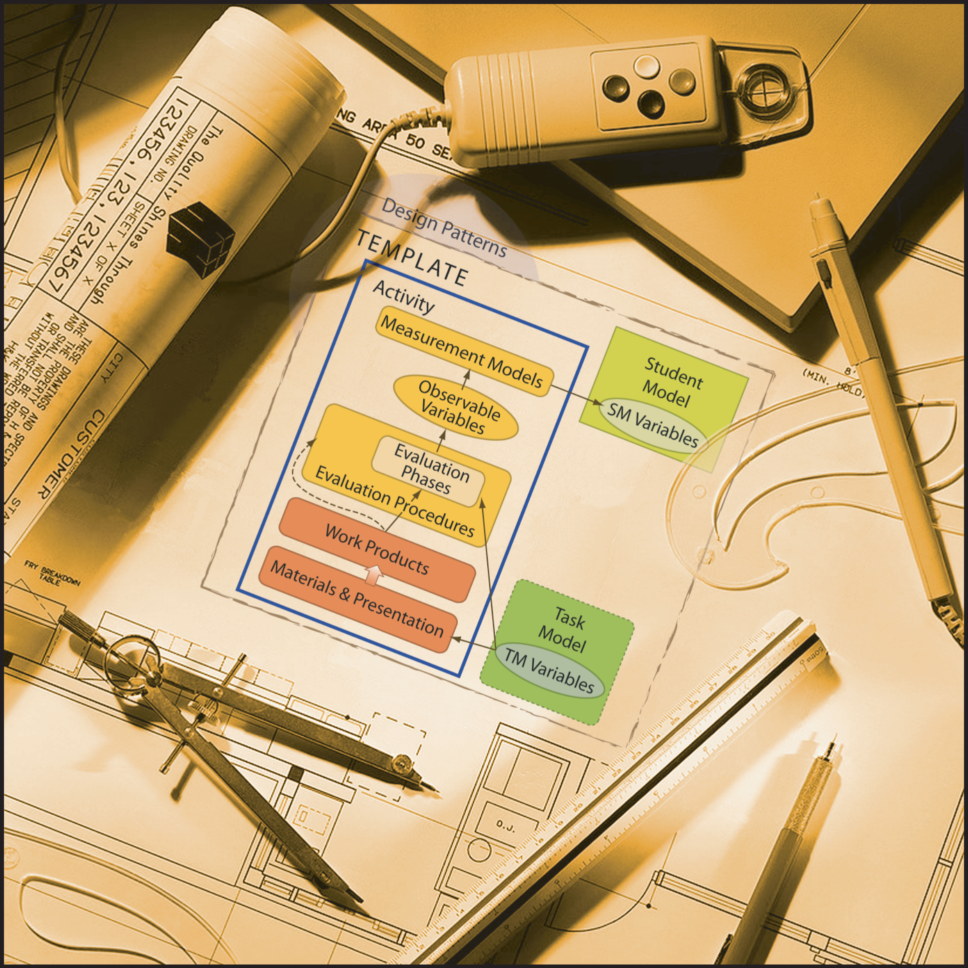


PADI Technical Report 3 | February 2005

An Introduction to PADI Task Templates

PADI | Principled Assessment Designs for Inquiry

Michelle M. Riconscente, University of Maryland
Robert J. Mislevy, University of Maryland
Lawrence Hamel, CodeGuild, Inc.
PADI Research Group

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

TECHNICAL REPORT 3

An Introduction to PADI Task Templates

Prepared by:

Michelle M. Riconscente, University of Maryland

Robert J. Mislevy, University of Maryland

Lawrence Hamel, CodeGuild, Inc.

PADI Research Group

Acknowledgment

PADI is supported by the Interagency Educational Research Initiative (IERI) under grants REC-0089122 (PADI Planning Grant) and REC-0129331 (PADI Implementation Grant).

1.0 Introduction **1**

2.0 The PADI Project **3**
2.1 Principled Assessment Designs for Inquiry 3
2.2 An Analogy for the Layers and Objects in the PADI Framework 5

3.0 Overview of PADI Design Patterns **8**
3.1 The Origins of Design Patterns 8
3.2 The Building Analogy, Continued 8
3.3 Details of PADI Design Patterns 9

4.0 PADI Templates **12**
4.1 The Role of Templates 12
4.2 The Structure of Templates 12
4.2.1 ECD and the PADI Template Object System 12
4.2.2 Defining Task Features 14
4.2.3 Gathering Evidence 15
4.2.4 Making Inferences about Student Proficiency 15
4.2.5 Activities and Attributes 16
4.2.6 Putting the Pieces Together 16
4.3 The Extensibility of Templates 17

5.0 A Sample Template: EDMS 738 **19**

6.0 Sample Task specs from the EDMS 738 Template **28**

7.0 Conclusion: Some Observations on Structures and Knowledge Representations in Assessment Design **34**

References **36**

Appendix A **37**
An Evidence-Centered Framework for Assessment Design and Delivery **37**
1.0 An Overview of the ECD Layers 37
2.0 The Conceptual Assessment Framework 38
3.0 Delivery System Architecture 40
References 42

Appendix B **43**
Definitions of the PADI Object Model **43**
Reference 52

Figures

Figure 1.	Simplified View of the Layers in the ECD Framework	1
Figure 2.	Basic (Generic) Template Structure	13
Figure 3.	Expanded (Generic) Template Structure	13
Figure 4.	Houses Generated From the Same Template Structure	17
Figure 5.	“Model Elaboration” Design Pattern Object in PADI Design System	20
Figure 6.	Basic “EDMS 738 Assignments” Template Object in the PADI Design System	22
Figure 7.	Graphic Representation of Basic “EDMS 738 Assignments” Template	22
Figure 8.	Expanded “EDMS 738 Assignments” Template	24
Figure 9.	Details of “Final Version of Essay” Activity	25
Figure 10.	Observable Variable “Project Grade” in the PADI Design System	26
Figure 11.	“EDMS 738” Task Spec I in PADI Design System	30
Figure 12.	Setting “Topic Area” TMV Value in the PADI Design System	31
Figure 13.	Displaying TMV Settings for Task Spec II	32
Figure 14.	“EDMS 738 Final Essay” Task Spec Object in PADI Design System	33

Glossary of Acronyms

CAF	*Conceptual Assessment Framework*
ECD	Evidence-Centered Assessment Design
IERI	Interagency Educational Research Initiative
IMS/QTI	Instructional Management Systems/Question-Test Interoperability
KSAs	Knowledge, Skills, and Abilities
MM	Measurement Model
M&P	Materials and Presentation
MRCMLM	Multidimensional Random Coefficients Multinomial Logit Model
OSEA	On the Structure of Educational Assessments (Mislevy, Steinberg, & Almond, 2003)
OV	Observable Variable
PADI	Principled Assessment Designs for Inquiry
SMV	Student Model Variable
TMV	Task Model Variable
UML	Unified Modeling Language
WP	Work Product
XML	Extensible Markup Language

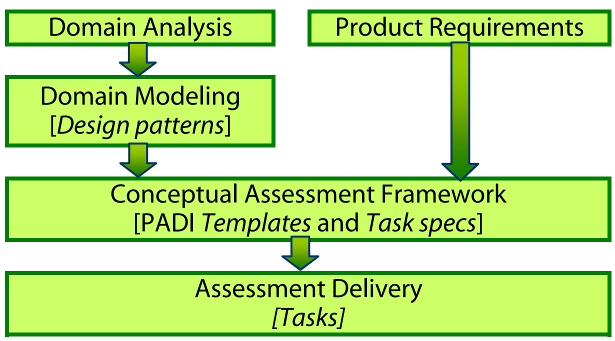
ABSTRACT

Principled Assessment Designs for Inquiry (PADI) is a project supported by the National Science Foundation to improve the assessment of inquiry in science learning. PADI is developing a design framework for assessment tasks, with a particular focus on tasks that stress concepts and problem solving, building and using models, or cycles of investigation. Previously PADI developed structures called *design patterns* for laying out in a narrative form the elements for assessment arguments. This report introduces PADI *task templates* and *task specifications*. *Task templates* provide an object model framework for the more nuts-and-bolts level of design, and *task specifications* are blueprints for individual tasks expressed in this framework. A *task template* articulates a conceptual assessment argument in terms of the elements and processes of operational assessment tasks. The focus in this report is on the rationale and structure of *templates*, illustrated with a relatively simple example. The issues of design processes, authoring, implementation, and operation are discussed elsewhere. This report is accompanied by two appendices that provide more detailed treatments of both evidence-centered design and the UML object model, which have evolved through the work of PADI.

1.0 Introduction

An educational assessment is a special kind of evidentiary argument. It is a way of gathering information in the form of a handful of things students say, do, or make under particular circumstances, to make inferences about what they know, can do, or have accomplished as more broadly conceived. In a paper titled "On the Structure of Educational Assessments" (abbreviated below as OSEA), Mislevy, Steinberg, and Almond (2003) proposed a general model for the layers of activities in the design and administration of assessments: layers for marshalling information about the domain and the purposes and constraints of the assessment, for explicating the assessment argument, for structuring the elements of the assessment, and for the operation of the implemented assessment. Figure 1 is a simplified graphic of these layers, referred to in OSEA as *domain analysis*, *domain modeling*, the *conceptual assessment framework* (CAF), and *assessment delivery*. The PADI objects representative of each layer are indicated in brackets, and proceed from *design patterns* → *task templates* → *task specifications* → *tasks*. Appendix A provides a fuller summary of the structures in the evidence-centered design (ECD) assessment framework.

Figure 1. Simplified View of the Layers in the ECD Framework



OSEA discussed the roles of these layers and described major elements and key relationships within and between layers, and Almond et al. (2002) elaborated the delivery system architecture. The level of discussion remained fairly abstract, however, without the details or structures that would be needed to design and implement assessments within the framework. The framework could be instantiated in many ways, all consistent with the overview provided in OSEA. The NSF-supported project Principled Assessment Designs for Inquiry (PADI) is developing an object model with supporting software that elaborates portions of the ECD model at the level of *domain modeling* and the CAF, with a focus on developing tasks to assess inquiry in science, including problem solving, building and using models, or involving cycles of investigation. A previous report (Mislevy et al., 2003) discussed PADI *design patterns*, a conceptual tool at the *domain modeling* layer. This presentation focuses on PADI objects called *task templates* and *task specifications*, which reside at the CAF layer.

To set the context for the discussion of PADI *templates*, Section 2.0 provides an overview of the PADI project. Section 3.0 briefly reviews *design patterns*, in order to distinguish their structure and use from those of *templates* and to introduce the example we will use to illustrate *templates*. That example is based on essay assignments in EDMS 738, a graduate-level course in the foundations of assessment design. Section 4.0 presents a generic description of *templates* and their constituent elements, along with a UML object representation. A distinction is made between two closely related PADI objects: *templates*, which provide a design framework for families of related tasks, and *task specs* (short for "*task specifications*"), which fill in the details of a *template* to provide a blueprint for implementing a specific task. Section 5.0 works through the details of the EDMS 738 *template* and Section 6.0 gives two examples of a *task spec* and implemented task instantiated from the *template*. Section 7.0 closes with some observations about the roles of structures and knowledge representations in assessment design.

2.0 The PADI Project

2.1 Principled Assessment Designs for Inquiry

The Principled Assessment Designs for Inquiry (PADI) project is supported by the Interagency Educational Research Initiative (IERI), under the auspices of the U.S. Department of Education, the National Science Foundation, and the National Institutes of Health. The goal of IERI, broadly speaking, is to promote educationally useful research that supports the learning of increasingly complex science content, with a particular emphasis on scaling up innovations that have proven successful on a smaller scale. 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 assessments 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 the 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).

PADI 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 the ways students come to know them can be taken into account.

The IERI goals of replicability and scalability require this effort up front, working 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, PADI is developing a focused implementation of the evidence-centered assessment design (ECD) framework summarized in OSEA. The ECD framework explicates the interrelationships among substantive arguments, assessment designs, and operational processes. In particular, PADI is developing object models, supporting software, and worked-out examples of structures in the *domain modeling* and CAF layers, as well as providing an optional scoring engine that users may wish to include in their assessment application. (This scoring engine is based on the work of Wilson and his colleagues with a general psychometric model called the Multidimensional Random Coefficients Multinomial Logit Model, or MRCMLM; e.g., Adams, Wilson, & Wang, 1997. The MRCMLM includes, as special cases, Rasch models for dichotomous and partial-credit responses, the linear logistic test model, Stegelmann's multivariate Rasch model, and Andersen's multidimensional Rasch model for nominal categories.)

Figure 1 showed major layers in a framework for the design and delivery of an 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—for example, how it is learned, how it is used, situations in which it is used, and how you know it when you see it. 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.

Design patterns lie in the layer in the assessment system called *domain modeling*. While *domain analysis* is identifying the knowledge and skills to be assessed, *domain modeling* organizes this information into the form of assessment arguments. PADI *design patterns* are an example of a *domain modeling* tool. The domains of interest in PADI combine science content and inquiry processes. A *design pattern* specifies the elements of an assessment argument, bridging content expertise with the measurement expertise needed to create an operational assessment.

The technical layers of the assessment design process are where the details of psychometric models, scoring rubrics or algorithms, presentation of materials, interactivity requirements, and so on, are specified. In OSEA, the major design structure at this layer is called the *conceptual assessment framework*, or CAF. The technical work is shaped 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 in service of the intended use of the assessment tasks. The particular structures in the PADI framework in which this technical level of design is expressed are called *templates* and *task specs*. They are used in turn to implement actual tasks in accordance with the requirements of the operational assessment system. The *template* structures are defined at a general level that allows them to be used to describe, in a common language and common framework, the elements of assessments of very different kinds—from classroom projects to traditional standardized tests to intelligent tutoring systems—even though the elements of the tasks, when implemented, can take radically different forms.

When working through the many structures and details of the PADI framework, it is useful to keep in mind a quotation from Messick (also see Mislevy, Almond, & Lukas, 2003):

> A construct-centered approach would begin by asking what complex of knowledge, skills, or other attributes should be assessed, presumably because they are tied to explicit or implicit objectives of instruction or are otherwise valued by society. Next, what behaviors or performances should reveal those constructs, and what tasks or situations should elicit those behaviors? Thus, the nature of the construct guides the selection or construction of relevant tasks as well as the rational development of construct-based scoring criteria and rubrics. (Messick, 1994, p. 16)

Messick's questions transect layers of the design process, resulting in design objects that move from broad knowledge about the target domain and purpose, through increasing detail and structure, to the nuts and bolts of an operational assessment. The overview posed by these guiding questions helps us understand how the issues addressed at each layer relate to those at other layers, all toward the goal of a coherent design that serves the assessment's purpose.

To summarize the key points for our discussion of PADI objects, *design patterns* are structures in the *domain modeling* layer of assessment design, where the substantive argument and essential elements of schemas for assessment tasks are laid out. *Templates* and *task specs* are structures in the CAF layer, at which operational elements and processes of tasks are specified (as opposed to *implemented*; the following section introduces an analogy with architectural design that helps clarify this distinction). Implemented tasks are instantiations of the objects described in general in *templates* and described in particular in *task specs*; these are the parameters, stimulus materials, instructions, and so on, that actually function in an operational assessment.

2.2 An Analogy for the Layers and Objects in the PADI Framework

An analogy to the process of constructing a building is helpful for both understanding the importance of these layers and extending the discussion to the objects in the PADI design framework. Using an apartment building as an example, consider the various phases that take place from the initial conceptualization to the final product. In the first stage (analogous to the *domain analysis* layer), the would-be apartment building exists only as an idea, far from complete but already reflecting information and decisions such as the overall purpose of the structure, the demands of future residents, geographic location, and budget constraints. The plan for the future apartment building would then start to be fleshed out with more detail in conversations among people including the developer, the architect, and the general contractor, though not yet with sufficient precision to begin construction. At this level, akin to the *domain modeling* layer, a number of decisions are made that focus the design and provide a rough sketch of the final product—possibly several variations, each an initial exploration of plausible directions that might be followed up. For example, the style of the building (luxury high-rise or student housing) may be fixed. As alternatives are debated and more design decisions are finalized, a set of blueprints would eventually be generated in which all the particulars are fully determined, such as number of units, floor plans, HVAC specifications (for heating, ventilation, and air-conditioning), locations and sizes of windows, elevators, stairs, and material requirements for each. This design work corresponds to the CAF layer in assessment design. The final stage of development is the actual construction of the building, which corresponds to creating the operational elements of the assessment. The functioning of the actual building elements—the elevators, the HVAC, the revolving doors in the lobby—correspond to the delivery, analysis, and reporting in an operational assessment, in patterns organized around the delivery system architecture.

Consider as well the range of actors involved in the overall construction process. The owner of the land may have in mind the overall goal of the enterprise, as well as constraints such as budget and timeline, whereas others (e.g., architects, construction workers, electricians, plumbers) have the expertise to turn the idea into reality, identifying key decisions to be made and options to be weighed. Many of these players never meet each other and possibly do not even know of the other's existence or role. However, the overall scope of the building project, as well as the needs of the individuals involved, can still be met if appropriate processes and knowledge-representation forms are in place to support the endeavor. Knowledge representations such as blueprints help organize and coordinate the work of these actors to design and construct a strong and safe building that serves the purposes for which it is intended. Knowledge representations such as PADI *design patterns* and *templates* play an analogous role in assessment design.

Most of the assessments that are created and administered every day throughout the world, of course, do not go through design processes as detailed, explicit, or structured as the ones described in this report. Nor do they need to, as a practical matter; it would be excessive to go through all of these activities to build a doghouse, although one could.

At this point, then, a few words are in order about the rationale for proposing what appears at first to be a rather complicated structure for what seems to be a rather simple job, namely, writing assessment tasks. We would argue that the principles and relationships that underlie familiar assessments are tacit in processes that have evolved over a century for assessing students. The processes and the artifacts may seem simple because they are familiar, but the principles are not. This fact comes to the fore whenever we try to define new kinds of assessments (e.g., problem solving in simulation environments) or work with different kinds of data (e.g., traces of actions in open-ended cycles of inquiry). It is productive in such cases to drop back to first principles and identify at a higher level of generality the elements and relationships that seem to be common among successful assessments of all types. Leverage is gained when unifying concepts and relationships appear.

One methodology for laying out systems at a fundamental level is called *object modeling*[1] in software engineering and *enterprise modeling* in the analysis of business systems (Booch, Rumbaugh, & Jacobson, 1999). This is the approach suggested with respect to the assessment design elements in the ECD models and implemented by the objects in the PADI framework. Because an object modeling framework allows individuals involved in different areas of the design process to do their jobs well without necessarily knowing how the other components function, a wide range of expertise can be brought to bear on the outcome. The investment in developing principled models affords efficient and effective communication across components.

[1] A brief definition from Yourdon and Constantine (1979): OOA [object-oriented analysis] views the world as objects with data structures and behaviors and events that trigger operations, or object behavior changes, that change the state of objects. The idea that a system can be viewed as a population of interacting objects [i.e., the object model], each of which is an atomic bundle of data and functionality, is the foundation of object technology and provides an attractive alternative for the development of complex systems.

3.0 Overview of PADI Design Patterns

3.1 The Origins of Design Patterns

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. Although *design patterns* can reference knowledge and skills addressed in content and inquiry standards, they are different in that they are organized specifically to ground assessment arguments. Less detailed than the technical specifications for assessment tasks found in *templates* and *task specs*, *design patterns* are intended to communicate with educators and assessment designers in a nontechnical 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).

Similar tools or schemas have been generated in other disciplines that provide useful analogies for explaining the role of *design patterns* in assessment design. Architect Christopher Alexander (Alexander et al., 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." Computer scientists picked up on Alexander's work when they noticed patterns recurring in their designs (Gamma et al., 1994). These patterns 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, essential elements include the Problem and the Solution. The Problem indicates when to apply the pattern, describing in generic terms both the problem and the context. It may additionally include a list of conditions that must be met before it makes sense to apply the pattern. The Solution specifies the elements that make up the design, relationships, responsibilities, and collaborations. A pattern is not a concrete design or implementation, but rather an abstract description of how a general arrangement of elements, applicable in many situations, solves a problem.

3.2 The Building Analogy, Continued

In PADI, *design patterns* lay out the elements of a chain of reasoning, from evidence to inference. Assessment *design patterns* reflect the PADI tenet that complex assessments should 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. These details appear at a later stage of the process in the form of *templates* and *task specs*.

If we continue in the metaphor of the building, we could imagine a “concept plan” for the apartment building. Such a plan would have a title, a brief summary of the purpose of the building, and some descriptive information, such as number of units and location. It might specify whether it is to be a luxury high-rise or university housing. We could also distinguish between a general framework for buildings and a specific plan for a building by thinking of “slots” of information. The general framework might have slots for “kind of building,” “zone,” “style,” and “climate.” In a specific plan, these slots would be filled in with appropriate values (e.g., kind of building = high-rise apartment; zone = residential; style = art deco; climate = tropical).

We can think of PADI’s *design patterns* in an analogous way. As a framework, *design patterns* consist of a set of attributes (like “slots”) that guide planning for the key elements of the design models in the ECD *conceptual assessment framework* (i.e., student, evidence, and task models; see Appendix A for details). For example, among *design pattern* attributes are characteristic task features and variable task features. The values that are set for these attributes begin to shape the context in which the work will be produced, while still leaving many of the more specific and technical decisions to be made later in the design process. In the PADI design system, *design pattern objects* are created by filling in the attributes of the *design pattern* structure.

3.3 Details of PADI Design Patterns

Table 1 shows the primary elements of the *design pattern* on which our running EDMS 738 example is based. Later, we will contrast the level of generality of the elements of this *design pattern* with the greater specificity of the elements of the *templates* discussed in Sections 4.0 and 5.0. This example suffices to illustrate the nature and the elements of *design patterns*, but the reader wanting more details and examples is referred to PADI Technical Report 1 on *design patterns* (Mislevy et al., 2003). A summary of key ideas follows.

In the PADI system, a *design pattern* helps the assessment designer structure a coherent assessment argument by making explicit three essential elements. These three elements presage the more technical components of the student, evidence, and task models from the CAF:

1. The knowledge, skills, and abilities (which we abbreviate as KSAs for now, without making any commitment to their nature) that express the aspects of students’ capabilities with respect to inquiry that are the target of inference in the assessment.
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 revolve around 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. With *design patterns*, a key step is made that links knowledge about what is important in the domain of interest to the essential components of assessment. *Templates* then draw on the narrative form of *design patterns* to make decisions about what actual assessment tasks will look like and how they will be quantified to result in inferences about student proficiency.

Table 1. Design Pattern Attributes

Attribute	Description
Rationale	Explains why this item is an important aspect of scientific inquiry and explicates the chain of reasoning connecting the inference of interest about student proficiency to potential observations and work products.
Focal knowledge, skills, and abilities	The primary knowledge/skills/abilities targeted by this design pattern.
Additional knowledge, skills, and abilities	Other knowledge/skills/abilities that may be required by this design pattern.
Potential observations	Some possible things one could see students doing that would yield evidence about the knowledge/skills/abilities.
Potential work products	Modes, like a written product or a spoken answer, in which students might produce evidence about knowledge/skills/abilities.
Potential rubrics	Some evaluation techniques that may apply.
Characteristic features	Aspects of assessment situations that are likely to evoke the desired evidence.
Variable features	Aspects of assessment situations that can be varied in order to shift difficulty or focus.
I am a kind of	Associations with other objects ("my parents") that are more abstract or more general than this object.
These are kinds of me	Associations with other objects ("my children") that are more concrete or more specialized than this object.
These are parts of me	Associations with other objects that contain or subsume this one. For example, a windshield is a part of an automobile.
Educational standards	Associations with (potentially shared) Educational standard objects.
Templates	Associations with (potentially shared) template objects.
Exemplar tasks	Associations with (potentially shared) task exemplar objects.
Online resources	Relevant items that can be found online (URLs).
References	Notes about relevant items, such as academic articles.

KSAs (knowledge, skills, and abilities) are the terms in which we want to talk about students to determine evaluations, make decisions, or plan instruction. In the context of PADI, 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. Because KSAs describe student proficiency, they are written accordingly as nouns (e.g., ability to evaluate scientific data, understanding of principles of chemical processes, ability to use grammar appropriately).

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 see directly 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 include seeing students identify outliers or inconsistencies in the data, explain strategies they use for error checking, propose explanations for anomalies, or reexpress data in a different representational form to reveal anomalies.

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.

An assessment task could correspond to a single *design pattern* or a sequence or assemblage of *design patterns*. For instance, the *design pattern* about evaluating the quality of scientific data could be linked with other *design patterns* that require students to design their own investigation and collect their own data. Having students assess the quality of the data they collected could be a later stage of the task.

4.0 PADI Templates

4.1 The Role of Templates

As with constructing an apartment building, having a general plan in place is a necessary and important step, but there is still a long way to go before construction can begin. Substantially more detail is required, both within and across the systems (e.g., structural, electrical, plumbing) that will make up the finished product. The conceptual plan, or *design pattern*, although constraining the outcome in some aspects, leaves the technical details for later decision. The development of PADI *task templates* is meant to support this step in the design process. *Templates* can be conceived of as “pre-blueprints,” which can be used to generate multiple specific blueprints that, while differing in some aspects, share a common framework. At the *template* level, we begin mapping out details, such as where the walls will go, how many windows a particular wall might have, and where doors will be placed to connect rooms to each other. As with a *design pattern*, some attributes of the *template* will be fixed, and others will be open, or constrained to a set of options. Some components will support others (floor, walls, supporting columns), whereas others will require certain conditions (a window or door is placed in a wall). Once set in place, supporting walls or the shape of the foundation cannot be changed, whereas windows and walls could be combined in different configurations.

As noted in Section 3.0, PADI *templates* provide a more specific object model for the primary design objects at the level of the CAF. Although they are more detailed than *design patterns* and have a more deeply hierarchical structure, it is helpful to keep in mind that *templates* share the same underlying approach, as sketched in the Messick quotation cited earlier. The *template* level identifies the specific objects and the relationships among them that need to be in place to make the underlying student, evidence, and task models of the CAF functional.

4.2 The Structure of Templates

4.2.1 ECD and the PADI Template Object System

The diagram in Figure 2 shows how the basic structure of *templates* corresponds to the three basic ECD design objects—Student, Evidence, and Task models—denoted by the colors blue, yellow, and pink, respectively. Each object shown can be thought of as a building block in an assessment “construction kit” of sorts. Making the connection to the apartment building analogy, the objects in the kit might include a “window” or a “wall” or an “elevator.” Each object has its own set of attributes (e.g., a window has dimensions, number of panes, type of insulation, screens, ways of opening). In addition, each has particular modes of connecting to the overall object, which is the apartment building. Windows must be placed in walls, which can hold multiple windows. Further, the blueprint for an apartment building will contain many copies of the window and wall objects. The same is true of the PADI *template* object system. Each object has its own internal consistency and ways of connecting to some of the other objects. A specific *template* may also include multiple copies of a certain object, each with its own unique attributes, as suggested in the more complex generic *template* illustrated in Figure 3.

Figure 2. Basic (Generic) Template Structure[2]

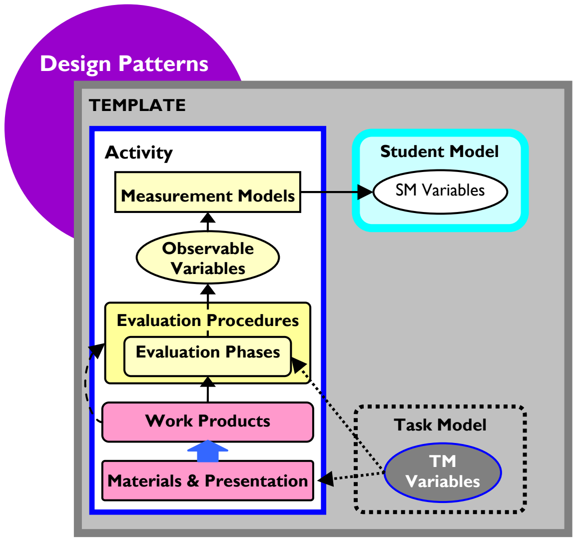
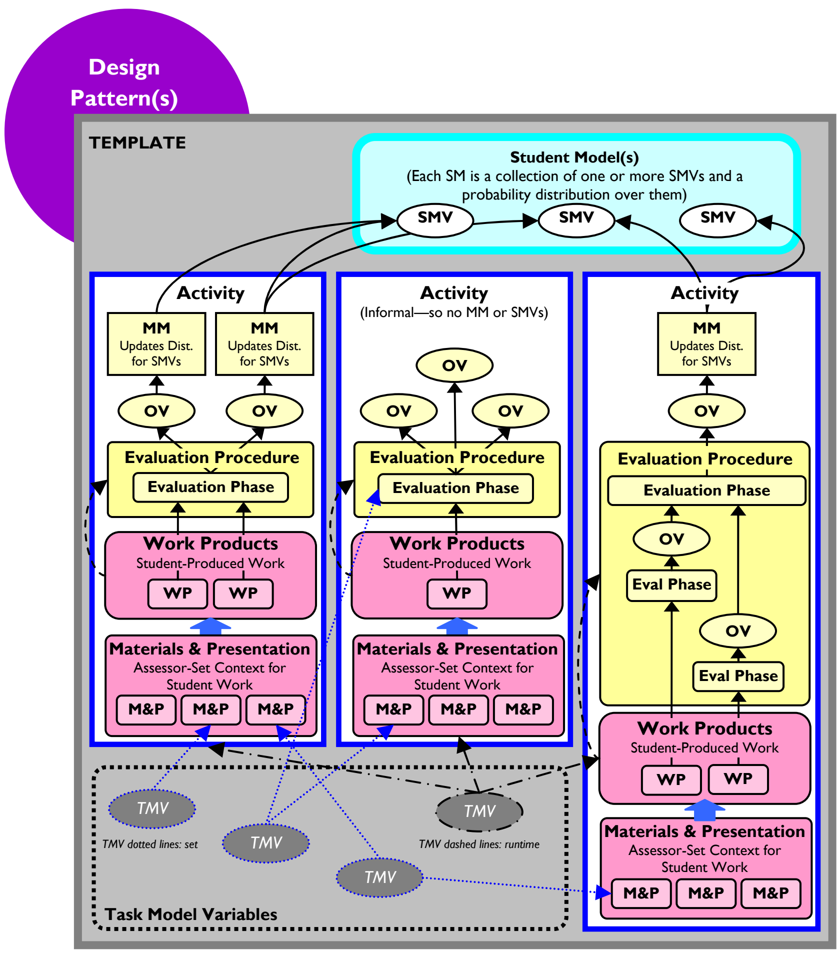


Figure 3. Expanded (Generic) Template Structure[2]



[2] The ECD Student Model is represented by blue objects, the ECD Evidence Model by yellow objects, and the ECD Task Model by pink objects.

Since *templates* are close to the level of implementation, portions of their structure correspond closely to the ECD task model. Recall that *design patterns* suggest characteristic features of tasks that students will produce in order to provide evidence about some targeted KSA, such as an aspect of capabilities for inquiry. In *templates*, these features are described in more detail, in terms of a set of design objects. As shown in Figure 2, Materials & Presentation and Work Products are the primary objects comprising the Task Model. Each Materials and Presentation object denotes some feature of the environment in which students will produce the evidence for the assessment argument. Examples include descriptions of a set of readings and requirements for a kind of assignment, such as an essay or oral presentation. Work Products are what is captured from among things students will say, do, or make, as elicited by the task. Each Work Product object describes a specific thing, such as a marked response, an essay, a sequence of trouble shooting actions, or a rationale for an investigation.

4.2.2 Defining Task Features

Task Model Variables (TMVs) describe key features of stimulus materials or relationships among them, tools and affordances made available to students, or other aspects of the environments in which students work. TMVs identify particular dimensions along which tasks can vary and either indicate a range of variation or specify a particular value along that dimension. Some Task Model Variables concern features of particular stimulus materials, others concern relationships among stimulus materials, and still others may concern relationships between a task and characteristics of an examinee's background (such as familiarity with a topic). TMVs can play a number of roles in the design and operation of assessment tasks, such as providing information to task authors, controlling task difficulty or focus of attention, guiding task selection, and moderating parameters in the statistical model (Mislevy, Steinberg, & Almond, 2002).

Some Task Model Variables concern an entire *template* (e.g., content area, type of assessment), whereas others will concern only particular stimulus materials or local conditions within a *template* (e.g., length of essay). The specification of Task Model Variables always occurs during *template* construction, but the actual values each Task Model Variable assumes can be preset either during the *template* construction phase or in subsequent phases of *task specification* or implementation. Once they are set in a given *template*, TMV values are available to all related objects used within that *template*. For example, an assessment designer may select three TMVs while creating a specific *template*: topic area, length of essay, and content area. In the case that all assessments generated from this *template* will address, say, "Cognitive Psychology" as the topic area, the *template* would specify both the TMV and its setting.[3] The length of essay TMV may be left open to the teacher; thus, this TMV, although indicated at the *template* level, will receive its setting at the *task specification* level. Finally, since the teacher has decided that each student is free to select a content area for the assessment, this final TMV would be set at implementation by each student's choice.

[3] While each TMV is an object in its own right, the settings it assumes are always "owned" by the *template*. Different *templates* can use the same TMV but specify that it be set to different values as they are appropriate to the different *templates*. Thus, among the attributes of a template are Task Model Variable Settings. The TMV objects carry within themselves the range of possible values they can take on, but it is the *template* that owns the settings it needs for each of these.

4.2.3 Gathering Evidence

In an operational assessment, the Work Products a student produces are input to the scoring processes laid out in the ECD Evidence Model. *Design patterns* sketch *potential observations* as dimensions of quality a particular Work Product may exhibit and offer examples of potential scoring *rubrics* by which these qualities might be evaluated. *Templates* specify the exact forms and details of scoring procedures by means of a web of several interrelated objects. Evaluation Procedures contain a sequence of Evaluation Phases, which channel relevant Work Products through a series of steps that assess their salient qualities in the form of values of Observable Variables. As depicted in Figures 2 and 3, a Measurement Model serves as a bridge to the Student Model in terms of a statistical or psychometric model for values of an Observable Variable, given values of its "parent" Student Model Variable(s).[4] The current Measurement Models in PADI are all special cases of the MRCMLM (although the PADI object model can be extended to accommodate measurement models for alternative scoring engines). A Measurement Model may provide information about several Student Model Variables, depending on the relevance of the skill or knowledge being assessed to the overall Student Model. By creating "item bundles" in a preceding Evaluation Phase, an assessment designer can model conditionally dependent observable variables (Wilson & Adams, 1995). Item bundles make it possible to weigh evidence from multiple responses that are conditionally dependent and, when appropriate, to retain distinctions of student proficiency across multiple dimensions. Item bundles would be important, for instance, when both accuracy of student responses and quality of explanation for the response are important for the inferences of interest. Although we must assume some relationship between these two pieces of information, each aspect of the response nevertheless contributes unique information to our assessment of student proficiency.

4.2.4 Making Inferences about Student Proficiency

Recall that the Student Model in *design patterns* is cast in terms of KSAs, that is, students' knowledge, skills, and abilities related to the assessment, our final goal. In *templates*, considerations such as context, use, and purpose determine how to move from the narrative level of KSAs to the formal statistical entities labeled Student Model Variable(s) and an overall Student Model. Each Student Model Variable corresponds to a specific dimension of the overall Student Model we wish to assess. Student Model Variables are described by attributes, including type (e.g., continuous, categorical), minimum and maximum values, and categories of possible values. For example, an assessment targeting students' ability involving both science content knowledge and inquiry skills (as in the case of BioKids; Songer & Wenk, 2003) may operationalize KSAs as two Student Model Variables—say, Biodiversity Knowledge and Building Explanations. Although an Observable Variable will assume a particular value once a student's performance is evaluated, knowledge about a Student Model Variable is never certain but rather is expressed by a *probability distribution* expressing current beliefs about a student's value for that variable. As noted above, Measurement Models are the objects in *templates* that

[4] Specifically, by an application of Bayes Theorem. Before data arrive, the Student Model contains a prior distribution expressing what is known about the possible values of a given student's Student Model Variables. This may be a population distribution or a noninformative distribution. The student's responses, in the form of values of Observable Variables, induce a likelihood function for the Student Model Variables. The form of the likelihood function depends on the particular psychometric model being used and its parameters, which are indicated in the Measurement Model object associated with each task.

contain the information about how values of Observable Variables should update the distribution in a Student Model.

4.2.5 Activities and Attributes

As illustrated in Figures 2 and 3, the group of related presentation materials, work products, evaluation rules, observable variables, and measurement models described above is itself an object in the PADI system called an Activity. A *template* can contain one or many activities; the decisions of whether to have several activities and how to define the scope of an activity are up to the assessment developer. For example, an inquiry task may have several distinct stages of investigation, and such stages act as a natural partitioning of the task into activities. As a second example, a task consists of a list of "mix and match" subtasks, from which a student may choose; each of the integral subtasks could be cast as an activity within a single *template*. All activities within a *template* will update Student Model Variables found within a single student model, as specified by the *template*.

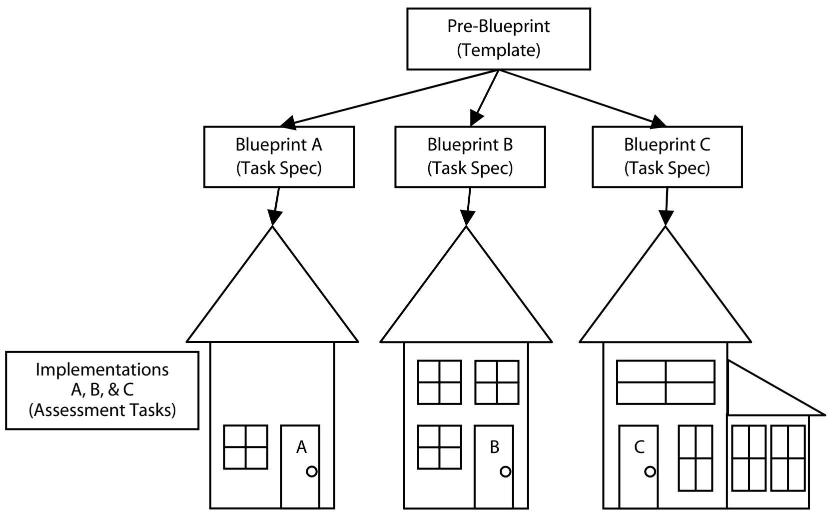
In Section 3, *design patterns* were described as a set of "slots" filled with content. In the similar "slot-like" structure of *templates*, most slots are filled in with instantiations of the objects described above. However, *templates* also include attributes, similar to those in *design patterns*, that are unique to a *template*, which are filled in with narrative. A Summary attribute serves as an introduction to the overall aims of a given *template*, complemented by further summary "slots" containing text descriptions of the Student Model, Measurement Model, Evaluation Procedures, Work Product, Task Model Variable, and Activities that make up each *template*. In addition to a Title attribute, each *template* includes a Type attribute indicating whether the object is a *template* (abstract and general) or a *task specification* (finished, complete, concrete). The process by which *task specifications* are generated from *templates* is detailed in Section 6.0 below.

4.2.6 Putting the Pieces Together

The various objects in the *template* system can be related in many configurations corresponding to different kinds of tasks and analyses, so the *template* structure also defines the possible relationships that can exist among objects and the nature of those relationships (e.g., one-to-many, one-to-one). For example, whereas some evaluation procedures will consist of a simple configuration of a single evaluation phase and observable variable, others will use several in a multistep evaluation process. Further, a unique set of "slots" defines each object in the PADI system. As we saw in the case of *templates*, for many of these objects (e.g., Evaluation Procedures) some "slots" will contain other objects in the system, and others will contain text descriptors. For other objects (e.g., educational standards), no references are made to other objects, and only text (which may include URLs to materials on the Internet) fills the set of "slots." A comprehensive table of objects and their structure is presented in Appendix B. Section 5.0 uses a sample *template* to illustrate some of the variations of arrangements of the objects in a *template*.

Figure 2 represents the simplest configuration of the main *template* objects. As mentioned earlier with respect to *design patterns*, it is important to distinguish between *structure* and *object*. In PADI, references to a task *template* imply that specific objects have been selected from the "assessment construction kit," arranged in a particular configuration, and assigned values or ranges of values. Extending the apartment building analogy, consider a development in which there are three or four variations of apartment buildings, which have many fundamental aspects and configurations in common. We might imagine a "pre-blueprint" capable of generating a range of blueprints, one for each of the various building types in the development. Furthermore, each blueprint could be used to construct multiple copies of a specific building, as depicted in Figure 4. Mapping the PADI terms to this example, the *templates* serve the purpose of the "pre-blueprint." *Task specifications* (*task specs*), in which the options represented in the *template* are fixed, are parallel to blueprints for specific buildings. All that remains is to construct the actual assessment by implementing a *task spec,* analogous to constructing a physical building from a blueprint.

Figure 4. Houses Generated From the Same Template Structure



4.3 *The Extensibility of Templates*

The structures of the PADI *template* described here, illustrated in Section 5.0, and detailed in Appendix B, are meant to be extensible. That is, the *template* object model will be made accessible to users, who can specialize objects (i.e., subclass the existing classes of PADI objects), add to the collection of classes, and include PADI objects or private extensions of PADI objects in an object model for their own assessment design systems and applications.[5] (Indeed, XML representations of PADI design objects for data and measurement objects are themselves extensions of the protocols of the IMS/QTI project on international standards for electronic learning and assessment objects.)

There are two reasons for providing an open architecture and encouraging extensions. The first reason is that to enable the PADI object model to support as wide a variety of assessment tasks as possible while maintaining the same general argument structure and forms for expressing constituent elements, the object model will not be optimized for any particular assessment design project. Designers interested in a particular kind of task or supporting the work of particular kinds of designers (e.g., researchers, classroom teachers, commercial test publishers,

[5] Uses of the PADI system and/or object model described herein must be consistent with license agreements, as provided for in the conditions of the PADI grant and PADI ownership rights.

developers of learning systems) can extend, constrain, or wrap the PADI objects within an interface more specifically suited to their users.

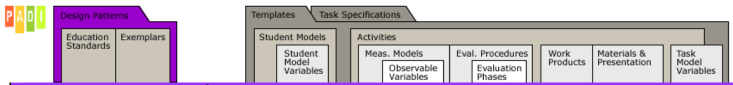
The second reason is that the openness allows extensions beyond the kinds of tests and processes that can be dealt with by the current object model, including ones that do not exist at present. For example, the scoring engine PADI is developing to accompany the design system accommodates psychometric models that can be expressed as special cases of the MRCMLM, and the PADI student model and measurement model design objects are compatible with these models. But developers can provide scoring engines of any other type, and if the existing student- and measurement-model class definitions are not sufficient, private extensions to these classes can add parameters or connections to other design objects as required.

5.0 A Sample Template: EDMS 738

A *template* created for a series of assignments in a graduate course in the fundamentals of assessment ("EDMS 738") offers a concrete example for exploring the PADI design system. In this course, which is built around a model for evidence-centered assessment design, students are expected to understand and be able to apply the principles of evidentiary reasoning and design-under-constraints common across assessment forms and purposes. As a class, students explore how a coherent assessment design builds on each of several disciplines, including cognitive and situative psychology, evidentiary reasoning, measurement models, and Bayesian statistical inference. Students then work individually to explicate an assessment (preferably one in which they have a professional interest) in terms of ECD and delivery system models. Although this example does not share PADI's focus on scientific knowledge per se, it illustrates the domain-free nature of the PADI object system and the ECD approach to assessment.

As described in Section 3.0, *design patterns* make explicit the theoretical underpinnings of assessment and guide the structure and content of the assessments they inform. The construction of the EDMS 738 *template* is informed by the "Model Elaboration" *design pattern* displayed in Figure 5 . The general process of mapping particular situations to scientific schemas or models addressed in this *design pattern* is instantiated here by students' analyzing assessments of their choice from the perspective of ECD.

Figure 5. “Model Elaboration” Design Pattern Object in PADI Design System



Model elaboration ǀ Design Pattern 84

Component	Value
Title	Model elaboration
Summary	This design pattern concerns working with mappings and extensions of given scientific models. *Comment—A central element of scientific inquiry is reasoning with models. This DP focuses on model elaboration, as a perspective on assessment in inquiry and problem-solving.*
Focal Knowledge, Skills and Abilities	- Establishing correspondence between real-world situation and entities in a given model - Finding links between similar models (ones that share objects, processes, or states) - Linking models to create a larger, more encompassing model - Within-model conceptual insights
Rationale	Scientific models are abstracted schemas involving entities and relationships, meant to be useful across a range of particular circumstances. Correspondences can be established between them and real-world situations and other models. Students use, and gain, conceptual or procedural knowledge working with an existing model. *Comment—Students' work is bound by the concept of an existing model (or models) so their work includes an understanding the constraints of the problem.* *Even though model elaboration does not involve the invention of new objects, processes, or states, it does entail sophisticated thinking and is an analogue of much scientific activity.* *Even though model elaboration does not involve the invention of new objects, processes, or states, it does entail sophisticated thinking and is an analogue of much scientific activity.*
Additional Knowledge, Skills and Abilities	Familiarity with task type (e.g., materials, protocols, expectations) Subject-area knowledge
Potential observations	- Catenating models across levels (e.g., individual-level and species-level models in transmission genetics) - Determining the degree to which observations correspond with predictions. - Explanation of modifications, in terms of data/model anomalies - Identifying ways that a model does not match a situation (e.g., simplifying assumptions), and characterizing the implications. - Mapping out the corresponding elements between a real-world situation and a scientific model.
Potential work products	- Correspondence mapping between elements or relationships of model and real-world situation - Correspondence mapping between elements or relationships of overlapping models - Elaborated model - Written/Oral Explanation of reasoning behind elaboration
Potential rubrics	
Characteristic features	Real-world situation and one or more models appropriate to the situation, for which details of correspondence need to be fleshed out. Addresses correspondence between situation and models, and models with one another.
Variable features	Is problem context familiar? Model given to student(s), vs. model to elaborate produced by student(s) themselves. Must experimental work or supporting research be carried out in order to ground the elaboration? Single model to elaborate, vs. establishing correspondence among models at different levels or with different focus?
I am a kind of	
These are kinds of me	
These are parts of me	
Educational standards	
Templates	
Exemplar tasks	
Online Resources	
References	Biomass project http://www.education.u… Marshall, S.P. (1995). Schemas in problem solving. Cambridge: Cambridge University Press. NSES standards Stewart, J., & Hafner, R. (1994). Research on Problem Solving: Genetics. In D. Gabel (Ed.), Handbook of Research on Science Teaching and Learning (pp 284-300). New York: MacMillan. White, B. Y., & Frederiksen, J. R. (1998). Inquiry, Modeling, and Metacognition: Making Science Accessible to All Students. Cognition and Instruction, 16(1), 3-118.

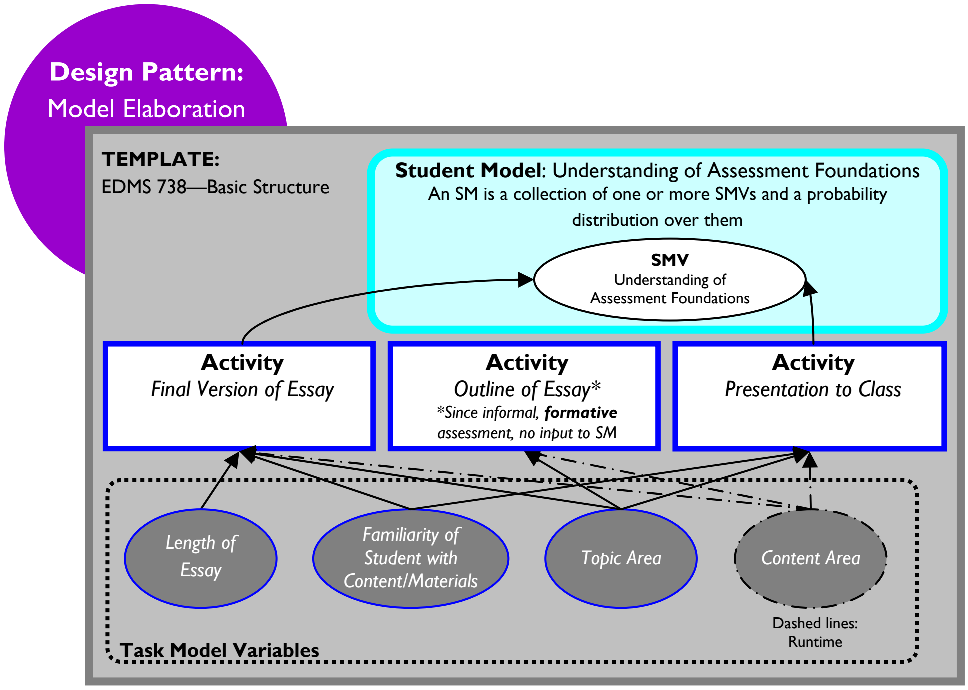
The representation of this *template*, “EDMS 738 Assignments,” is shown in Figure 6. To facilitate explanation, Figure 7 shows a graphic representation that illustrates the basic configuration of this *template*, consisting of a single Student Model and three Activities. A range of assessments for this course will be generated from this *template*. Most of the Task Model Variables (TMVs) represented indicate decisions that an assessment designer must make before deploying this assessment. In this case, “Topic Area” is such a characteristic, which will be assigned by the instructor. This is an example of a *template*-wide TMV, since it applies to the entire *template* and thus to all Activities and students. Alternatively, some TMVs apply only to certain Activities, such as the “Length of Essay” TMV shown in Figure 7.

Figure 6. Basic “EDMS 738 Assignments” Template Object in the PADI Design System

EDMS 738 Assignments | Template 78

Component	**Value**
Title	**EDMS 738 Assignments**
Summary	Assessments for Bob Mislevy’s course in Fundamentals of Assessment at U Maryland. Topics are assigned by instructor, in connection with the study, readings, and discussion of those topics through the course. Students have choice about the particular actual assessment (i.e., the ‘content area’) that they will analyze in their essay. The aspect(s) of assessment design, analysis, or implementation they will address in the assignment (i.e., the topic) is determined by the instructor. *Comment—required: knowledge of how to create word document on PC*
Type	
Student Model Summary	One overall summary variable of proficiency
Student Models	EDMS Overall Proficiency Model. Defines a univariate student model, with a continuous variable that signifies proficiency in applying concepts and knowledge representations of assessment design to assessment of student’s choosing.
Measurement Model Summary	univariate
Evaluation Procedures Summary	Generic rubrics *Comment—There are rubrics associated with the activity phases that can be used across specific topic areas.*
Work Product Summary	Essay in MS Word format is main work product. Optional activities can produce draft outline, and in-class presentation with charts.
Task Model Variable Summary	Real-world situation and one or more models appropriate to the situation, for which details of correspondence need to be fleshed out. Addresses correspondence between situation and models, and models with one another.
Template-level Task Model Variables	Topic area. Topics for essay about assessment Content area. Specific domain content under consideration Amount of scaffolding. The task can guide students to think about certain concepts or can help students structure their ans… Familiarity of student with content/materials. EDMS Assignment Type. The desired kind of EDMS assignment. The list of possible responses should be the list of templates …
Task Model Variable Settings	
Materials and Presentation Requirements	Optional draft outline is take-home activity, can include unlimited use of materials and resources, two weeks in duration. Main activity is take-home essay, one week duration, open book. Optional class presentation is 10-minute oral presentation, with PowerPoint projection available for student’s use.
Template-level Materials and Presentation	
Materials and Presentation Settings	
Activities Summary	1. (optional) review by instructor of outline by examinee 2. final draft 3. (optional) presentation to class
Activities	Outline of essay. An outline of the essay is turned in to the instructor, and formative feedback is provided back to t… Presentation to class. Presentation of key points in essay to the class. Final version of essay. This is the final essay that is turned in for a grade.
Tools for Examinee	computer with MS Word textbook, readings for course
Exemplars	
Educational Standards	
Design Patterns	Model elaboration. This design pattern concerns working with mappings and extensions of given scientific models.
I am a kind of	
These are kinds of me	EDMS 738 Task Spec I - Psych and Your Assessment. An assessment for Bob Mislevy’s course in Fundamentals of Assessment at U. Maryland. The topic is “… EDMS 738 Task Spec II - Final Essay. This is the final assessment for Bob Mislevy’s course in Fundamentals of Assessment at U. Maryland, …
These are parts of me	EDMS 738 Final Version of Essay. Assessments for Bob Mislevy’s course in Fundamentals of Assessment at U Maryland. This template is … EDMS 738 Outline of Essay. Assessments for Bob Mislevy’s course in Fundamentals of Assessment at U Maryland. This template is … EDMS 738 Presentation to Class. Assessments for Bob Mislevy’s course in Fundamentals of Assessment at U Maryland. This template is …
Online Resources	http://www.education.u… Comment – Used in EDMS 738 Fall 2002, “Cognitive psychology and educational assessment”
References	

Figure 7. Graphic Representation of Basic "EDMS 738 Assignments" Template



Because the scope of this particular *template* is to estimate a student's overall proficiency in the foundations of assessment, the Activities will generate relevant information to update the Student Model. However, the second activity ("Outline of Essay") does not link to the Student Model because of the informal character of that Activity, which is designed to provide feedback to students but not result in a summative evaluation.

Having reviewed the general composition of the EDMS example, we are ready to unpack each Activity. An expanded view of this *template* appears in Figure 8, and Figure 9[6] reveals the details of Activity "Final Version of Essay," the outcome of which will inform the Student Model "EDMS Overall Proficiency Model," which in this example contains only a single SMV named "Understanding of Assessment Foundations." As was true of TMVs, multiple instantiations of any given object may be used to assemble an Activity. This Activity uses three Materials and Presentation objects to set the context in which students will produce their Work Product, "Final Essay." As briefly mentioned in Section 4.0, each of these objects is described by a set of attributes. In the case of Materials and Presentation, these attributes include a title and summary, as well as type of materials (e.g., text document, images), the role of stimulus (e.g., directive, nondirective), and the TMVs that influence it. Work Products are in the PADI objects that can be completely described by text and do not reference other objects in their attributes. That is, title, summary, and type attributes (e.g., essay, painting, demonstration), together with possible examples, online resources, and references, are sufficient to describe any given work product. The diagram elaborates on the attributes relevant to this particular activity.

[6] As in Figures 2 and 3, in Figures 8 and 9 the colors blue, yellow, and pink are used to denote objects that correspond to the ECD Student, Evidence, and Task Models, respectively.

Figure 8. Expanded "EDMS 738 Assignments" Template

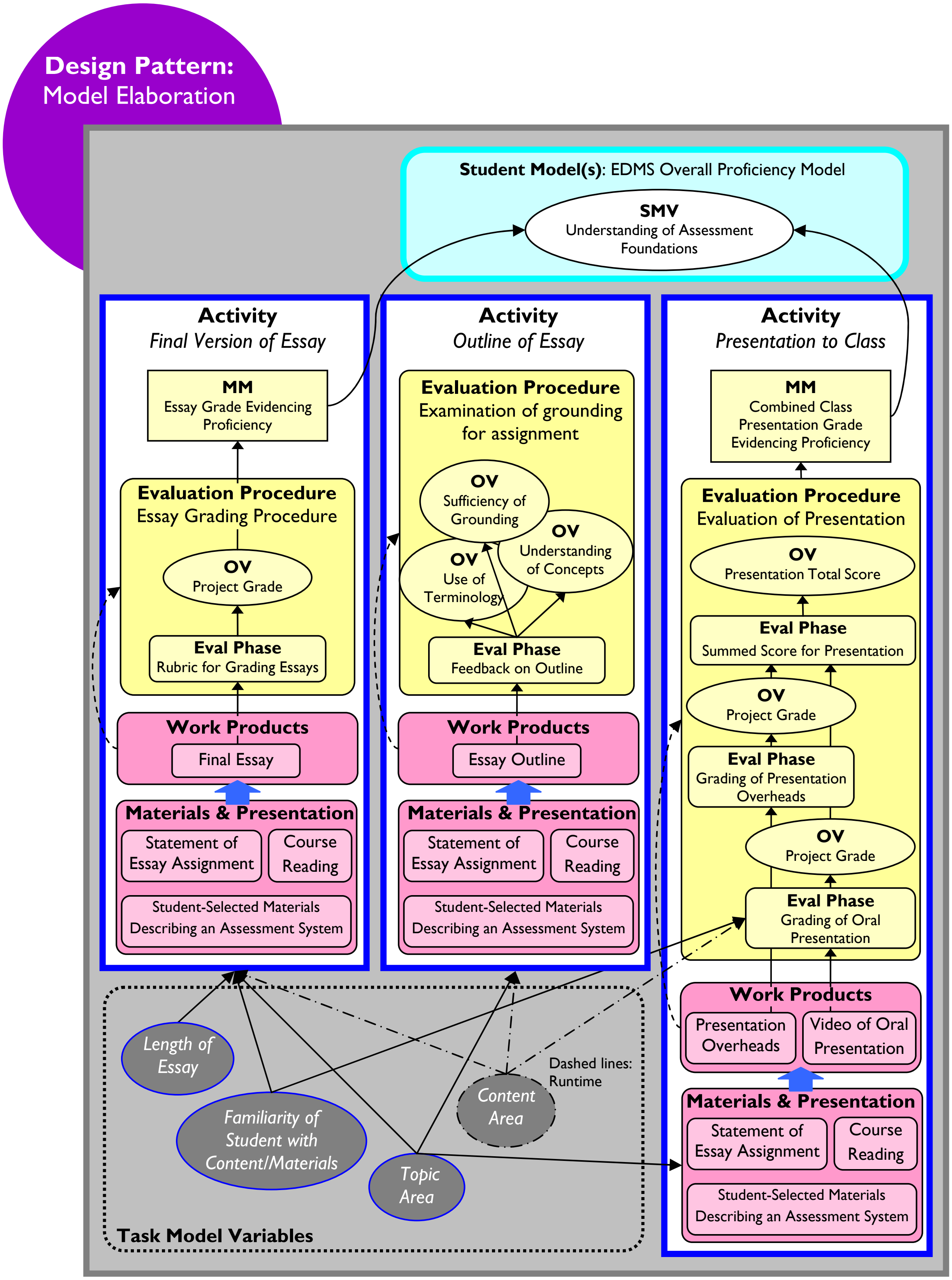
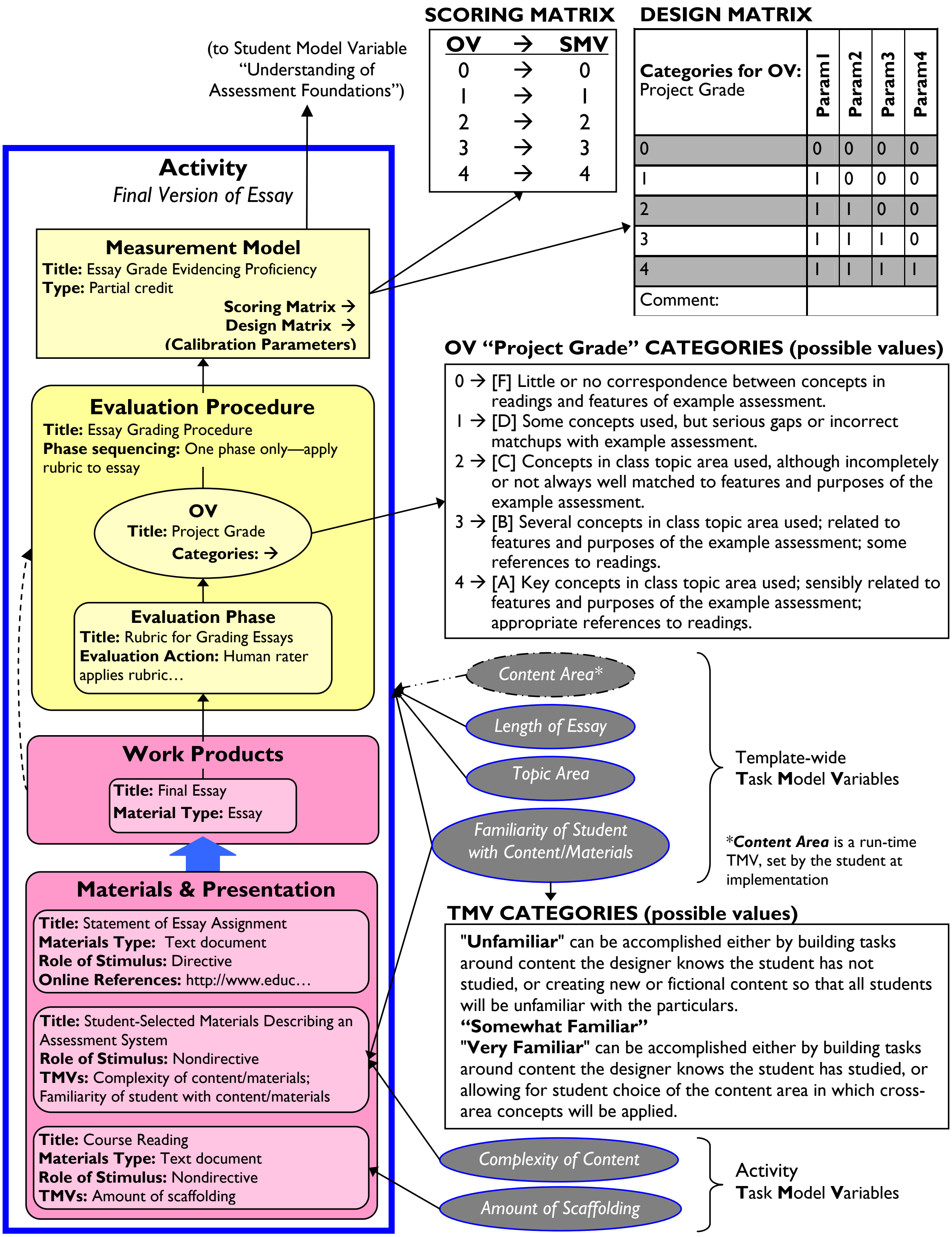
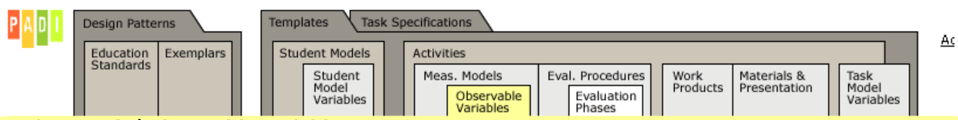


Figure 9. Details of "Final Version of Essay" Activity



Once handed in, the Final Essay Work Product undergoes an Evaluation Procedure, which specifies a set of evaluation phases and the sequence in which they are implemented. In the case of this activity, the evaluation procedure is a simple process consisting of only one Evaluation Phase, "Rubric for Grading Essays," which assesses the work product. This phase is similar to the classroom grading process familiar to most people, in which an essay would be assigned a grade from A to F or a score from 1 to 4, based on a set of criteria. In the PADI system, the Observable Variables are evaluative summaries of the key aspects of a performance—essentially a generalization of the more familiar concept of a project grade or item score. As shown in the details in Figure 9, among the attributes of each Observable Variable is a description of possible categories it can assume. Within the PADI design system, an assessment designer may use the Comment field to provide explanations of the meaning of each categorical value, as shown in Figure 10.

Figure 10. Observable Variable "Project Grade" in the PADI Design System



Project Grade | Observable Variable 92

Component	**Value**
Title	Project Grade
Summary	Grade for how well the student has used the course concepts to analyze the targeted aspect(s) of their example assessment. *Comment – Students are provided feedback in terms of a letter grade label: A, B, C, D, F. These correspond to values of 4, 3, 2, 1, 0.*
Categories (possible values)	0 *Comment – Little or no correspondence between concepts in readings and features of example assessment.* 1 *Comment –Some concepts used, but serious gaps or incorrect matchups with example assessment.* 2 *Comment –Concepts in class topic area used, although incompletely or not always well matched to features and purposes of the example assessment.* 3 *Comment –Several concepts in class topic area used; related to features and purposes of the example assessment; some references to readings.* *4* *Comment –Several concepts in class topic area used; related to features and purposes of the example assessment; some references to readings.*
Online Resources	
References	
I am a part of	Essay Grade Evidencing Proficiency. (Measurement Model) Rubric for grading essays. (Evaluation Phase)

A further step is needed to update the Student Model. The Student Model takes the form of a probability distribution over the Student Model Variables, which can synthesize information over more than one activity. This is accomplished by using the Measurement Model "Essay Grade Evidencing Proficiency." This Measurement Model indicates that the form of the psychometric model to be used for the updating is a univariate Rasch partial-credit psychometric model, and it conveys the parameters of the model. The PADI scoring engine uses this information and a student's value of the Observable Variable "Project Grade" to update the probability distribution for that student's Student Model Variable "Understanding of Assessment Foundations."

The Student Model Variable "Understanding of Assessment Foundations" is contained in the Student Model object, in this example named "EDMS Overall Proficiency Model." This Student Model contains only this one Student Model Variable, although a Student Model can contain several Student Model Variables. A Student Model also defines slots for a probability distribution over the Student Model Variables. A Measurement Model such as the one described above contains all the information necessary for updating a Student Model distribution via Bayes Theorem, when a value for the Observable Variable in the Measurement Model is realized. Thus, a Measurement Model describes the Observable Variable that can be used as input to a scoring process that updates the Student Model distribution, in accordance with the form and the parameters of the psychometric model that the Measurement Model describes.

The relationship among objects for the "Final Essay" Activity reflects a fairly simple configuration. The two other Activities for this *template* offer examples of other ways the objects can combine at the *template* level to describe an eventual assessment task. In contrast to the "Final Version of Essay" Activity, the "Outline of Essay" Activity is intended only to provide formative feedback to the student by means of three Observable Variables (i.e., use of terminology; sufficiency of grounding; understanding of concepts). Since student achievement on this Activity is not meant to contribute to the overall grade for the course, there are no connections from this Activity out to the Student Model, and thus there is no Measurement Model.

The third Activity, "Presentation to Class," represents a more complex use of the PADI design system objects, particularly with respect to evaluation procedures. Whereas the Evaluation Procedures in each of the first two Activities contained exactly one evaluation phase, this Activity uses three phases. The first two phases each evaluate different Work Products (i.e., presentation overheads, video of oral presentation), resulting in two project grades, one for each Work Product. These two variables are conditionally dependent, as aspects of performance in the same context. In this example, we account for this dependence by combining their information into a single total Observable Variable. This is the role of the third Evaluation Phase for this Activity, which simply sums the values of the other two Observable Variables to produce a third Observable Variable, "Presentation Total Score." In the present example, the two scores being summarized are equally weighted. One potential variation would be to assign more weight to one Observable Variable in calculating a total score for the presentation. Another would be to map each possible pair of scores on the two component Observable Variables to a specific score on the total Observable Variable.

Thus far, we have explored the EDMS sample *template* as an illustration of both the range of possible configurations of objects and the initial shaping of what will ultimately become a concrete assessment. Having articulated the *template*, the next step toward this concrete actualization is to make a further set of decisions from among the range of options specified in the *template*. In the PADI design system, this is done by creating a *task specification* (*task spec*) from the *template*.

6.0 Sample Task specs from the EDMS 738 Template

Some of the benefits of developing a complete *template* have already been accrued by this point—in particular, the conceptual benefits of explicating the rationale and the elements for the assessment argument that this task is meant to embody. We begin to reap operational benefits when it comes time to generate *task specs*, which share the same essential structure as *templates*. Referring back to our building analogy, we are now ready to generate specific blueprints for a specific house. Likewise, a *task spec* is a blueprint for implementing a specific task, one of the family of tasks that could be generated from the same *template*, all sharing the same essential structure and evidentiary argument. What distinguishes a *template* from a *task spec* is not its constituent objects nor the ways they relate to one another, but rather the specificity of the settings for the defining attributes and objects. A *task spec* inherits all information and settings of its parent *template*. In the transition, some contents will remain unchanged, some will be slightly edited, and others will have settings or options selected.

Table 2 maps this transition according to the extent of changes made. To facilitate an understanding of this transition, we will consider how the existing *template* would have to change to specify a two- to three-page writing assignment about "psychology and your assessment."

Table 2. Template to Task Spec Transitions

Unchanged	Possibly Edit (Text)	Set Or Select
▪ Student Model Summary ▪ Student Model ▪ Student Model Variables (perhaps only a subset of the SMVs will be used, as determined by activities) ▪ Measurement Model Summary ▪ Design Patterns	▪ Title ▪ Summary ▪ Evaluation Procedures Summary ▪ Work Product Summary ▪ TMV Summary ▪ Presentation Environment Requirements ▪ Activities ▪ Tools for Examinee ▪ Online Resources ▪ References	▪ Type (becomes Task Spec) ▪ TMV Settings ▪ Materials & Presentation ▪ Activities ▪ Exemplars ▪ I am a Kind of (indicates generating template)

The Student Model in which we are interested is still "EDMS Overall Proficiency Model," which this specific assignment on the topic of "psychology and your assessment" will inform. Consequently, as shown in the "Unchanged" column of Table 2, the Student Model Summary and the Student Model do not vary from the originating *template*. The inherited attributes and objects in the "Unchanged" column remain unchanged at the *task spec* level. This is also true of the Measurement Model and *design pattern(s)* that inform the assessment. Although the range of Student Model Variables will not increase in the transition from *template* to *task spec* in general, it is possible that by selecting a subset of Activities, the relevant SMVs will be just a subset of those identified by the *template*.

Although all information from the *template* will be inherited by the *task spec,* some of this information will require editing to reflect the specifics of the particular assessment we are describing. The "text edits" column identifies the attributes that may undergo changes in the creation of a *task spec* from a *template*. As an illustration, compare the three "Summary" descriptions in Table 3. The Summary for the *task specs* will be a specialization, or narrowed-down version, of the Summary at the *template* level.

Table 3. *Template* and *Task spec* Summary Samples

Summary		
***Template*: EDMS 738 Assignments**	***Task spec* I: Psychology & Your Assessment**	***Task spec* II: Final Essay**
Assessments for Bob Mislevy's course in Fundamentals of Assessment at U Maryland. Topics are assigned by instructor, in connection with the study, readings, and discussion of those topics through the course. Students have choice about the particular actual assessment (i.e., the 'content area') that they will analyze in their essay. The aspect(s) of assessment design, analysis, or implementation they will address in the assignment (i.e., the topic) is determined by the instructor.	An assessment for Bob Mislevy's course in Fundamentals of Assessment at U Maryland. The topic is "psychology and your assessment," in connection with the study, readings, and discussion of those topics through the course. Students have choice about the particular actual assessment (i.e., the 'content area') that they will analyze in their essay.	This is the final assessment for Bob Mislevy's course in Fundamentals of Assessment at U Maryland, for which individual students produce an integrated discussion of an assessment based on the course study, readings, and discussion. Students select the particular actual assessment (i.e., the 'content area') they will analyze in their essay.

Finally, a number of decisions are made at the *task spec* level with regard to relevant Task Model Variables, Activities, and other objects. The third column of Table 2 identifies the objects that must be either preset or selected from the range offered by the *template*. In the "psychology and your assessment" assessment, the TMVs of "length" and "topic" have been preset by the instructor to "short 2-3 pages" and "psychological underpinning," respectively. Further, of the three Activities listed in the *template*, only the "final essay" object is retained in the "psychology and your assessment" *task spec.* These choices dictate further selections from among the possibilities expressed by the *template*. For example, since the Presentation to Class Activity is not retained in this *task spec,* the Observable Variables associated with that Activity will not be used in this particular assessment.

The decisions described above result in a *task specification* that has inherited all the theoretical underpinnings of its "parent" *template*, yet is one step closer to the implementation layer. Indeed, with these selections in place, all the information necessary to instantiate this *task spec* is in place. A screen shot of this *task spec* as it appears in the PADI system is presented in Figure 11.

Figure 11. "EDMS 738 Task Spec I" in PADI Design System

EDMS 738 Task Spec I – Psych and Your Assessment	Task Specification 308
Component	**Value**
Title	EDMS 738 Task Spec I- Psych and Your Assessment
Summary	An assessment for Bob Mislevy's course in Fundamentals of Assessment at U. Maryland. The topic is "psychology and your assessment", in connection with the study, readings, and discussion of those topics through the course. Students have choice about the particular actual assessment (i.e., the 'content area') that they will analyze in their essay.
Type	*Comment – Though still a work-in-progress*
Student Model Summary	One overall summary variable of proficiency
Student Models	EDMS Overall Proficiency Model. Defines a univariate student model, with a continuous variable that signifies proficiency in applying concepts and knowledge representations of assessment design to assessment of student's choosing.
Measurement Model Summary	univariate
Evaluation Procedures Summary	Generic rubrics *Comment – There are rubrics associated with the activity phases that can be used across specific topic areas.*
Work Product Summary	Essay in MS Word format is main work product.
Task Model Variable Summary	
Template-level Task Model Variables	Length of essay. Long or short assignments Topic area. Topics for essay about assessment
Task Model Variable Settings	
Materials and Presentation Requirements	Main activity is take-home essay, one week duration, open book.
Template-level Materials and Presentation	
Materials and Presentation Settings	
Activities Summary	
Activities	Final version of essay. This is the final essay that is turned in for a grade.
Tools for Examinee	- computer with MS Word - textbook, readings for course *Comment – knowledge of how to create word document on PC*
Exemplars	Foundations of Assessment - IEA Civic Education Assessment. The assessment design of the IEA Civic Education Study emphasized the perspective of the situative a… Foundations of Assessment - Multicultural Counseling Inventory. Paper available online at: http://www.education.umd.edu/EDMS/mislevy/FoundationsFall2003/PastExamp… Foundations of Assessment - NAEP 8th grade mathematics. Student paper examining the Mathematics framework for the 1996 and 2000 National Assessment of Educa…
Educational Standards	
Design Patterns	Model elaboration. This design pattern concerns working with mappings and extensions of given scientific models.
I am a kind of	EDMS 738 Assignments. Assessments for Bob Mislevy's course in Fundamentals of Assessment at U Maryland. Topics are assign… EDMS 738 Final Version of Essay. Assessments for Bob Mislevy's course in Fundamentals of Assessment at U Maryland. This template is …
These are kinds of me	
These are parts of me	
Online Resources	http://www.education.u… Comment – Used in EDMS 738 Fall 2002, "Cognitive psychology and educational assessment"
References	
I am a part of	EDMS assessment. (Template)

In terms of its place in the EDMS 738 course, this assignment is of moderate intensity—as evidenced by the short length of the essay and the narrow scope of the topic—and intended to refine students' thinking about how the assessment they selected (represented by the runtime TMV "content area") reflects multiple psychological traditions. By the end of the semester, students would be expected to demonstrate a broader and more critical understanding of the foundations of assessment. Let us then craft a *task spec* from our EDMS 738 *template* that would be appropriate for this second, more comprehensive task.

From a copy of the EDMS 738 Assignments *template*, we start by modifying the summary text to describe a final essay assignment. We will also change the "type" attribute from "abstract *template*" to "concrete specification" to identify this object as a *task spec* rather than a *template*. The remaining summaries will not change, so we proceed to the Task Model Variable Settings. In contrast to the first *task spec,* the appropriate length of this essay is long, so we set the "length of essay" to the option for "Long: 15-25 pages." (The short and long labels and page specifications were defined when we created this TMV object. Other levels of this object could be offered by modifying that TMV attribute.) Since this assignment is intended to integrate all course topics, we set the "topic area" TMV to "entire span of assessment structure," as shown in Figure 12. After we have indicated these settings, the PADI design system will display them from the "TMV Settings" screen for this *task spec,* as shown in Figure 13.

Figure 12. Setting "Topic Area" TMV Value in the PADI Design System

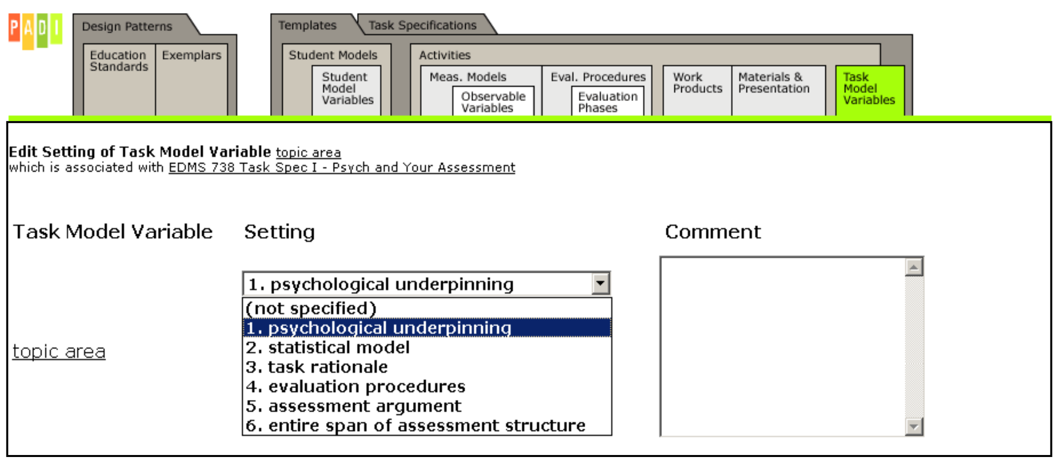


Figure 13. Displaying TMV Settings for Task Spec II

Design Patterns | Education Standards | Exemplars | Templates | Task Specifications | Student Models | Student Model Variables | Activities | Meas. Models | Observable Variables | Eval. Procedures | Evaluation Phases | Work Products | Materials & Presentation | Task Model Variables

Display settings of Task Model Variables (TMVs)

The settings below apply to the template EDMS 738 Task Spec I - Psych and Your Assessment only. Although TMV objects may be associated with multiple Template objects and multiple Action and Materials/Presentation objects, the settings for TMVs are remembered by the template only. In other words, a TMV may be shared, but the setting of the TMV is an owned attribute of the template, and the setting applies to all parts of the template.

TMV	Setting	Comment
Amount of scaffolding	(not specified)	
Complexity of content/materials	(not specified)	
Content area	RUNTIME - Set by students	
Familiarity of student with content/materials	Somewhat familiar	
Length of essay	Short: 2-3 pages	
topic area	1. psychological underpinning	

TMVs associated with Activity Final version of essay

Content area

Familiarity of student with content/materials

Length of essay

topic area

TMVs associated with Materials and Presentation Student-selected materials describing an assessment system

Complexity of content/materials

Familiarity of student with content/materials

TMVs associated with Materials and Presentation Course reading

Amount of scaffolding

Proceeding to the Presentation Environment Requirements attribute, we modify the text to discuss only the final essay and outline Activities, since for this assignment we will not include a presentation activity. This decision is also reflected in the Activities attribute, from which we remove the "presentation to class" object. The final changes we make are to the Materials & Presentation Settings and Activity-level TMV settings. Thus, the specific course readings and statement of essay assignment values are set to reflect this particular assignment. However, since the "content area" TMV is determined by the student, note that this TMV remains "unspecified" even at the *task spec* phase. This run-time TMV will only be set at the actual implementation. The *task spec* generated by these settings (*Task Spec* 2) is shown in Figure 14.

Figure 14. "EDMS 738 Final Essay" Task Spec II Object in PADI Design System

EDMS 738 Task Spec II – Final Essay | Task Specification 340

Component	Value
Title	EDMS 738 Task Spec II – Final Essay
Summary	This is the final assessment for Bob Mislevy's course in Fundamentals of Assessment at U. Maryland, for which individual students produce an integrated discussion of an assessment based on the course study, readings, and discussion. Students select the particular actual assessment (i.e., the 'content area') they will analyze in their essay.
Type	
Student Model Summary	One overall summary variable of proficiency
Student Models	EDMS Overall Proficiency Model. Defines a univariate student model, with a continuous variable that signifies proficiency in applying concepts and knowledge representations of assessment design to assessment of student's choosing.
Measurement Model Summary	univariate
Evaluation Procedures Summary	Generic rubrics *Comment – There are rubrics associated with the activity phases that can be used across specific topic areas.*
Work Product Summary	Essay in MS Word format is main work product.
Task Model Variable Summary	
Template-level Task Model Variables	Length of essay. Long or short assignments Topic area. Topics for essay about assessment
Task Model Variable Settings	
Materials and Presentation Requirements	Optional draft outline is a take-home activity, can include unlimited use of materials and resources, two weeks in duration. Main activity is take-home essay.
Template-level Materials and Presentation	
Materials and Presentation Settings	
Activities Summary	
Activities	Final version of essay. This is the final essay that is turned in for a grade. Presentation to class. Presentation of key points in essay to the class.
Tools for Examinee	- computer with MS Word - textbook, readings for course *Comment – knowledge of how to create word document on PC*
Exemplars	
Educational Standards	
Design Patterns	Model elaboration. This design pattern concerns working with mappings and extensions of given scientific models.
I am a kind of	EDMS 738 Assignments. Assessments for Bob Mislevy's course in Fundamentals of Assessment at U Maryland. Topics are assign… EDMS 738 Final Version of Essay. Assessments for Bob Mislevy's course in Fundamentals of Assessment at U Maryland. This template is …
These are kinds of me	
These are parts of me	
Online Resources	http://www.education.u… *Comment – Used in EDMS 738 Fall 2002, "Cognitive psychology and educational assessment"*
References	
I am a part of	EDMS assessment. (Template)

7.0 Conclusion: Some Observations on Structures and Knowledge Representations in Assessment Design

In describing the structure of *templates*, as well as the larger assessment system in which they are embedded, we have hinted at wide-ranging benefits of engaging in what can appear an exaggerated investment of time and resources. Thus, we conclude this presentation by explicating the generative, reusable, and duplicable features of the main products of the PADI system: namely *design patterns*, *templates*, *task specs*, and task objects themselves. By generative, we refer to the value of the PADI data structures and objects for generating related objects that retain the theoretical underpinnings of the "generator." Reusability refers to the ability of PADI objects to be used multiple times, in multiple assessments, while retaining the substance and thinking that went into their development. Finally, duplicable refers to the functional capacity of the PADI design system to make copies of existing objects. Minor changes to objects (e.g., activities, task model variables) are facilitated at the implementation level by this feature.

Consider how reusability in particular plays out for the above-mentioned key objects in the system. The importance of reusability becomes clear when a specific assessment task is considered. Traditionally, an assessment is created when a designer sits down and writes each task itself. An expert in task design will have worked through a process similar to the development of a PADI *design pattern*. But although exemplary tasks may result from this process, the thinking that led there, the design challenges recognized, and the solutions reached are not captured. They are but tacit in the particular task that was produced, invisible to and inaccessible by another designer (or even the same designer) to help create subsequent tasks. Without a framework such as the PADI design system, modifying a particular task to meet a slightly different need or context requires rediscovering or reinventing the same kind of thinking and problem solving in which the original task designer engaged. In contrast, creating a task within the PADI approach takes place within an explicit conceptual framework. The PADI framework supplies theory about the structure of the evidentiary arguments that underlie assessment tasks, while the theoretical work and identification of key aspects of the substantive content of the task are present in the design structures that support the task. In the development of this task, the work, thinking, and rationale have been retained for subsequent task development.

Regarding *task specs*, reusability applies in particular to the thinking, concepts, and strategies for assessing certain inquiry skills. These task "blueprints" can be used repeatedly to generate tasks grounded in domain and assessment principles. Further, the consistency across *task specs* maximizes opportunities for comparability within and across students. The expression of *task specs* as XML code in IMS/QTI standards and documented extensions promotes development and use of tasks across different agencies, developers, and applications.

At the *template* layer, reusability is manifest at the nuts-and-bolts level of expressing concretely and in more detail the operational elements that instantiate an assessment argument described in *design patterns* for a given subject area, for a given purpose. This further level of detail calls for objects such as the student model variables and evaluation rules described in earlier sections. As was shown, *templates* can be used to generate a number of varied tasks but are substantially more specialized and targeted than the thinking made visible in *design* *patterns*. Like *design patterns*, *templates* are reusable and generative, but at a level closer to operational development than to conceptual design.

In addition to generating different kinds of objects (e.g., *design patterns* → *task templates* → *task specifications* → *tasks*), which we could describe as "horizontal reusability," objects in the PADI system feature a "vertical reusability" as well. That is, *templates* can be developed such that one *template* reflects a generalization of a set of more specific *templates*. The parent-child paradigm has been used to describe these relationships. The structure of *templates* and *design patterns* allows for partial development to any degree and with respect to any portions of the structure that might prove useful in a given application. In a large testing program, for example, some test developers could have the job of sketching a number of "master" *templates*, which would be copied and fleshed out by other test developers. Still others might author *task specs* for individual tasks from the more-filled-out *templates*. Developing *templates* and *design patterns* in vertical hierarchies captures and maximizes reusability of the thinking that informs the assessment argument.

Finally, we underscore the importance of building a shared language and forms for communicating and eliciting thinking regarding new problems that operating within a shared framework affords. Apt representational forms (such as blueprints in architecture and *design patterns* in software engineering) organize thinking, embody key principles in a domain, and enable people to coordinate their actions across time, space, and specializations. Indeed, the PADI project itself can be viewed as a response to the challenge laid out by Gitomer and Steinberg in 1999:

> [T]he primary need in supporting a disciplined, integrated, and comprehensive process of assessment design is a tool-based methodology that begins with capturing essential features of the domain content and performance expectations in representations that can be understood by those who must use them—representations that are useful "mediators" in acquiring knowledge and cognitive understanding of the domain.
>
> The dominant tradition in assessment has been psychometric measurement, with its associated set of representations. As assessments address a more diverse and ambitious set of purposes, there comes an increasing need for inferencing that can address complex task performance and instructionally useful information. Thus, the focus is shifting to evidence—what the standards of evidence are for given domains and what the representational possibilities are for achieving those standards. (pp. 367-368)

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. (2002). Enhancing the design and delivery of assessment systems: A four-process architecture. *Journal of Technology, Learning, and Assessment, 1*(5). http://www.bc.edu/research/intasc/jtla/journal/v1n5.shtml

Booch, G., Rumbaugh, J., & Jacobson, I. (1999). *The Unified Modeling Language user guide.* Reading, MA: Addison-Wesley.

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

Gitomer, D. H., & Steinberg, L. S. (1999). Representational issues in assessment design. In I. Sigel (Ed.), *The development of representational thought: Theoretical perspectives* (pp. 351-369). Hillsdale, NJ: Lawrence Erlbaum.

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., Almond, R.G., & Lukas, J. (2003). A brief introduction to evidence-centered design. *ETS Research Report* RR-03-16. Princeton, NJ: Educational Testing Service.

Mislevy, R. J., Chudowsky, N., Draney, K., Fried, R., Gaffney, T., Haertel, G., et al. (2003). Design patterns for assessing science inquiry (*PADI Technical Report No. 1)*. Menlo Park, CA: SRI International.

Mislevy, R. J., Steinberg, L. S., & Almond, R.G. (2002). On the roles of task model variables in assessment design. In S. H. Irvine & P. C. Kyllonen (Eds.), *Item generation for test development* (pp. 97-128). Hillsdale, NJ: Lawrence Erlbaum.

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

Songer, N., & Wenk, A. (2003, April). *Measuring the development of complex reasoning in science*. Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL.

Wilson, M., & Adams, R. J. (1995). Rasch models for item bundles. *Psychometrika, 60*, 181-198.

Yourdon, E., & Constantine, L. (1979). *Structured design*. Englewood Cliffs, NJ: Prentice Hall.

Appendix A

An Evidence-Centered Framework for Assessment Design and Delivery

Evidence-centered assessment design (ECD) is a program of research aimed at developing a principled framework for designing, producing, and delivering educational assessments (Almond, Steinberg, & Mislevy, 2002; Mislevy, Steinberg, & Almond, 2003). The term evidence-centered underscores the central idea that all assessments are a particular kind of evidentiary argument. The ECD phases provide structures for laying out an assessment argument, building design elements to embody the argument, and arranging the processes that instantiate it. This appendix provides a brief overview of the layers and principal models in the ECD framework.

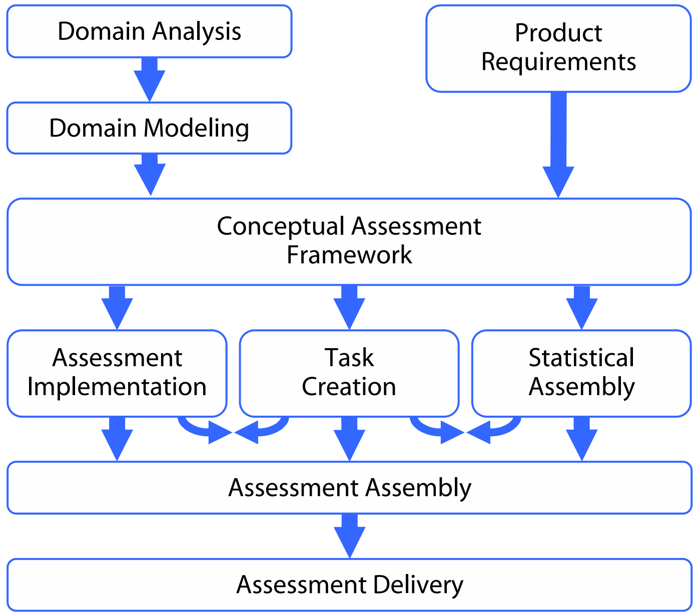
The intent of the ECD framework is to express at an abstract level the elements and relationships that an assessment needs to ground a coherent assessment argument. As a conceptual model, it can serve as the basis for a design tool in the form of an object model. Instantiating instances of the objects as needed for a particular assessment ensures that an assessment will have the functionality it needs and the components will work together. Such a framework promotes reusability of objects and processes. A key idea is that different kinds of objects are not defined for different kinds of tests; rather, the same general kinds of objects are tailored and assembled in different ways to meet different purposes.

The following section provides a brief overview of the layers in assessment design that are expressed in the ECD structures. Little more than names and purposes of layers are given here; the interested reader is referred to Mislevy, Steinberg, and Almond (2003) for details and rationale of the design phases, and to Almond, Steinberg, and Mislevy (2002) regarding delivery architecture. The overview is followed by sections that lay out the main elements of what amounts to a blueprint for assessment elements, or the *conceptual assessment framework* (CAF), and the main processes of the delivery system architecture.

1.0 An Overview of the ECD Layers

Figure A1 is a schematic of the ECD design and delivery framework. In the layer called *domain analysis*, designers analyze the domain from a number of perspectives, including cognitive research, available curricula, expert input, standards and current testing practices, test purposes, and various requirements, resources, and constraints to which the proposed product might be subject. They gather information from a variety of sources and identify concepts and relationships that can play roles in assessment arguments in the domain, such as aspects of knowledge, situations in which knowledge is used, knowledge representations, features of situations that make performing harder or easier, and key characteristics of performances.

Figure A1. Schematic of Design and Implementation



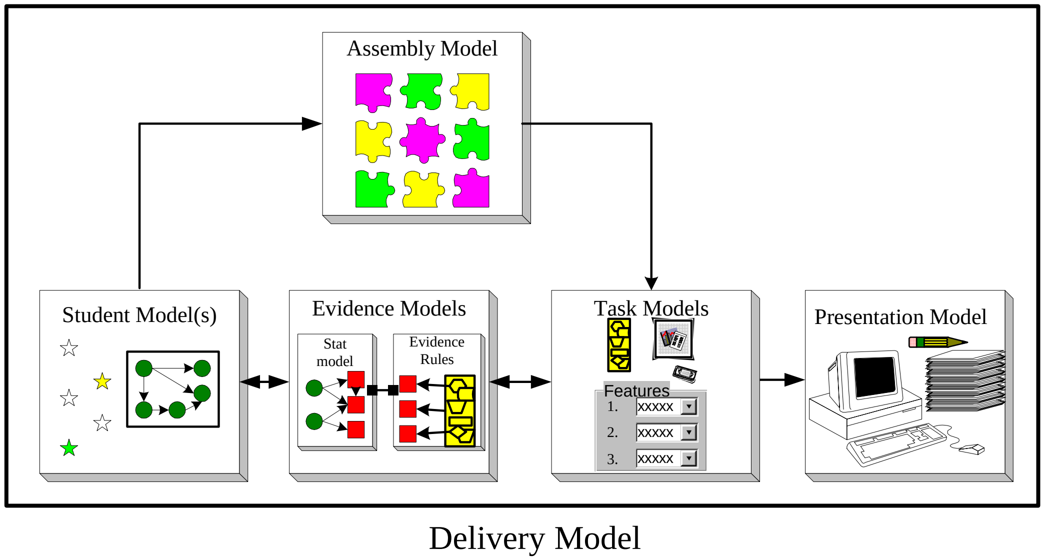
In the *domain modeling* phase, the designers use information from the domain analyses to establish relationships among proficiencies, tasks, and evidence. They explore different approaches and develop high-level sketches that are consistent with what they have learned about the domain so far. To convey these complex relationships, they can create graphic representations and schemas such as Toulmin diagrams for argument structures (Mislevy, 2003), PADI *design patterns* (Mislevy, et al, 2003), and BEAR construct maps (Wilson, 2004). They may develop prototypes to test their assumptions.

The *conceptual assessment framework* (detailed in the next section) is where more technical elements of an assessment are laid out, such as psychometric models, scoring rubrics, descriptions of stimulus materials, and administration conditions (e.g., affordances and properties of a simulation system with which an examinee will interact). The three smaller boxes in Figure A1 (*assessment implementation*, *task creation*, and *statistical assembly*) represent the activities of actually constructing all the necessary pieces, and the *assessment assembly* box represents coordinating and packaging the elements into a ready-to-run composite. The *assessment delivery* box (also see Figure A2) represents the operation of the implemented assessment.

2.0 *The Conceptual Assessment Framework*

The assessment argument is embodied in the *conceptual assessment framework*. The objects and specifications created here provide a blueprint for the operational aspects of work, including the creation of assessments, tasks, and statistical models, and the delivery and operation of the assessment. Figure A2 is a high-level schematic of the basic models in the *conceptual assessment framework* and objects they contain.

Figure A2: The Principal Design Objects of the Conceptual Assessment Framework (CAF)



The Student Model: What knowledge, skills, and 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. These are the terms in which we want to determine evaluations, make decisions, or plan instruction—but we don't get to see the values directly. We see instead what students say or do and must interpret what we see as evidence about these student model variables. The number and nature of student model variables in an assessment also depend on its purpose. A single variable characterizing overall proficiency might suffice in an assessment meant only to support a pass/fail decision. But a coached practice system to help students develop the same proficiency might require a finer-grained student model, to monitor how a student is doing on particular facets of skill and knowledge for which we can offer advice or suggest practice.

The student model in Figure A2 depicts student model variables as circles. The arrows represent important empirical or theoretical associations. We use a statistical model to manage our knowledge about a given student's unobservable values for these variables at any given point in time, expressing current knowledge as a probability distribution that can be updated in light of new evidence.

Evidence Models: What behaviors or performances should reveal those constructs, and what is the connection? An evidence model lays out the argument about why and how the observations in a given task situation constitute evidence about student model variables. Figure A2 shows that there are two parts to the evidence model. **The evaluation submodel, which contains evaluation (or evidence) rules,** concerns extracting the salient features of whatever the student says, does, or creates in the task situation—the "work product" represented by the jumble of shapes in the rectangle at the far right of the evidence model. It is a unique human production, perhaps as simple as a response to a multiple-choice item or as complex as repeated evaluation and treatment cycles in a patient-management problem. The three squares coming out of the work product represent "observable variables," evaluative summaries of whatever the designer has determined are the key aspects of the performance in light of the assessment's purpose. Evaluation rules map unique human actions into a common interpretive framework, effectively laying out the argument about what is important in a performance. These rules can be as simple as determining whether the response to a multiple-choice item is correct or as complex as an evaluation of multiple aspects of an unconstrained patient-management solution. There can be several stages of evaluation and synthesis. The rules can be automated, demand human judgment, or involve both in combination.

The measurement (or statistical) submodel of the evidence model expresses how the observable variables depend on student model variables. This is effectively the argument for synthesizing evidence across multiple tasks or from different performances. Figure A2 shows that the observables are modeled as depending on some subset of the student model variables. Familiar models from test theory, such as item response theory and latent class models, are examples of statistical models in which values of observed variables depend probabilistically on values of unobservable variables.

Task Models: What tasks or situations should elicit those behaviors? A task model provides a framework for constructing and describing the situations in which examinees act. Task model variables play many roles, including structuring task construction, focusing the evidentiary value of tasks, guiding assessment assembly, implicitly defining student model variables, and conditioning the statistical argument between observations and student model variables. A task model includes specifications for the environment in which the student will say, do, or produce something — for example, characteristics of stimulus material, instructions, help, tools, and affordances. Here is where the ECD design framework connects with research in theoretically based task construction (e.g., Embretson, 1998) and automated item generation (Irvine & Kyllonen, 2000). The task model also includes specifications for the work product, or the form in which what the student says, does, or produces will be captured.

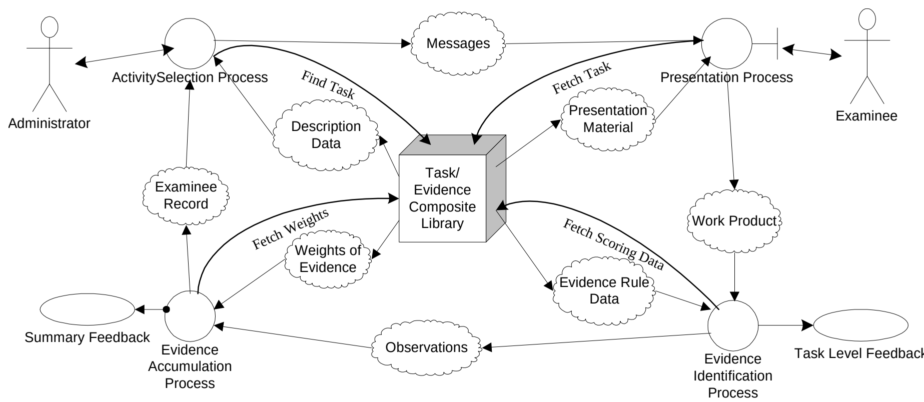
Also shown in Figure A2 are the Assembly Model, the Presentation Model, and the Delivery Model. The Assembly Model contains specifications for assembling individual tasks into a larger unit, such as an assessment or a subtest. This could be a table of specifications for a fixed test, constraints and optimizing targets for an adaptive test, or logic for the interaction of assessment and instruction in an instructional system. The Presentation Model contains requirements and specifications for the assessment's interaction with the examinee, such as hardware, software, and interface requirements for a computer-administered test. Considerations such as alternative presentation methods for tests taken by students needing accommodations are specified here. The Delivery Model provides requirements and specifications for the assessment as a whole, including protocols and mechanisms for the rest of the messages among delivery processes discussed in the next section.

3.0 Delivery System Architecture

Figure A3 sketches four principal processes that take place in an assessment. Some are compressed or implicit in familiar forms of assessment. Explicating them makes it easier to design reusable, interoperable components. This architecture for delivery, scoring, and reporting is compatible with the IMS (Instructional Management Systems) interoperability standards consortium's standards for assessment objects, or QTI (Question and Test Interoperability). In one simple way for these processes to interact, the Activity Selection

Process selects a task (or other activity) and instructs the Presentation Process to display it. When the examinee has finished interacting with the item, the Presentation Process sends the results (a Work Product) to the Evidence Identification Process. This process identifies key Observations about the results and passes them to the Evidence Accumulation Process, which updates the Examinee record. The Activity Selection Process then makes a decision about what to do next, based on the current beliefs about the examinee. Any pattern of interaction is possible, and quite different patterns can be required for applications such as intelligent tutoring systems, self-assessment, training drills, and multiple-stage investigations. This abstract design is open with regard to the means by which processes are implemented, their locations, and their sequence and timing (e.g., the interval between evidence identification and evidence accumulation could be measured in weeks or in milliseconds).

Figure A3: Processes and Messages in the Assessment Delivery Cycle



Ensuring that these processes interact coherently requires standards for the messages they must pass from one to another. The protocols for defining the forms and the contents of the messages in a given assessment—importantly, not the forms or the contents themselves—are specified in the evidence-centered object model. In this way, designing an assessment within the common evidence-centered framework ensures the coordination of operational processes. Analogously, fully specifying the assessment objects in the object model helps the assessment designer lay out specifications for task creation and statistical analyses.

References

Almond, R. G., Steinberg, L. S., & Mislevy, R. J. (2002). Enhancing the design and delivery of assessment systems: A four-process architecture. *Journal of Technology, Learning, and Assessment, 1*(5). Available from http://www.jtla.org.

Embretson, S. E. (1998). A cognitive design system approach to generating valid tests: Application to abstract reasoning. *Psychological Methods, 3*, 300-396.

Irvine, S. H., & Kyllonen, P. (Eds.). (2001). *Item generation for test development.* Mahwah, NJ: Lawrence Erlbaum Associates.

Mislevy, R. J., Chudowsky, N., Draney, K., Fried, R., Gaffney, T., Haertel, G., et al. (2003). Design patterns for assessing science inquiry (*PADI Technical Report No. 1)*. Menlo Park, CA: SRI International.

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

Mislevy, R. J., Steinberg, L. S., & Almond, R. G. (2003). On the structure of educational assessments. *Measurement: Interdisciplinary Research and Perspectives, 1*(1), 3-62.

Wilson, M. (2004). *Constructing measures: An item response modeling approach*. Mahwah, NJ: Lawrence Erlbaum Associates.

Appendix B

Definitions of the PADI Object Model

N.B.: Underlined text indicates other objects in the PADI Object Model.

Overview

PADI aims to provide a practical, theory-based approach to developing 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 various standards but difficult to assess.
Below is the PADI object model, a conceptual framework for representing complex assessment tasks. The PADI approach to standards-based assessment moves from statements of standards, through statements of the claims about the student capabilities 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, central science concepts 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 it when they see it.

Assessment Task

"A task is a goal-directed human activity to be pursued in a specified manner, context, or circumstance. Tasks may vary from relatively simple (e.g., responding to a multiple-choice item) to complex (e.g., conducting a symphony)" (Haertel & Wiley, 1993, p. 361).

In the PADI system, Assessment Tasks are generated by Task Specifications, which in turn are generated by Templates. Examples of Assessment Tasks are listed as Task Exemplars.

Activity

Activities constitute the major components of a task template and are used to structure the generation, collection, and scoring of evidence. An activity contains a group of related items, including presentation materials, work products, evaluation rules, observable variables, and measurement models. Activities can belong to multiple templates, and a template can have one or many activities; the decisions of whether to have several activities and how to define the scope of an activity are up to the assessment developer. For example, an inquiry task may have several distinct stages of investigation, and such stages act as a natural partitioning of the task into activities. As a second example, suppose a task consists of a list of "mix and match" subtasks, from which a student may choose; each of the integral subtasks could be cast as an activity within a single template. All activities within a template will update student model variables found within a single student model, as specified by the template.

Attributes of Activity:

1)	"**Measurement Models**" attributes are associations with (potentially shared) objects of type: Measurement Model.
2)	"**Evaluation Procedures**" attributes are associations with (potentially shared) objects of type: Evaluation Procedure (rubric).
3)	"**Work Products**" attributes are associations with (potentially shared) objects of type: Work Product.
4)	"**Materials and Presentation**" attributes are associations with (potentially shared) objects of type: Materials and Presentation.
5)	"**Presentation Logic**" attributes specify the order in which various materials should be presented and algorithmic logic that describes any desired looping or conditional presentation.

6) "**Task Model Variables**" attributes are associations with (potentially shared) objects of type: Task Model Variable.

7) "**Design Patterns**" attributes are associations with (potentially shared) objects of type: Design pattern.

8) "**Online Resources**" attributes are relevant items that can be found online (URLs).

9) "**References**" attributes are notes about relevant items, such as academic articles.

Continuous Zone

A zone or level in a continuous Student Model Variable that describes a distinct amount of ability, as judged by experts, usually with an empirical basis within a specific population of examinees. Each zone includes a lower and upper (minimum and maximum) cutoff value.

Attributes of Continuous Zone:

1) "**Minimum**" attributes specify the "cut-point" for the smallest value which should be included in this zone. For example, to create three zones in a range -4 to +4, the lowest zone would have a minimum of -4, the second zone would have a minimum of, say, -1, and the third zone might have a minimum of 1.

2) "**Advice to Next Level**" attributes offer advice to students about how to progress from this current level to the next one.

3) "**Maximum**" attributes specify the "cut-point" for the largest value that should be included in this zone. For example, to create three zones in a range -4 to +4, the highest zone would have a maximum of +4, the middle zone would have a maximum of, say, +1, and the lowest zone might have a maximum of -1. Note that the borders of zones may overlap, in which case the application must make a judgment. One common convention is to "round up" so that the higher zone gets any overlap.

Design Pattern

Design patterns are concepts that form a foundation for an assessment. The focus of design patterns is on the substance of the assessment argument rather than the technical details of operational elements and delivery systems. For example, some of the design patterns in PADI bridge knowledge about aspects of science inquiry with knowledge of the structures of a coherent assessment argument, in a format that guides task creation and assessment implementation.

Attributes of Design Pattern:

1) "**Focal Knowledge, Skills, and Abilities**" attributes are the primary knowledge/skill/abilities targeted by this design pattern.

2) "**Rationale**" attributes explain why this item is an important aspect of scientific inquiry and explicate the chain of reasoning connecting the inference of interest about student proficiency to potential observations and work products.

3) "**Additional Knowledge, Skills, and Abilities**" attributes are other knowledge/skills/abilities that may be required by this design pattern.

4) "**Potential Observations**" attributes are some possible things one could see students doing that would give evidence about the knowledge/skills/abilities.

5) "**Potential Work Products**" attributes are modes, like a written product or a spoken answer, in which students might produce evidence about knowledge/skills/abilities.

6) "**Potential Rubrics**" attributes are some evaluation techniques that may apply.

7) "**Characteristic Features**" attributes are aspects of assessment situations that are likely to evoke the desired evidence.

8) "**Variable Features**" attributes are aspects of assessment situations that can be varied in order to shift difficulty or focus.

9) "**I am a Kind of**" attributes are associations to other objects that are more abstract or more general than this object. For example, a dog is a specific kind of animal.

10) "**These are Kinds of Me**" attributes are associations with other objects that are more concrete or more specialized than this object. For example, animal is a general category that includes specific kinds of dogs.

12) "**These are Parts of Me**" attributes are associations with other objects that contain or subsume this one. For example, a windshield is a part of an automobile.

13) "**Educational Standards**" attributes are associations with (potentially shared) objects of type: <u>Educational Standard</u>.

14) "**Templates**" attributes are associations with (potentially shared) objects of type: <u>Template</u>.

15) "**Exemplar Tasks**" attributes are associations with (potentially shared) objects of type: <u>Task Exemplar</u>.

16) "**Online Resources**" attributes are relevant items that can be found online (URLs).

17) "**References**" attributes are notes about relevant items, such as academic articles.

Educational Standard

Educational standards are links to the most related educational standards, such as the National Science Education Standards (NSES).

Attributes of Educational Standard:

1) "**Online Resources**" attributes are relevant items that can be found online (URLs).
2) "**References**" attributes are notes about relevant items, such as academic articles.

Evaluation Phase

Evaluation phases are individual steps during an evaluation procedure.

Attributes of Evaluation Phase:

1) "**Preceding Evaluation Phase**" attributes are associations with objects that occur before this one and thereby feed into this one.
2) "**Work Products**" attributes are associations with (potentially shared) objects of type: <u>Work Product</u>.
3) "**Input Observable Variables**" attributes are intermediate observable variables that provide input to this phase. For example, a bundling phase might have inputs from previous phases in which raters evaluated different parts of an item.
4) "**Task Model Variables**" attributes are associations with (potentially shared) objects of type: <u>Task Model Variable</u>.
5) "**Evaluation Action Data**" attributes are data that assist with the Evaluation Action, e.g., a scoring key or other instructions to the scorer.

6) “**Evaluation Action**” attributes describe the algorithm, the actual steps, that should be used to convert work products into observable variables.

7) “**Output Observable Variables**” attributes are associations with (potentially shared) objects of type: Observable Variable.

8) “**Online Resources**” attributes are relevant items that can be found online (URLs).

9) “**References**” attributes are notes about relevant items, such as academic articles.

Evaluation Procedure (rubric)

Evaluation procedures (rubrics) are scoring schemes that turn students’ work products into observable variables (scores).

Attributes of Evaluation Procedure (rubric):

1) “**Evaluation Phases**” attributes are associations with (potentially shared) objects of type: Evaluation Phase.

3) “**Online Resources**” attributes are relevant items that can be found online (URLs).

4) “**References**” attributes are notes about relevant items, such as academic articles.

Materials and Presentation

Materials and Presentation specifications are requirements for the environment surrounding a student during an assessment, as well as things provided to the student. These materials are typically the stimuli for the tasks, such as the pictures, text, or other media that present a situation or problem to the student. Materials and Presentation specifications are abstract descriptions, complemented by concrete “settings,” stored outside the description. That is, a Materials and Presentation specification describes, but does not contain, the actual content (the text or image or whatever) of the material and/or presentation. That actual content is indicated outside the Materials and Presentation specification via a “setting” made in the Template or Task Specification. (Typically, an abstract Template has Materials and Presentations without settings, whereas a concrete Task specification includes both the description and the concrete settings.)

Attributes of Materials and Presentation:

1) “**Materials (MIME) Type**” attributes designate the kind of material, such as a picture on paper or an audio clip. Must be a MIME type. See http://www.iana.org/assignments/media-types/ for a list of established types. Please search well for an established type to describe your media. New entries can be created ad-hoc (but please follow MIME format).

2) “**Role of Stimulus**” attributes indicate whether this material is intended as a Directive, instructing students to do something, or intended as a Target, providing a model to match or emulate, or whether there is some other intended purpose for the material. “Non-directive” stimulus materials, such as charts, graphs, maps, tables, and pictures, present information that students can use in answering an assessment. There are various ways to use non-directive materials; their use is not highly specified.

3) “**Task Model Variables**” attributes are associations with (potentially shared) objects of type: Task Model Variable.

4) “**Online Resources**” attributes are relevant items that can be found online (URLs).

5) “**References**” attributes are notes about relevant items, such as academic articles.

Measurement Model

Measurement models handle associations between observable variables and student model variables. Each measurement model may associate with one or more student model variables but may associate with only one observable variable.

Attributes of Measurement Model:

1) "**Type of Measurement Model**" attributes indicate whether this measurement concerns Dichotomous (right/wrong) scoring, or Partial Credit scoring, or some other kind of scoring.
2) "**Observable Variable**" attributes are associations with (potentially shared) objects of type: Observable Variable.
3) "**Student Model Variables**" attributes are associations with (potentially shared) objects of type: Student Model Variable.
4) "**Scoring Matrix**" attributes refer to the MRCMLM scoring engine, wherein this matrix of values provides "loading" values about the weighting of a score (OV Category) with regard to all of the Student Model Variables (SMVs) in the measurement model. Each mapping entity is represented by a column of the matrix.
5) "**Design Matrix**" attributes refer to the MRCMLM scoring engine, wherein this matrix holds values that reflect the difficulty of moving from one score (OV Category) to another score. Each step-item entity is represented by a column of the matrix.
6) "**Calibration Parameters**" attributes refer to the MRCML scoring engine, wherein the parameters help tune the estimation of student proficiency by using data from previous experience with the measurement model. Provided here is a place to store these calibration parameters for a given population on a given examination. Keep in mind that such parameters should not be reused for a different assessment across a different population. An application system can override these values in that case, or use its own set of stored calibrations. The application system is ultimately responsible for associating the proper calibration. This storage of parameters in PADI is for convenience in using examples and should not be used indiscriminately.
7) "**Online Resources**" attributes are relevant items that can be found online (URLs).
8) "**References**" attributes are notes about relevant items, such as academic articles.

Observable Variable

Observable variables are the "scores" that result from an evaluation of a student's work product. Each observable variable is associated with exactly one measurement model.

Attributes of Observable Variable:

1) "**Categories (possible values)**" attributes are the possible scores for an observable variable.
2) "**Online Resources**" attributes are relevant items that can be found online (URLs).
3) "**References**" attributes are notes about relevant items, such as academic articles.

Student Model

Student models are collections of estimates of student proficiencies and contain one or more student model variables.

Attributes of Student Model:

1) "**Distribution Summary**" attributes are general descriptions of the statistical maps of proficiency estimates, in a form described by "Distribution Type".
2) "**Distribution Type**" attributes are the kind of probability model which is expected to describe the values for the Student Model Variables (SMVs) contained by this student model. The distribution can be Univariate Normal (a normal statistical distribution for a single variable), Multivariate Normal (normal distribution for multiple variables), or another kind of distribution.
3) "**Covariance Matrix**" attributes are measures of the amount of dependency between Student Model Variables. A cell value of 1 indicates that the two variables (the row and column variables for this matrix cell) are completely dependent — the two variables vary perfectly in tandem. In contrast, a cell value of 0 indicates that two variables are completely independent — the two variables have no relation.
4) "**Means Matrix**" attributes are median values, considering all the students' values within the distribution of Student Model Variables within this Student Model.
5) "**Student Model Variables**" attributes are associations with (potentially shared) objects of type: Student Model Variable.
6) "**I am a Kind of**" attributes are associations with other objects that are more abstract or more general than this object. For example, a dog is a specific kind of animal.
7) "**These are Kinds of Me**" attributes are associations with other objects that are more concrete or more specialized than this object. For example, animal is a general category that includes specific kinds of dogs.
9) "**These are Parts of Me**" attributes are associations with other objects that contain or subsume this one. For example, a windshield is a part of an automobile.
10) "**Online Resources**" attributes are relevant items that can be found online (URLs).
11) "**References**" attributes are notes about relevant items, such as academic articles.

Student Model Variable

Student model variables are individual estimates of one facet of student proficiencies. A student model variable is a part of at least one, and possibly more than one, student model.

Attributes of Student Model Variable:

1) "**Type of Student Model Variable**" attributes describe whether the variable may take any continuous value between its endpoints (e.g., 3.156) or whether it is restricted to taking only a finite number of values (e.g., only 1, 2, 3, or 4).
2) "**Minimum**" attributes specify the lowest value possible.
3) "**Maximum**" attributes specify the highest value possible.
4) "**Finite Categories**" attributes distinguish levels of distinct ability for this finite student model variable. This SMV cannot have a fractional value like "1.35"; instead, this SMV must be set to one of the finite values specified here.

5) “**Continuous Zones**” attributes summarize a group of values within the full range of the SMV. For example, a variable for ability to lift weights might define zones with cutoff levels of 0 to 25% of body weight, 25% to 50%, etc. Typically, zone cutoffs (zone minimum and zone maximum) are determined empirically by the distribution of scores in some calibrated population.

6) “**Educational Standards**” attributes identify relevant Educational Standard(s). In other words, this Student Model Variable measures similar abilities as do the associated Educational Standard(s).

7) “**Online Resources**” attributes are items that pertain and can be found online (URLs).

8) “**References**” attributes are notes about relevant items, such as academic articles.

Task Exemplar

Samples of actual tasks; these assessments may be suitable as models.

Attributes of Task Exemplar:

1) “**Online Resources**” attributes are relevant items that can be found online (URLs).

2) “**References**” attributes are notes about relevant items, such as academic articles.

Task Model Variable

Task model variables are conditions in the assessment and its environment that are caused to vary, or vary because of the student, and thereby affect the assessment in a significant way. A task model variable can represent a decision that an assessment designer makes before deploying an assessment, like the difficulty level of an item, which may be adjustable to a given audience. Alternatively, if a student or the assessment environment changes the outcome of a work product such that the evaluation must adapt, that is considered a “runtime” task model variable. Task model variables are abstract descriptions, complemented by concrete “settings,” stored outside the description. That is, a Task model variable describes but does not contain the actual content (e.g., the decision between choices) of the variable. That actual concrete choice is indicated outside the Task model variable in a “setting” made in the Template or Task Specification. (Typically, an abstract Template describes things like Task model variables in general terms, while a concrete Task specification includes both the description and the concrete setting.)

Attributes of Task Model Variable:

1) “**TMV Type**” attributes specify the values that may be used for the task model variable. Often, the designer supplies a set of discrete choices, represented in the menu. Or the variable may be allowed to have any free-form entry. Yet another type of TMV is determined at runtime, according to the behavior of the student or other environmental factors.

2) “**TMV Category (possible value**)” attributes specify discrete values, suitable for putting in a menu, that the variable may take. Categories are appropriate only for a TMV of type: Discrete, Menu-Chosen.

3) “**I am a Kind of**” attributes are associations with other objects that are more abstract or more general than this object. For example, a dog is a specific kind of animal.

4) “**These are Kinds of Me**” attributes are associations with other objects that are more concrete or more specialized than this object. For example, animal is a general category that includes specific kinds of dogs.

5) “**Online Resources**” attributes are relevant items that can be found online (URLs).

6) “**References**” attributes are notes about relevant items, such as academic articles.

Task Specification

A Task specification is a "blueprint" for creating an Assessment Task. Task Specifications are the final, most concrete form of Templates. When every variable in a Template is decided and specified for a particular assessment, the Template becomes a Task Specification.

Template

Templates are blueprints for assessment tasks that combine task environment information with evidence evaluation logic; templates are also known as "task-evidence shells." Templates can vary from abstract, general ideas to concrete specifications, ready to generate assessments. A template generally retains some flexibility, some unspecified aspects, such as Task Model Variables that have not been specified yet. When every variable in a template is decided and specified for a particular assessment, the template becomes a Task Specification, something that is ready for use in generating assessments.

Attributes of Template:

1)	"**Type**" attributes indicate whether the object is a finished, complete, concrete Task specification or an abstract and general Template.
2)	"**Student Model Summary**" attributes describe the student models in the template in outline form.
3)	"**Student Models**" attributes are associations with (potentially shared) objects of type: Student Model.
4)	"**Measurement Model Summary**" attributes describe an outline of the requirements for measurement models.
5)	"**Evaluation Procedures Summary**" attributes describe a general outline of requirements for evaluation procedures.
6)	"**Work Product Summary**" attributes describe an outline of the things created by the student.
7)	"**Task Model Variable Summary**" attributes describe an outline of all the task model variables that are used by this template.
8)	"**Task Model Variable Settings**" attributes are the exact choices made from among those allowed for each task model variable (TMV). In other words, the designer has specified a given task model variable, and it is no longer variable. The template is "pinned" to use this setting. Settings apply to the template/TMV combination. The same TMV may have different settings in different templates if it is associated with more than one template. Templates may also have associated Activities, and these Activities may have associated TMVs, but any setting for an "activity" TMV is still controlled by the template. Settings apply to the template, not to individual Activities, even though a TMV may show up under the Activity only.
9)	"**Presentation Environment Requirements**" attributes specify how the stimuli are presented to the student and any large-scale needs, like having a large room.
10)	"**Materials and Presentation Settings**" attributes are the exact choices made from among those allowed for each Materials and Presentation (M&P) item. In other words, the designer has specified a given Materials and Presentation choice, and it is no longer variable. The template is "pinned" to use this setting. Settings apply to the template/M&P combination. The same M&P may have different settings in different templates if it is associated with more than one template. Templates may also have associated Activities, and these Activities may have associated M&Ps, but any setting for an "activity" M&P is still controlled by the template. Settings apply to the template, not to individual Activities, even though an M&P may show up under the Activity only.
11)	"**Activities Summary**" attributes are an overview of all the activities included.
12)	"**Activities**" attributes are associations with (potentially shared) objects of type: Activity.

13) "**Template-Level Task Model Variables**" attributes are associations with (potentially shared) objects of type: Task Model Variable.

14) "**Tools for Examinee**" attributes are things provided to or permitted for use by the examinee.

15) "**Exemplars**" attributes are associations with (potentially shared) objects of type: Task Exemplar.

16) "**Educational Standards**" attributes are associations with (potentially shared) objects of type: Educational Standard.

17) "**Design Patterns**" attributes are associations with (potentially shared) objects of type: Design Pattern.

18) "**I am a Kind of**" attributes are associations with other objects that are more abstract or more general than this object. For example, a dog is a specific kind of animal.

19) "**These are Kinds of Me**" attributes are associations with other objects that are more concrete or more specialized than this object. For example, animal is a general category that includes specific kinds of dogs.

20) "**These are Parts of Me**" attributes are associations with other objects that contain or subsume this one. For example, a windshield is a part of an automobile.

21) "**Online Resources**" attributes are relevant items that can be found online (URLs).

22) "**References**" attributes are notes about relevant items, such as academic articles.

Work Product

Work products are the actual things created by the student during the assessment.

Attributes of Work Product:

1) "**Product Type**" attributes describe the kind of thing produced by the student's labor. For example, the work product may be a kind of menu choice, an audio transcript, or an essay on paper.

2) "**Examples**" attributes include references to online samples (URLs) or pictures or concrete, actual text produced by student labor.

3) "**Online Resources**" attributes are relevant items that can be found online (URLs).

4) "**References**" attributes are notes about relevant items, such as academic articles.

Relation: "Associated"

"**Associated**" are associations with other objects.

Relation: "Educational Standards"

"**Educational Standards**" are associations with the goals and specifications written by national educational councils and other standard-setting bodies.

Relation: "Exemplar"

"**Exemplar**" are associations with objects that exemplify this one.

Relation: "Exemplar Task"

"**Exemplar Task**" are associations with Tasks that exemplify this one.

Relation: "I am a Kind of"

"**I am a Kind of**" are associations with other objects that are more abstract or more general than this object. For example, a dog is a specific kind of animal.

Relation: "I am a Part of"

"**I am a Part of**" are associations with other objects that are components or steps within this one. For example, an automobile contains a windshield.

Relation: "Precedes and Feeds into Me"

"**Precedes and Feeds into Me**" are associations with objects that occur before this one, and thereby feed into, this one.

Relation: "Templates"

"**Templates**" are associations with Templates that fit this Design pattern.

Relation: "These are Kinds of Me"

"**These are Kinds of Me**" are associations with other objects that are more concrete or more specialized than this object. For example, animal is a general category that includes specific kinds of dogs.

Relation: "These are Parts of Me"

"**These are Parts of Me**" are associations with other objects that contain or subsume this one. For example, a windshield is a part of an automobile.

Stimulus: Directive

A directive provides a goal, instructing the examinee to act in some way.

Stimulus: Hint or Cue

A hint or cue provides some small assistance.

Stimulus: Manipulable (artifact)

A manipulable provides some concrete thing that can be inspected (for example, a frog in a dissection exam).

Stimulus: Non-Directive

A non-directive provides information that examinees can use in answering the assessment. Examples of non-directive stimulus materials are charts, graphs, maps, tables, and pictures.

Stimulus: Selection

A selection provides several stimuli that serve as alternative problem contexts or sources of information for an assessment. Respondents may select one or more of these when they are solving a problem.

Stimulus: Target (model for matching or emulation)

A target provides a model for matching or emulation.

Reference

Haertel, E. H., & Wiley, D. E. (1993). Representations of ability structures: Implications for testing. In N. Frederiksen, R. J. Mislevy, & I. I. Bejar (Eds.), *Test theory for a new generation of tests* (pp. 359-384). Hillsdale, NJ: Erlbaum.

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*