data and use it to answer questions that pertain directly to this material. The last part of this question addresses the design pattern **"Re-expressing data."** This question is a high level question because students are asked to look at data and then create their own graph (reinterpretation) of the given data.

APPENDIX C

FOSS Adaptation of "Viewing Real-World Situations from a Scientific Perspective"

Appendix C

FOSS Adaptation of "Viewing Real-World Situations from a Scientific Perspective"

Attribute	Value(s) (from GLOBE)	Comments	*Populations and Ecosystems* Specifics
Name	Viewing real-world situations from a scientific perspective		
Summary	A student encounters a real-world situation that lends itself to being framed from a scientific perspective. Does the student act in a way consistent with having done so?	Viewing a situation from a scientific perspective can be contrasted with, for example, personal political, social, or magical perspectives. This is a design pattern that is clearly appropriate for younger students. It is also appropriate for adults, once they are outside their areas of expertise.	Students study many different aspects of ecosystems. When faced with issues that affect real-world ecosystems, can they base decisions on a scientific perspective?
Rationale	A scientific perspective says that there are principles and structures for understanding real-world phenomena, which are valid in all times and places, and through which we can understand, explain, and predict the world around us. There are systematic ways for proposing explanations, checking them, and communicating the results to others.		Students are introduced to some of the systematic ways in which scientists study ecosystems. As adults, they may need to make important decisions about the status of certain ecosystems, and it is important that they be able to look at arguments made from a scientific perspective rather than merely emotional or personal.
Focal KSAs	Knowledge and understanding of how to view real-world phenomena from a scientific perspective.		Students need to understand several aspects of how ecosystems function. These include: population growth and limiting factors, how populations interact (e.g. food webs with producers, consumers, and decomposers), and energy transfer through a system.
Additional KSAs	Ability to structure setting so that knowledge of particular scientific content or models is required or is minimized.		Additionally students need to understand how scientists study ecosystems, and how to evaluate and predict the impact of stresses and pressures that might be introduced to the populations within them.

Attribute	Value(s) (from GLOBE)	Comments	*Populations and Ecosystems* Specifics
Potential observations	Posing a scientifically answerable question.	Question should be relevant, realistic, and potentially addressable in light of the situation.	Determine the important data needed to study a particular ecosystem and argue the issues of a given controversy from a scientific perspective.
	Explaining how to get started investigating the situation.		
	Identifying reasonable (read, scientific) next steps		
	Critiquing responses offered by other students, either predetermined or as they arise naturally.		
Potential work products	Verbal (oral or written) question, explanation of how to get started investigating the problem, etc.		Written responses, short answer, and multiple choice responses from items that are developed in the context of simple ecosystem scenarios.
	Diagram of the situation	Looking for relevant features, especially if there is particular substance or knowledge representations the student should be employing.	Ecoscenario report and presentation preparation. Teachers would make observations of the data students are gathering, the questions they are asking, and the general direction they are taking in creating their arguments. Students keep a log that identifies information they think they need that will be important to their taking a stand on one of the issues.
	Identification, from given possibilities, of those which reflect a scientific perspective.		• Ecoscenario Reports (written product)—gather data about a complete real-world system; report issues; take a stance. • Ecoscenario Class presentations (oral product)—present facts, report issues accurately, answer questions regarding issues and stance taken; not reverting to emotional or naïve conceptions.
Potential rubrics			See chart of observations that would help teachers identify a scientific view from a more naïve view. Also see generic FOSS scoring guide.

Attribute	**Value(s) (from GLOBE)**	**Comments**	***Populations and Ecosystems* Specifics**
Characteristic features	Motivating question or problem		Interesting situations explained in all ecosystems presented to students.
	Background information provided so student can provide a meaningful question and answer.		• Ecosenarios: just enough information provided to stimulate student interest in the ecosystem they have been assigned. Students required to gather additional information on their own.
Variable features	Amount of prompting or cueing	Less cueing gives better evidence about whether student is internally inclined to see situations from a scientific perspective; more cueing gives better evidence about whether student is able to proceed knowing that it is appropriate to think from a scientific perspective.	• Students given specs for report. • Some substantive information provided, but student must gather additional data and decide which are the important pieces with regards to arguing issues.
	Degree of substantive knowledge involved	"Content lean" vs. "content rich" in Baxter & Glaser's terms. Light content focuses evidence on inquiry perspective. Heavier content puts stress on knowledge of that content, and calls for seeing situation in terms of models/principles. This confounds the inquiry and content KSAs, but makes it possible to get evidence about whether the student sees situations scientifically with respect to given content. [note: connects with diSessa research-see links entry below]	Try to find a balance between inquiry perspective and substantive knowledge so that feedback can be provided for both.
	Amount of substantive knowledge provided	When substantive knowledge, such as models, formulas, knowledge representation, tools, or terminology is required for an appropriate response, to what degree are they provided? Providing them reduces the load on the substantive KSAs. Not providing them means the response requires, conjunctively, the substantive KSA and the focal inquiry KSA.	This will vary depending on task.

Attribute	Value(s) (from GLOBE)	Comments	*Populations and Ecosystems* Specifics
I am a kind of…	Scientific reasoning (or model-based reasoning)	This paradigm is part of a more encompassing pattern of assessing students' articulating between specific real-world situations and representations of those situations in terms of scientific concepts, models, and principles.	
Kinds of me	Planning solution strategies		
I am part of…	Conducting investigations	Viewing real-world problem and situation can be a first phase of an investigation.	Designing and conducting investigations.
Parts of me			
Educational standards			
Unifying concepts	*Evidence, models, and explanations*		*Evidence, models, and explanations*
Science as inquiry standards	*Ability to ask scientific questions*		*• Different kinds of questions require different kinds of investigation.* *• Develop descriptions, explanations, predictions, and models using evidence.*
Templates (task/ evidence shells)	GLOBE generic template	Posing a question, one of the kinds of observations that bears on the focal KSA, is the first step in a GLOBE investigation.	
Exemplar tasks	[various GLOBE tasks]		
Online resources	GLOBE home page		
References	diSessa, A. (1982). Unlearning Aristotelian physics: A study of knowledge-based learning. *Cognitive Science*, *5*, 37-75.	Harvard physics students solve complicated mechanics problems in the classroom, but fall back on naïve explanations when asked what will happen next with kids on playground equipment—even though exactly the same models apply.	

Appendix D

PADI Use-Case

Appendix D

PADI Use-Case

Edwina works with Edwin, building the library of design patterns. She logs into the PADI design system [Screen 1] and is presented with an initial screen that has some instructions along with a list of the existing top-level design patterns in the library [Screen 2].

Screen 1: Login

Welcome to the PADI Assessment Design System!
Username: Edw ina
Password: ******
Login

Screen 2: Design Pattern List

Welcome back, Edwina.
View, edit, or create a design pattern below.
Design Pattern Name
Summary
Action

Conducting investigations
(placeholder)
delete - edit

Planning solution strategies
In this design pattern, students are presented with an open-ended problem to investigate and must generate a plan for solving the problem. Do students generate coherent plans that are guided by an adequate representation of the problem situation and possible procedures and outcomes?
delete - edit

Scientific Reasoning
This design pattern concerns a scientific problem to solve or investigate. Do they effectively plan a solution strategy, carry out that strategy, monitor their own performance, and provide coherent explanations?
delete - edit

Viewing real-world situations from a scientific perspective
A student encounters a real-world situation that lends itself to being framed from a scientific perspective. Does the student act in a way consistent with having done so?
delete - edit

View Complete List | Create New Paradigm

Edwina clicks on the **"Planning solution strategies"** link to get a feeling for what a design pattern looks like [Screen 3]. Since Edwina has rights to edit design patterns, an "Edit" button appears at the top of the view page, in case she wants to edit what she is viewing. Note that reviewers can add their feedback, which appears at the bottom of screen 3.

Screen 3: Viewing a Design Pattern

1.1.1.1.1.1 **Design Patterns** | **view: "Planning systematic solution strategies"**
Edit this design pattern

Attribute Name
Value
Comments

Name
Planning systematic solution strategies

Summary
In this design pattern, students are presented with an open-ended problem to investigate and must generate a plan for solving the problem. Do students generate coherent plans that are guided by an adequate representation of the problem situation and possible procedures and outcomes?

Rationale
Cognitive studies of expertise have shown that planning an approach before employing a solution strategy is one of the characteristics that differentiate more competent from less competent problem-solvers in a content domain.
Competent performers qualitatively assess the nature of a problem and construct a mental model or internal representation prior to initiating a solution strategy. This representation is used to anticipate alternative outcomes to various actions

Focal KSA
Ability to plan solution strategies (procedures and possible outcomes).

[Wanted to narrow things down to looking at only two variables. If a given problem has more than two variables we would say that this design pattern occurs multiple times in the same problem]

Additional KSAs
Content knowledge

Inquiry skills
e.g., Strategies for control of variables

Verbal abilities, if response mode is verbal

Potential observations

- Completeness of plan
- Integrity of procedures
- Comparison of students' plan to expert plan

Potential rubrics

Potential work products
Written plan
"How are you going to go about solving this problem?"

Students' oral presentations on how they are going to approach problem

Selecting best solution plan from given possibilities
Rough outline of plan developed by students in a small group

Observations of students as they work in groups brainstorming how to approach problem

Characteristic features
Motivating problem to solve/investigate
Open ended; little/no cueing

Variable features
Complexity of inquiry activity
Some investigations may be quite complex, with multiple variables to control

Focus on process vs. content

Process: emphasis on how students approach the problem
Content: how students bring to bear their content knowledge in coming up with a plan

Domain-specific vs. general knowledge
Specific: knowledge specific to domain (e.g., conservation of energy)
General: principles that cut across scientific domains (e.g., control of variables)

I am a kind of
Viewing real world situations from a scientific perspective

These are kinds of me

I am part of...
Scientific reasoning

Links to NSES Standards
Unifying Concepts
Evidence, models, and explanation

Science as Inquiry standards
Abilities necessary to do scientific inquiry
- *Design and conduct a scientific investigation.*
- *Use appropriate tools and techniques to gather, analyze, and interpret data.*

- *Communicate scientific procedures and explanations.*

Understandings about scientific inquiry
- *Kinds of questions and investigations.*
- *Methods of different scientific domains.*
- *Use of technology*.

Links to Template (task/evidence shells)

Links to exemplar tasks
Mystery Powders (Baxter, Glaser, & Elder)
In this performance assessment students are asked to investigate which of three white powders (salt, baking soda, and cornstarch)—individually or in combination—are contained in each of six bags

Online resources

References
Baxter, G. P., Elder, A. D., & Glaser, R. (1996). Knowledge-based cognition and performance assessment in the science classroom. *Educational Psychologist*, *31*(2), 133-140.

Feedback/Comments from reviewers
Contributor
Comment
Date

Joe Smith
This design pattern is nifty.
3 Jan. 2002

Mary Jones
Educational Goals should not include xyz.
22 Feb. 2002

Add Feedback

Edwina then clicks to edit a design pattern from the previous listing [screen 2] or **"Edit this design pattern"** link in screen 3 and is presented with a form [screen 4] by which she can update the contents of a design pattern. (Note: Some of the screen shots include the term "paradigm," which at an early stage of the project we were using instead of the less controversial term "Design pattern," which we eventually settled on.)

Screen 4: Editing a Design Pattern

Design Patterns | Edit: " Planning systematic solution strategies "

Attribute

Value

Comments

Name

Planning systematic solution strategies

This title is tentative.

Summary

In this paradigm, students are presented w ith an ope

summary taken from source xy

Educational standards which are associated

Standard

Association

Change Type of association

Delete association

Unifying Concepts

direct

edit

delete

Science as Inquiry standard

peripheral

edit

delete

Add Associated Standard

Task-evidence templates
Template
Association
Change Type of association
Delete association

abc template
direct
edit
delete

xyz template
peripheral
edit
delete

Add Associated Template

Relations to other design patterns
Design Pattern
Relation
Change Type of Relation
Delete Relation

Viewing real world situations from a scientific perspective
I am a kind of
edit
delete

Scientific reasoning
I am part of...
edit
delete

Add Related Paradigm

should w e include the abc para

Update Paradigm

Sponsor
The National Science Foundation, Grant REC-0129331

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

Subgrantees
University of Maryland
University of California, Berkeley. *Berkeley Evaluation & Assessment Research (BEAR) Center* and *The Full Option Science System (FOSS)*
University of Michigan. *BioKIDs*