



Large-Scale Assessment Technical Report 1 | May 2009

PADI Online Assessment Design System and Minnesota Science Assessment Glossary of Terms



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

Daniel Zalles, SRI International
Larry Hamel, Codeguild, Inc.
Dennis Fulkerson, Pearson

Report Series Published by SRI International





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

Technical Report Series Editors

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

Copyright © 2009 SRI International. All Rights Reserved.

APPLICATION OF EVIDENCE-CENTERED DESIGN TO STATE
LARGE-SCALE SCIENCE ASSESSMENT
TECHNICAL REPORT 1

PADI Online Assessment Design System and Minnesota Science Assessment Glossary of Terms

Prepared by:
Daniel Zalles, SRI International
Larry Hamel, Codeguild, Inc.
Dennis Fulkerson, Pearson

Abstract

This paper presents a Glossary of terms about features of the Application of Evidence-Centered Design to State Large-Scale Science Assessment project, which is partnering with the state of Minnesota. Each term is used in either the PADI (Principled Assessment Designs for Inquiry) design system or in the Minnesota science assessment and is classified as such in the Glossary. The PADI terms cover broad foundational characteristics of the PADI system as well as actual tools embedded in the system. When appropriate, definitions of the components of specific tools are included. The Minnesota assessment terms cover the assessment's broad attributes and alignments to the Minnesota State standards. The glossary will be available in an electronic database that permits the user to browse terms alphabetically or by their affiliation to the design system or to the Minnesota assessment.

PADI Online Assessment Design System and Minnesota Science Assessment Glossary of Terms

Activity (*PADI Design System*)

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. Consider one activity for any single stimulus, where a stimulus is, for example, one paper test. If there is any chance of conditional dependence among responses, all the responses should be part of the same activity, to allow for bundling as necessary. 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 left 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:

- 1) "**Design Patterns**" attributes are associations with (potentially shared) objects of type: Design Pattern
- 2) "**Evaluation Procedures**" attributes are associations with (potentially shared) objects of type: Evaluation Procedure (rubric).
- 3) "**Materials and Presentation**" attributes are associations with (potentially shared) objects of type: Materials and Presentation.
- 4) "**Measurement Models**" attributes are associations with (potentially shared) objects of type: Measurement Model.
- 5) "**Online resources**" attributes are links to online demos, research reports, press releases, and other relevant resources on the World Wide Web.
- 6) "**Presentation Logic**" attributes specify the order in which various materials should be presented and algorithmic logic that describes any desired looping or conditional presentation.
- 7) "**References**" attributes are notes about relevant items, such as academic articles.

- 8) "**Task Model Variables**" attributes are associations with (potentially shared) objects of type: Task Model Variable.
- 9) "**Work Products**" attributes are associations with (potentially shared) objects of type: Work Product.

Assessment Argument (*PADI Design System*)

Evidence-centered design views an assessment as an evidentiary argument: An argument from what we observe students say, do, or make in a few particular circumstances, to inferences about what they know, can do, or have accomplished more generally

Assessment Task (*PADI Design System*)

A task is a goal-directed human activity to be pursued in a specified manner, context, or circumstance. An assessment task is a type of task that is used to measure the learner's attainment of knowledge or skills. 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|#TASK_SPECIFICATION], which in turn are generated by Templates. For examples of assessment tasks, see Task Exemplars.

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

Benchmark (*Minnesota Science Assessment*)

The third and most specific division of content in the Minnesota science standards. Each test item is aligned to a benchmark.

Cognitive Level (*Minnesota Science Assessment*)

The level of cognitive complexity required to correctly respond to an item.

Minnesota recognizes 3 cognitive levels based on a modified version of Bloom's taxonomy. Each test item is assigned a cognitive level.

Cognitive Level A: Knowledge

Cognitive Level B: Understanding (Comprehension)

Cognitive Level C: Application, Analysis, Synthesis, and Evaluation

Conceptual Assessment Framework (*PADI Design System*)

A layer in the process of evidence-centered design that fleshes out the technical details of the assessment argument. The conceptual assessment framework is comprised of the student, evidence, and task models which are represented in the PADI object model.

Continuous Zone (*PADI Design System*)

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:

- 1) "**Advice to next level**" attributes offer advice to a student about how to progress from this current level to the next one.
- 2) "**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.
- 3) "**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.

Data Structures (see also Reusable data structures) (*PADI Design System*)

Data structures (such as schemas and object models) are ways of representing knowledge in terms of recurring concepts and relationships.

Design Pattern (*PADI Design System*)

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 the PADI Project 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:

- 1) "**Additional Knowledge, Skills, and Abilities**" attributes are other knowledge/skills/abilities that may be required by this design pattern.
- 2) "**Characteristic features**" attributes are aspects of assessment situations that are likely to evoke the desired evidence.
- 3) "**Educational standards**" attributes are associations with (potentially shared) objects of type: Educational Standard.
- 4) "**Exemplar tasks**" attributes are associations with (potentially shared) objects of type: Task Exemplar.
- 5) "**Focal Knowledge, Skills, and Abilities**" attributes are the primary knowledge/skill/abilities targeted by this design pattern.
- 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) "**Online resources**" attributes are relevant items that can be found online (URLs).
- 8) "**Potential observations**" attributes are some possible things one could see students doing that would give evidence about the knowledge/skills/abilities.
- 9) "**Potential rubrics**" attributes are some evaluation techniques that may apply.
- 10) "**Potential work products**" attributes are modes, like a written product or a spoken answer, in which students might produce evidence about knowledge/skills/abilities.
- 11) "**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.

- 12) "**References**" attributes are notes about relevant items, such as academic articles.
- 13) "**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.
- 14) "**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.
- 15) "**Templates**" attributes are associations with (potentially shared) objects of type: Template.
- 16) "**Variable features**" attributes are aspects of assessment situations that can be varied in order to shift difficulty or focus.

Drag-and-Drop (*Minnesota Science Assessment*)

A type of figural response item in which students respond by moving graphic elements from one location on the screen to another location.

Educational Standard (*PADI Design System*)

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

Attributes:

- 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.

Essential Benchmark (*Minnesota Science Assessment*)

A benchmark that must be assessed every year.

Evaluation Phase (*PADI Design System*)

Evaluation phases are individual steps during an evaluation procedure.

Attributes:

- 1) "**Evaluation Action Data**" attributes are data that assist with the Evaluation Action, e.g., a scoring key, or other instructions to the scorer
- 2) "**Evaluation Action**" attributes describe the algorithm, the actual steps, that should be used to convert work products into observable variables. In the case where there are Input OVs which should be bundled into an Output OV, a special editing panel can be involved to specify how to collapse or otherwise translate input categories into output categories.

- 3) "**Input Observable Variables**" attributes are intermediate OVs that provide input to this phase. For example, a bundling phase might have two inputs from previous phases that are combined to form the output OV. Note: the GradeBook application takes special action to process scores if the following conditions are met: 1) the MM has type 'bundle' 2) an evaluation phase has both input and output OVs 3) there is only one output OV In this case, the GradeBook will accept 'raw' scores provided for the input OV(s) (also called 'intermediate' OV(s)), and all permutations of these combinations will be mapped into the categories of the output (also called 'final') OV. Additionally, some of the input permutations may be 'collapsed' according to instructions on the Eval Phase Action. See the help text for that attribute for more information on how categories may be collapsed, which aids both calculation and interpretation of results.
- 4) "**Online resources**" attributes are relevant items that can be found online (URLs).
- 5) "**Output Observable Variables**" attributes are associations with (potentially shared) objects of type: Observable Variable.
- 6) "**Preceding Evaluation Phase**" attributes are associations with objects that occur before this one and feed into this one.
- 7) "**References**" attributes are notes about relevant items, such as academic articles.
- 8) "**Task Model Variables**" attributes are associations with (potentially shared) objects of type: Task Model Variable.
- 9) "**Work Products**" attributes are associations with (potentially shared) objects of type: Work Product.

Evaluation Procedure (rubric) (*PADI Design System*)

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

Attributes:

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

Evidence-Centered Design: PADI Instantiation of General Principles and Stages (*PADI Design System*)

Evidence-centered assessment design is a framework for organizing the activities and elements of assessment design in the service of an assessment argument. It moves in layers from the analysis of the targeted domain, to structuring an assessment argument, to planning then implementing the technical machinery of the assessment, to delivery processes.

Evidence-Centered Assessment Design	Purpose/Description of Stage	PADI Framework for Assessing Science Inquiry
I. <i>Domain Analysis</i>	<ul style="list-style-type: none"> ● Nature of knowledge, how people acquire it, how they use it. ● Definition of competence ● Development of competence/understanding ● Purpose of assessment 	<ul style="list-style-type: none"> ● Definition of Inquiry from standards documents ● Inquiry assessments used by curriculum developers and researchers ● Discussions with subject matter experts and review of literature on the development of inquiry
II. <i>Domain Modeling</i>	<ul style="list-style-type: none"> ● Systematic structure for organizing information gathered in <i>domain analysis</i> stage. ● Narrative description of proficiencies of interest, ways of getting observations that evidence proficiency, and ways of arranging situations in which students provide evidence of targeted proficiencies. 	<p><i>Design patterns</i>—narrative description of connections between inquiry standards and ways of obtaining evidence of what students know about inquiry.</p> <ul style="list-style-type: none"> ● Pointers to other relevant information (e.g., exemplar tasks, other <i>design patterns</i>, reference materials). ● Content and grade independent.
IIIA. <i>Conceptual Assessment Framework</i> <ul style="list-style-type: none"> ● Student Model ● Task Model ● Evidence Model <ul style="list-style-type: none"> — Evaluation — Measurement 	<ul style="list-style-type: none"> ● Expression of targeted knowledge as variables ● Identification of features of eliciting situations as variables in task schemas ● Identification and summary of evidence: <ul style="list-style-type: none"> — Task level scoring — Summary scoring 	<p><i>Templates</i>—detailed, technical description, blueprint, or specs for creating a family of tasks.</p> <ul style="list-style-type: none"> ● Specifies student and task model variables, rules for evaluating performance (e.g., rubrics), and psychometric

		measurement models.
IIIB. <i>Compilation</i> ● Task Creation ● Statistical Assembly ● Assessment Implementation	● Models for schema-based task authoring. ● Protocols for fitting and estimation of psychometric models. ● Strategies and algorithms for adaptive and nonadaptive test construction.	Outside the PADI project, with the exception of: ● <i>Exemplary tasks</i> produced by FOSS and BioKIDS partners in the PADI project ● Reference to the Berkeley Evaluation & Assessment Research Center's Item Calibration procedures for optional PADI scoring engine
IV. <i>Four-Process Delivery Architecture</i> ● Presentation ● Response Scoring ● Summary Scoring ● Activity Selection	● Data structures and processes for implementing assessments. ● Desire for interoperable processes and assessment objects	PADI object models promote design of assessment elements and processes to common IMS/SCORM standards. Optional PADI scoring engine available for users to incorporate in their assessment applications.

Extended Constructed Response (ECR) Item (*Minnesota Science Assessment*)

A type of item in which students must respond by typing an answer in one or more response boxes. ECR items limit student responses to 700 characters.

Figural Response (FR) Item (*Minnesota Science Assessment*)

A type of item in which students must respond by manipulating on-screen graphics. FR items include graphing items, hot spot items, and drag-and-drop items.

Graphing (*Minnesota Science Assessment*)

A type of FR item in which students respond by plotting points or manipulating bars on a graph.

Hot Spot (*Minnesota Science Assessment*)

A type of FR item in which students respond by selecting (i.e., "clicking on") one or more areas on a graphic.

MCA-II (*Minnesota Science Assessment*)

"Minnesota Comprehensive Assessments Series II," the title of the state's general assessment program.

MCA-II Science Assessment (*Minnesota Science Assessment*)

The MCA-II Science assessment is an online, innovative, scenario-based test assessing science content and skills at grades 5-8, and once in high school. Scenarios incorporate textual, visual, and audio components to present science content in simulated real-world contexts. Items are embedded within scenarios and may require students to manipulate on-screen graphics. Item types include multiple-choice (MC), figural response (FR), short constructed response (SCR), and extended constructed response (ECR).

Materials and Presentation (*PADI Design System*)

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:

- 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/>² 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) "**Online resources**" attributes are relevant items that can be found online (URLs).

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

4) "**Role of stimulus**" attributes indicate whether this material is intended as a Directive, instructing students to do something, or intended as a Directive, 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.

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

Measurement Model (*PADI Design System*)

Measurement models associate observable variables (OVs) with student model variables (SMVs) in order to describe how to change proficiency estimates as observations are obtained. Each measurement model may associate one or more student model variables with exactly one observable variable. (The rule for a single OV was established for computational simplicity within PADI.) Some of the attributes below have to do with a specific psychometric model, the Multidimensional Random Coefficient Multinomial Logit Model (MRCMLM; see http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&list_uids=9711023&dopt=Abstract). The MRCML Model is a Rasch-based item response model generalized for use with polytomous items to produce multivariate estimates. The scoring engine provided with PADI uses this model, so this Design System makes it convenient to provide information for MRCMLM. Other psychometric models can be somewhat accommodated by ignoring the MRCMLM attributes (Scoring, Design, and Calibration matrices). Alternatively, the PADI Design System is flexible, so the attributes themselves could be adapted as necessary.

Attributes:

1) "**Calibration Parameters**" attributes refer to the MRCMLM scoring engine, wherein the calibration parameters (typically thought of as item or

step difficulties, depending on the Type of Measurement Model) help tune the estimation of student proficiency by using data from previous experience with the measurement model. This attribute provides a place to store these calibration parameters for a given population on a given assessment instrument. Keep in mind that such parameters should not be reused for a different instrument across a different population. In any case, an application system (like the Gradebook) can override these values, or use its own set of stored calibrations when it combines student scores and assessment design information. The application system is ultimately responsible for using the proper calibration. This storage of calibration parameters in PADI is for convenience in examples, and should not be used indiscriminately.

2) "**Design Matrix**" attributes refer to the MRCMLM scoring engine, wherein this matrix holds values that reflect the difficulty in moving from one score (OV Category) to another score. Each category of the associated OV is represented by a row and each Calibration Parameter (see below) associated with the OV is represented by a column of the matrix.

3) "**Observable Variable**" attributes are associations with (potentially shared) objects of type: Observable 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 each of the Student Model Variables (SMVs) in the measurement model. Each category of the associated OV is represented by a row and each SMV is represented by a column.

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.

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

8) "**Type of Measurement Model**" attributes indicate whether this measurement concerns Dichotomous (right/wrong) scoring, Partial Credit scoring, or Rating Scale scoring. This definition is related to the way the Design Matrix (described below) is interpreted. (In this sense, the type of measurement model is constrained to be within the MRCML frame of reference.)

MIME (Multipurpose Internet Mail Extensions) *(General)*

An Internet standard that supports the format of a file sent over the Internet

Multiple-Choice (MC) Item *(Minnesota Science Assessment)*

A type of item in which students must respond by selecting the correct answer out of four answer options.

Narrative Structure *(Minnesota Science Assessment)*

The underlying narrative framework for a storyboard. There are currently 6 recognized narrative structures: General to specific or whole to parts, specific to general or parts to whole, investigation, topic with examples, change over time, and cause & effect.

Nonessential Benchmark *(Minnesota Science Assessment)*

A benchmark that must be assessed at least one time every four years.

Object Model *(PADI Design System)*

The Object Model of the PADI Assessment Design System provides 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 capabilities of students' 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.

Observable Variable (*PADI Design System*)

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:

- 1) "**Categories (possible values)**" attributes are the possible scores for an observable variable. By convention, list these from smallest at the top of the display, to largest at the bottom of the display, unless you have a special case. (Adhering to this ordering convention allows the Scoring and Design matrices to look more conventional too.)
- 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.

PADI Assessment Design System (*PADI Design System*)

The PADI Assessment Design System is an online system that supports the development of learning assessments in science education. "PADI" stands for Principled Assessment Designs in Inquiry. The purpose of the PADI System is 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.

Reusable Data Structures (*PADI Design System*)

Structures are reusable if either the form of the structure can be used in many circumstances, or the contents of the structure can be used in many circumstances. An example of the former is a PADI evidence model, with variables and conditional distributions that can be used to express the relationship between observable variables and student model variables for many tasks and in many projects. An example of the latter is a PADI design pattern, which, once created, helps task designers make design choices to build many tasks in the same or multiple assessment projects.

Rooted Items (*Minnesota Science Assessment*)

Items associated with and embedded in specific scenes. Rooted items can be either MC or FR types.

Scenario (*Minnesota Science Assessment*)

The context for a group of items. Scenarios may include text, visual, and audio, components. Scenarios consist of 3-5 scenes. Each scene supports 1-2 embedded (rooted) items.

Scene (*Minnesota Science Assessment*)

The basic unit of a storyboard or scenario. Storyboards/scenarios consist of 3-5 scenes. Scenes must support items.

Short Constructed Response (SCR) Item (*Minnesota Science Assessment*)

A type of item in which students must respond by typing an answer in one or more response boxes. SCR items limit student responses to 400 characters.

Storyboard/Scenario Type (*Minnesota Science Assessment*)

Storyboards/scenarios are classified as phenomenon, investigation, data analysis, or a combination of these.

Strand (*Minnesota Science Assessment*)

The broadest division of content in the Minnesota science standards. Science strands include History & Nature of Science (Strand I), Physical Science (Strand II), Earth & Space Science (Strand III), and Life Science (Strand IV). Each strand consists of multiple sub-strands.

Sub-strand (*Minnesota Science Assessment*)

The second division of content in the Minnesota science standards. Each sub-strand consists of multiple benchmarks.

Student Model (*PADI Design System*)

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

Attributes:

- 1) "**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.
- 2) "**Distribution Summary**" attributes are descriptions of the statistical maps of proficiency estimates. The form of the distribution is described by the next attribute, "Distribution Type". Please include information about the source of the distribution: if it is based on a population of students, mention the collection date and number of students, etc. If it is based on a theoretical distribution, name the theory. Otherwise, indicate that the distribution uses default, uncalibrated values.
- 3) "**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.
- 4) "**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.
- 5) "**Means Matrix**" attributes are median values, considering all the students' values within the distribution of Student Model Variables within this Student Model
- 6) "**Online resources**" attributes are relevant items that can be found online (URLs).
- 7) "**References**" attributes are notes about relevant items, such as academic articles.
- 8) "**Student Model Variables**" attributes are associations with (potentially shared) objects of type: Student Model Variable.
- 9) "**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.
- 10) "**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.

Student Model Variable (*PADI Design System*)

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:

- 1) "**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.
- 2) "**Educational Standards**" attributes identify like-minded Educational Standard(s). In other words, this Student Model Variable measures similar abilities as do the associated Educational Standard(s).
- 3) "**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.
- 4) "**Minimum**" attributes specify the lowest value possible.
- 5) "**Maximum**" attributes specify the highest value possible.
- 6) "**Online resources**" attributes are relevant items that can be found online (URLs).
- 7) "**References**" attributes are notes about relevant items, such as academic articles.
- 8) "**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 finite values (e.g., only 1,2,3 or 4).

Summary Items (*Minnesota Science Assessment*)

Items given at the end of a scenario. Summary items are usually broad, "big-picture" items and can be FR, SCR, or ECR types.

Task Exemplars (*PADI Design System*)

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

Attributes:

- 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 (*PADI Design System*)

Task model variables (TMVs) 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 TMV. Task model variables are abstract descriptions, complemented by concrete settings that are stored outside the TMV. That is, a TMV describes, but does not contain the decision between choices of the variable. Any concrete choice or setting is stored outside the TMV in a Template or Task Specification setting. See a Template sample and locate the 'Task model variables settings' attributes. Typically, an abstract Template describes Task model variables and has few settings, whereas a concrete Task Specification includes both the description and the concrete setting.

Attributes:

- 1) "**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.
- 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.
- 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) "**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".
- 6) "**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.

Task Specification (see also Template) (*PADI Design System*)

Task Specifications are fully specified blueprints for creating specific Assessment Tasks. 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 (*PADI Design System*)

Templates are blueprints for assessment tasks that combine task environment information with evidence evaluation logic. *Unlike task specifications, which are fully specified blueprints, templates are blueprints with slots to fill.* 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:

- 1) "**Activities**" attributes are associations with (potentially shared) objects of type: Activity.
- 2) "**Activities Summary**" attributes are an overview of all the activities included.
- 3) "**Design Patterns**" attributes are associations with (potentially shared) objects of type: Design Pattern.
- 4) "**Educational Standards**" attributes are associations with (potentially shared) objects of type: Educational Standard.
- 5) "**Evaluation Procedures Summary**" attributes describe a general outline of requirements for evaluation procedures.
- 6) "**Exemplars**" attributes are associations with (potentially shared) objects of type: Task Exemplar.
- 7) "**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.
- 8) "**Materials and Presentation Requirements**" attributes specify how the stimuli are presented to the student and any large-scale needs like having a large room.
- 9) "**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.

10) **Measurement Model Summary**" attributes describe the nature and requirements for measurement models used in this template. For example, one could mention whether the template requires a multidimensional model, or whether items have dependencies.

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

12) **Presentation Environment Requirements**" attributes specify how the stimuli are presented to the student and any large-scale needs, like having a large room.

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

14) **Student Models**" attributes are associations with (potentially shared) objects of type: Student Model.

15) **Student Model Summary**" attributes describe an outline of the student models in the template.

16) **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.

17) **Task Model Variable Summary**" attributes describe an outline of all the task model variables that are used by this template.

18) **Template-level Materials and Presentation**" attributes are associations with (potentially shared) objects of type: Materials and Presentation.

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

20) **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.

21) "**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.

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

23) "**Type**" attributes indicate whether the object is a finished, complete, concrete Task Specification or a Template which is abstract and general.

24) "**Work Product Summary**" attributes describe an outline of the things created by the student.

Work Product (*PADI Design System*)

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

Attributes:

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

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

3) "**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.

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

Relation Types (*PADI Design System*)

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 to other objects which 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 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 (*PADI Design System*)

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

1) Stimulus: **Hint or Cue**

A hint or cue provides some small assistance.

2) Stimulus: **Manipulable** (artifact)

A manipulable provides some concrete thing that can be inspected, like a frog in a dissection examination.

3) 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.

4) 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.

5) Stimulus: **Target** (model for matching or emulation)

A target provides a model for matching or emulation.



Sponsor

The National Science Foundation, Grant DRL - 0733172

Prime Grantee

SRI International. *Center for Technology in Learning*

Subgrantees

University of Maryland

Pearson

