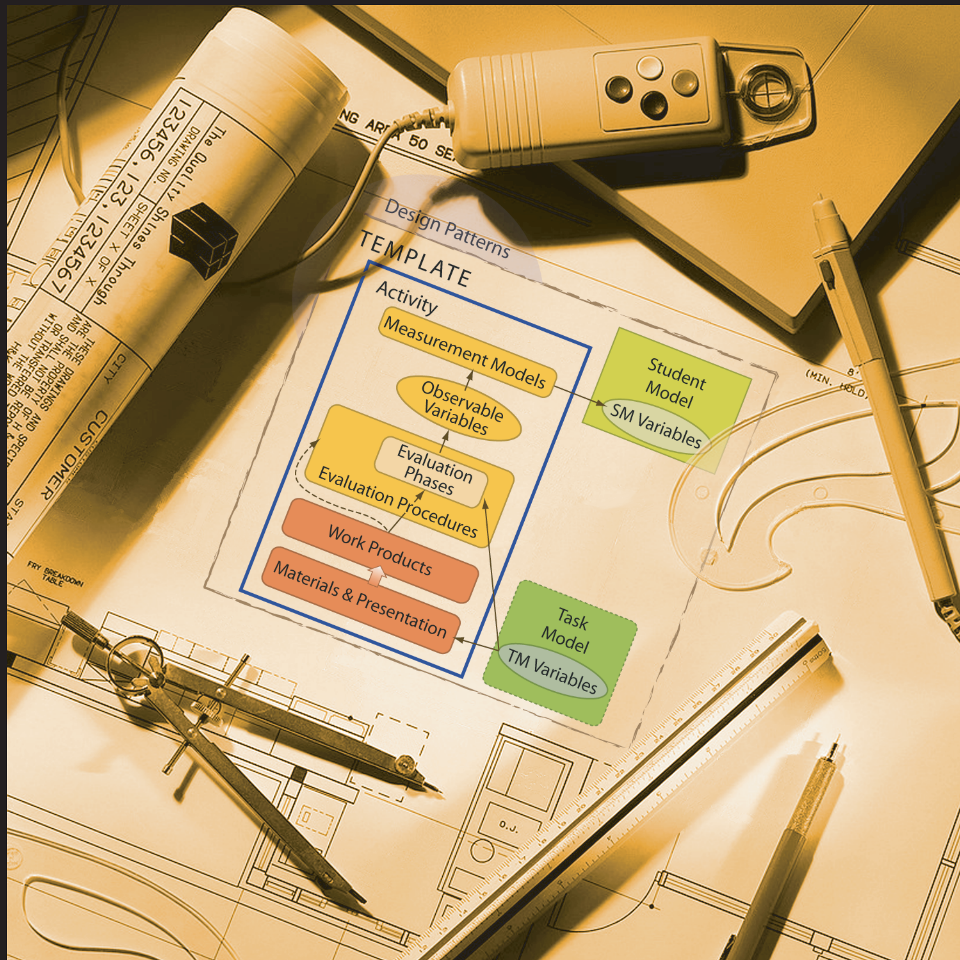


PADI Technical Report 8 | June 2005

An Example-Based Exploration of Design Patterns in Measurement

PADI | Principled Assessment Designs for Inquiry

Angela Haydel DeBarger, SRI International

Michelle M. Riconscente, University of Maryland

with **Alissa L. Morrison**, **Patricia Verdines-Arredondo**, University of Maryland
Joy Barnes, **Kia Johnson**, **René Lawless**, **Duanli Yan**, Educational Testing Service

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 8

An Example-Based Exploration of Design Patterns in Measurement

Prepared by:
Angela Haydel DeBarger, SRI International
Michelle M. Riconscente, University of Maryland

With:
Alissa Morrison and Patricia Verdines-Arredondo, University of Maryland
Joy Barnes, Kia Johnson, René Lawless, and Duanli Yan, Educational Testing Services

Acknowledgment
PADI is supported by the Interagency Educational Research Initiative (IERI) under grant REC-0129331 (PADI Implementation Grant).
Disclaimer
Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

CONTENTS

FIGURES

ABSTRACT

This paper extends the work conducted by the Principled Assessment Designs for Inquiry (PADI) project to investigate more deeply the application of assessment *design patterns*. By using examples of *design patterns* in several domains, such as science, mathematics, and studio art, this paper describes their role in assessment development. Three key benefits of *design patterns* are discussed and illustrated with the examples: (1) *design patterns* facilitate decision making about assessment design; (2) *design patterns* explicate the assessment argument; and (3) *design patterns* afford flexibility in usage for assessment design. In addition, the examples show how *design patterns* can vary in their generality, in their scale, and in the psychological perspective that they represent.

1.0 Introduction

Recent advances in cognitive psychology and learning, statistics, measurement, and technology can substantially enhance our ability to develop complex assessments of student learning. The Principled Assessment Designs for Inquiry (PADI) project is defining and implementing a set of structures to facilitate the orchestration of these areas of expertise in service of high quality operational assessments. PADI *design patterns* are schemas for organizing information about some aspect of knowledge in terms of assessment arguments.

The primary purpose of this paper is to describe the role that *design patterns* play in the assessment design process through a series of sample *design patterns* from several domains. These examples of *design patterns* apply to multiple content areas and assessment formats and thus illustrate the adaptability of the PADI system. As an introduction to these examples, we first sketch the work of PADI and the origin of *design patterns*. We then describe their role in assessment design and provide the rationale for their creation and use in assessment development. For additional information about *design patterns*, the reader is referred to the PADI Technical Report 1, *Design Patterns for Assessing Science Inquiry* (Mislevy et al., 2003)

2.0 Principled Assessment Designs for Inquiry (PADI)

The work of PADI is guided by an evidenced-centered design (ECD) framework (Mislevy, Steinberg, & Almond, 2003), which articulates the interrelationships among substantive arguments, assessment designs, and operational processes. ECD embodies a conception of assessment as reasoning from the particular things students say, do, or make to more broad inferences about what students can say, do, or make, as suggested by Messick (1994):

> 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. (p. 16)

PADI goes beyond Messick's description by identifying structures that capture commonalities across a set of problems or situations at different stages of the assessment process. These structures afford the design and implementation of assessments that may vary greatly in terms of surface features, but retain an underlying assessment argument that links the inference of interest to the evidence garnered in its support. Among these structures are *design patterns*, a term first used by architect Christopher Alexander (Alexander, Ishikawa, & Silverstein, 1977). Alexander et al. (1977) stressed the importance of structures that emerge naturally through population growth, such as health centers, accessible greens, roads and paths, and formalized a way of describing these patterns of building and community designs in a "pattern language."

3.0 Design Patterns

A *design pattern* addresses both a problem that occurs repeatedly in the environment and the core of the solution to that problem—but at a level of generality in which the solution can be applied repeatedly without being the same in its particulars. For example, in architecture, the *design pattern* perspective can be applied to the structure of a city (patterns for a park and a transportation center), a building (patterns for a museum or a restaurant), or a single room (the schema of a work triangle for a kitchen). More recently, software designers have crafted *design patterns* to develop sophisticated software applications based on commonalities across development processes, as well as across software packages themselves (e.g., a *design pattern* for an “object generator”).

PADI uses *design patterns* as a schema or structure for conceptualizing the components of assessment arguments and their interrelationships. The role of *design patterns* is to rationalize the assessment argument by identifying in narrative form the student knowledge, skills and abilities (KSAs), potential observations, work products and rubrics that test designers may want to use, as well as characteristics and variable features of potential assessment tasks.

The rationale for the use of *design patterns* in assessment starts from a need to extend thinking from individual assessment tasks to prototypical ways of obtaining evidence about the acquisition of various aspects of knowledge. Thinking through a task at the level of a *design pattern* grounds the subsequent detailing of the operational elements, or “nuts and bolts”, of assessments, such as psychometric models, evaluation procedures, and specific stimulus materials. In addition to supporting the identification of aspects of knowledge that are similar across content areas or skill levels, this approach affords the identification of reusable schemas for obtaining evidence about such knowledge. Further, echoing Alexander et al. (1977), *design pattern* structures are considered useful in the discussion of and planning for assessments among both content experts and measurement experts, rather than representing top-down conceptualizations imposed on the actual processes and kinds of information that contribute to the creation of assessments. The structure and content of *design patterns* are expected to emerge naturally from assessment development processes as they evolve.

Thinking at the level of *design patterns* is integral in the development of complex assessments because it enables content and measurement experts to share a coherent view about the substantive argument needed to create a principled assessment. *Design pattern*s can advance the current approaches used by subject area specialists in designing assessments by incorporating the advances in cognitive psychology and learning, statistics, measurement, and technology. Subject matter specialists using such *design patterns* gain access to these newer approaches to assessment rather than being constrained to familiar item formats and simple measurement models.

As they have evolved in PADI, *design patterns* are essentially a set of attributes that, when completed, represent an assessment argument. Each *design pattern* has a Title, Summary, and Rationale, which provide an overview of the target inferences addressed by this *design pattern*, as well as a rationale for using certain kinds of information about student performance as evidence of the targeted Focal and Additional KSAs. Table 1 contains an exhaustive set of *design pattern* attributes with brief descriptions of each.

Table 1. Attributes of a PADI Assessment Design Pattern

Attribute	Description	Comments
Title	A short name for referring to the design pattern.	
Summary	An overview of the kinds of assessment situations students encounter in tasks that are instantiations of this design pattern and what one wants to know about students' knowledge, skills, and abilities (KSAs).	
Rationale	Explanation why this item is an important aspect of scientific inquiry.	
Focal KSAs	The primary KSAs targeted by this design pattern.	
Additional KSAs	Other KSAs that may be required by this design pattern.	These could be nuisance skills, for example, background knowledge the student must be provide, or knowledge intended to be assessed jointly with the focal KSAs. Additional KSAs make assessment designesr aware that other KSAs beside the focal one are often addressed by an assessment task and that determining which ones to include is a design choice that should be made purposefully.
Potential observations	Some possible things students do that would give observable evidence about the KSAs.	Potential observations differ from work products (below) in that work products are what students produce, while observations are qualities that assessors discern and evaluate in work products.
Potential work products	Modes, like a written product or a spoken answer, in which students might produce evidence about KSAs.	
Potential rubrics	Some evaluation techniques that may apply.	These may include links to relevant scoring rubrics and procedures (algorithms, guidelines, and/or examples of ways to ascertain values of observations from student work products).
Characteristic features	Aspects of assessment situations that are likely to evoke the desired evidence.	These are features of situations (tasks) that are required so that students can provide evidence of the KSAs of interest. If a focal KSA is problem-solving with algebraic representations in ill-structured problems, then a characteristic feature of tasks to assess this KSA would be that the situation must present a problem that is amenable to algebraic representation and solution—possibly several different ones—but the approach and the representation must be developed by the student rather than provided by the assessor.

Table 1. Attributes of a PADI Assessment Design Pattern *(continued)*

Attribute	Description	Comments
Variable features	Aspects of assessment situations that can be varied in order to shift difficulty or focus.	Given that all the tasks that might be generated from a given design pattern are alike at some level in terms of characteristic features, variable features specify ways in which they might vary to increase or decrease difficulty, focus of information, put more or less demand on various additional KSAs, etc.
I am a kind of	Associations to other objects ("my parents") which are more abstract or more general than this object.	
These are kinds of me	Associations to other objects ("my children") which are more concrete or more specialized than this object.	
These are parts of me	Associations to other objects that contain or subsume this one. For example, an automobile contains a windshield.	
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.	These may include links to sample assessment tasks that are instances of this design pattern.
Online resources	Relevant items that can be found online (URLs).	These items may illustrate or provide background for this design pattern.
References	Notes about relevant items, such as academic articles.	

A variety of approaches can be used to create a *design pattern*. One is to start from an existing assessment and work backwards to extract a more general *design pattern* that may be used to generate similar kinds of assessments. Another strategy consists of beginning with a set of learning outcomes to be included in the Student Model and proceeding from there to identify appropriate Potential Observations, Work Products, and Rubrics. As we become involved in creating tasks that provide a context for eliciting those learning outcomes (and later as we field test the assessment with students), we often develop new insights into their nature and limitations. These insights may in turn lead to modifications of the KSAs, Potential Observations, Work Products, or other *design pattern* attributes. Because all *design pattern* attributes are related, the creation of *design patterns* is an iterative process that involves cycling through all of the attributes (perhaps multiple times) to ensure that they cohere.

4.0 Examples of Design Patterns

The assessment *design pattern* (DP) examples presented here were created by graduate students in a course on cognitive psychology and assessment taught by Robert Mislevy at the University of Maryland in Fall 2003. As part of their course work, students were charged with the task of analyzing an existing assessment of their choice (e.g., National Assessment of Educational Progress [NAEP], university degree program portfolio system) through the perspective of *design patterns*. Students were invited to contribute their *design patterns* to this technical report. The examples that follow are presented as submitted by the students, with minor editing. These *design patterns*, largely reverse-engineered from previously-existing tasks, reflect a range of domains (e.g., mathematics, science, art) and assessment formats and thus vary accordingly in their detail, focus, and formatting.

Each *design pattern* is presented with a brief overview. In the final section on "Benefits of Design Patterns," we provide integrative comments and discussion to tie these examples together. Although PADI is focused on science inquiry, *design patterns* are not limited to this area; the range of examples below demonstrates how characteristics of *design patterns* are viable across domains and purposes. As with the science inquiry *design patterns* crafted in PADI, the present examples illustrate how the use of *design patterns* supports the clear articulation of the assessment argument and facilitates subsequent development of assessment tasks capable of providing the necessary evidence to inform the claims of interest regarding student KSAs.

List of *design pattern* examples:

- DP-1: Reflective Assessment in BioKIDS (Alissa Morrison)
- DP-2: Understand and Apply Rate, Time, and Distance Concepts in Word Problem (Duanli Yan)
- DP-3: Scientific Investigation—Establishing Experimental Controls (Joy Barnes)
- DP-4: Selected Math Items from the Standardized Achievement Test (SAT) I (Kia Johnson)
- DPs-5A, 5B, 5C: Portfolios for Performance Assessments (Patricia Verdines-Arredondo)
- DPs-6A, 6B, 6C, 6D: AP Studio Art Portfolio and Other Hypothetical Design Patterns (Michelle Riconscente)
- DP-7: Reasoning to Maximize the Difference Between Two Numbers for NAEP Grade 8 Mathematics (René Lawless)

4.1 DP-1: Reflective Assessment in BioKIDS (by Alissa Morrison)

A major goal of the BioKIDS: Kids' Inquiry of Diverse Species project (Songer & Wenk, 2003) is to support and study student learning of complex science data. Building on the rich foundation of research on how children learn (i.e., Bransford, Brown, and Cocking, 2000), the pedagogy of poverty in urban classrooms (Haberman, 1991), and how scientists study real-time events, the BioKIDS project is developing, testing, and organizing inquiry-focused, technology-rich science programs in biodiversity and other content areas to span from Grades 5 through 8.

One important development in science education has been the recognition that students' lack of ability to reflect on their own progress and to look critically at their thinking, reasoning, and final work product is partly responsible for their inability to grasp complicated scientific content knowledge. White and Frederiksen (1998), in their attempt to develop an instructional theory and materials to make science inquiry available to a wide range of students, emphasize the importance of developing metacognitive knowledge and skills. They describe a process of Reflective Assessment in which students reflect on their own and other's inquiry processes. This self and group reflection process is also an important component of the BioKIDS project. For example, when students are asked to swap their data sheets with other groups and critique their work, they are not simply looking for mistakes or errors; students are evaluating what good work should look like and at the same time developing and reinforcing such habits in their own work. A *design pattern* that identifies the key attributes of reflective assessment at a more general level would be an important tool for task development in the BioKIDS curriculum. Such a *design pattern* is sketched below. Although the *design pattern* is based on research in science inquiry, the pattern for reflective assessment could be applied to other subject areas as well.

DP-1: Reflective Assessment in BioKIDS

Attribute	Value(s)	Comments
Title	Reflective assessment	
Summary	Students are introduced to a process in which they learn to evaluate and assess their own and each other's research methods.	
Rationale	Reflective self-assessment helps students to develop the ability to simultaneously monitor and improve their own learning as well as acquire the subject matter. Additionally, understanding the criteria by which their work will be evaluated enables students to better understand the characteristics of good performance.	Reflective assessment directs learning as students begin to think more carefully about the qualities to strive for in a performance or product.
Focal KSAs	▪ Metacognitive skills	▪ Learning to monitor the quality of one's thought and the product of one's effort. The implicit overall goal is teaching how to think about thinking. The metacognitive skills should compliment each other and be applicable to a wide range of cognitive contexts.

DP-1: Reflective Assessment in BioKIDS *(continued)*

Attribute	Value(s)	Comments
Focal KSAs (continued)	▪ Understand instructional objectives	▪ Reflecting on what they have learned raises new questions.
	▪ Recognize the progress being made toward these objectives	▪ A critique of the process itself. Students can be given the means to understand how to do well in their performances.
	▪ Diagnose particular strengths and weaknesses	▪ Reflective assessment makes students aware of the strengths and weaknesses of their current system or model. Self-evaluation encourages continual change and improvement, thereby discouraging unexamined models and ideas.
Additional KSAs	▪ Self-awareness	▪ Often the simple task of rating oneself can lead to reflection about what one really knows or can do and what areas are in need of improvement or better understanding.
	▪ Communication and collaboration	▪ May or may not be required, depending on whether the designer wants to encompass collaborative activity around reflective assessment.
	▪ Subject area knowledge	▪ Some tasks may require a strong knowledge of the subject area, as understanding one's performance in that domain may not be measurable outside of the metacognitive skills.
Potential observations	▪ Explanation of rationale of process.	▪ For instance, a student explaining what s/he is doing when assessing own or group's products or performance.
	▪ Identification of next step in a thinking cycle	
	▪ Recognizing and resolving contradictions between one's own and a standard work product	
	▪ Applying generally stated qualities of a rubric to the specifics of own or group's work	▪ For instance, being able to map one's own work into the framework of evaluation.
Potential work products	▪ Self-assessment questionnaires	▪ Designed to be completed by the student to assess performance on a certain task.

DP-1: Reflective Assessment in BioKIDS *(continued)*

Attribute	Value(s)	Comments
Potential work products *(continued)*	▪ Critiquing a flawed experiment/project	▪ For instance, practicing reflective-assessment skills with work other than one's own, as a precursor to evaluating one's own work
	▪ Critique of audio or video recordings/ transcripts of own or group's work	▪ Allows the student to record a sample of behavior for subsequent self-analysis, off-line of having to do it while doing the work. Can be used as a form of scaffolding.
	▪ Student produced rubrics for self-evaluation	▪ Asking students to develop the rubric will highlight that they understand the processes they are looking for.
Potential rubrics	▪ Students recognize the cognitive requirements of a task	▪ [detailed rubrics could be developed]
	▪ Student identifies strengths and weaknesses in their own performance	▪ [detailed rubrics could be developed]
Characteristic features	▪ A shared understanding of "guidelines for judging work"	
	▪ Work for which the guidelines are applicable	▪ Typically one's own or group's work
Variable features	▪ Formality of assessment	▪ Reflective assessment can be more or less formal or informal. To highlight certain behaviors a more formal method is required, although more informal reflection can be encouraged for nearly any task. A more informal assessment may involve a conversation with the student about what steps they took, whereas a formal assessment could involve a questionnaire, presentation, etc.
	▪ Group vs. individual reflective assessment	▪ Assessment can be a social process where students can see how multiple perspectives can be applied in viewing one's own and others' work. Starting off as group work can also help students to practice, model for others, and internalize habits of reflection.

DP-1: Reflective Assessment in BioKIDS *(continued)*

Attribute	Value(s)	Comments
Variable Features *(continued)*	▪ Amount of substantive knowledge required	▪ Some tasks may require a strong knowledge of the subject area, as understanding one's performance in that domain may not be measurable outside of the metacognitive skills.
	▪ Formative vs. summative assessment	▪ Some tasks may have several stages, allowing students the opportunity for reflection and improvement.
	▪ Specificity of metacognitive skills to particular task	▪ Some skills, such as checking one's work, are more general cognitive skills, as opposed to some subject areas that require less generalizable skills.
	▪ Amount of prompting/cueing	▪ In the initial stages of self-reflection, students will need to be prompted to look for certain criteria in their own work. This scaffolding may be removed as students develop more metacognitive skills; at this point selecting the appropriate self-monitoring skill may be more important.
I am a kind of		
These are kinds of me		
These are parts of me		
Educational standards		
Templates		
Exemplar tasks		
Online resources		
References	White, B. Y., & Frederiksen, J. R. (1998). Inquiry, Modeling, and Metacognition: Making Science Accessible to All Students. *Cognition and Instruction, 16*(1), 3-118.	
I am a part of	Inquiry cycle	

4.2 DP-2: Understand and Apply Rate, Time, and Distance Concepts in Word Problems (by Duanli Yan)

NAEP Mathematics is a part of the National Assessment of Educational Progress, a congressionally mandated project of the U.S. Department of Education's National Center for Education Statistics, NAEP publishes The Nation's Report Card, which contains the results of periodic assessments in mathematics and other content areas. The purpose of this document is to inform the public about the nature of students' comprehension of the subject; to inform curriculum specialists about the level and nature of students' understanding; and to inform policy makers about factors related to schooling and its relationship to students' proficiency in mathematics. This *design pattern* reflects the NAEP 1996-2000 Mathematics for Grade 8 assessment from the information processing theory (schema knowledge) point of view and focuses specifically on one of the five content strands addressed by NAEP: Number Sense, Properties, and Operations.

DP-2: Understand and Apply Rate, Time, and Distance Concepts in Word Problems

Attribute	Value(s)	Comments
Title	Understand and apply rate, time, and distance concepts in mathematics word problems	
Summary	This assessment is based on a set of beliefs about the kinds of tasks or situations that will prompt students to say, do, or create something that demonstrate the important KSAs in mathematics.	
Rationale	Since explicit schemas are posited to be useful for word problems, this design pattern targets the assessment of schema knowledge on both how knowledge is stored in the schema (structure properties) and what knowledge is stored (content properties), though the design pattern does help on getting at the KSAs, and we want to ensure that the students acquire those schemas.	
Focal KSAs	*Mathematical Abilities: Conceptual Understanding, Procedural Knowledge, and Problem Solving* ▪ The knowledge and ability to understand the situation expressed as a word problem ▪ The skills to recognize the constraints for the schema including ratio and proportional thinking ▪ The skills to plan/set goals about the responses or actions, and the skills to execute the steps/procedures to carry out the plan for task solutions ▪ The skills could be focused individually for each schema component or combined to assess the schema contents together.	

DP-2: Understand and Apply Rate, Time, and Distance Concepts in Word Problems *(continued)*

Attribute	Value(s)	Comments
Focal KSAs *(continued)*	*Number Sense, Properties, and Operations* ▪ The knowledge about the definition of rate, time, distance, and the relationship among them ▪ The skills to extract the necessary information from the situation, such as recognizing the features of the rate, time, and distance components of the schema, including the definitions and relationship among them ▪ The knowledge about arithmetic elements in each of the components of the schema such as numbers (whole numbers, fractions, decimals, integers), properties and operations involving those numbers, estimations and use of ratios, and proportional thinking to represent situations involving quantity and their application to real world situations ▪ The means to carry out the computations indicate mastery *Mathematical Power (Reasoning, Connections, and Communication)* ▪ The overall ability to gather and use mathematical knowledge through exploring, conjecturing, and reasoning logically; through solving nonroutine problems; and through communicating about mathematics as said earlier. It is a function of students' prior knowledge and experiences and the ability to connect that knowledge in productive ways to new contexts ▪ The knowledge and ability to reason in settings involving the careful application of concept definitions of rate, time and distance, relations among them, or representations of either	
Additional KSAs	▪ The skills to understand and write in English	
Potential observations	▪ Correctness of responses to Feature Recognition tasks, such as definition of rate, time, and distance (e.g. rate is 8 miles per 10 minutes, distance is 8 miles) ▪ Correctness of responses to Constrain tasks, such as the relationship among rate, time, and distance, and difference in distance if two cars (e.g., distance = rate × time, the starting time and ending time, starting place and ending place, the difference in distance of two cars at the end time)	

DP-2: Understand and Apply Rate, Time, and Distance Concepts in Word Problems *(continued)*

Attribute	Value(s)	Comments
Potential observations *(continued)*	▪ Appropriateness/Correctness of responses to the Planning tasks, such as the formulations of lines of attack on problems and the way in which students' reason through situations ▪ Correctness of responses to the Execution tasks, such as written arithmetic expressions or explanations, or drawings or diagrams to illustrate their reasoning, or computations of rate, time and distance, conversion of units (time, distance) individually or combined	
Potential work products	▪ Written explanations ▪ Numeric responses to tasks assessing individual feature recognition for the components of the schema or mathematical expressions for the relationship among the components ▪ Drawings or tables of the traveling distance(s) ▪ Conversion results for the components of the schema produced by students	
Potential rubrics	▪ Correctness by matching a key, but maybe additionally whether the responses are plausible but incorrect, or implausible (e.g., tasks that assessing only feature recognition schema component) ▪ Rubrics for evaluating written explanations, drawings, or tables if for comparison of distance at different rates	
Characteristic features	▪ Instructions asking students to enter numeric responses for easy tasks, such as those to assess feature recognitions of the schema alone, or asking students to write/draw/explain their reasoning by steps to show their thinking ▪ Illustrations of the tasks for instructions, and illustrations or partial illustrations which students are required to complete	Tasks are word problem questions designed to ask students to make explicit their understanding of the concept of rate, distance, and time, the relationship among them, their reasoning, as well as elicit features of their schema knowledge and connections in this domain.

DP-2: Understand and Apply Rate, Time, and Distance Concepts in Word Problems *(continued)*

Attribute	Value(s)	Comments
Variable features	▪ Selection of the variable categories for rate (same for time and distance) with the variation of numbers (from integers to fraction, or decimals) to vary the difficulty of the tasks ▪ Selection of a variable categories for time unit with hour, minute, second, day, month, or year ▪ Selection of the variable categories for distance unit can include miles, meters, kilometers, yards, or feet ▪ Selection of more than one car for comparative reasoning, such as the example (compare the rate for two cars) with different variable units for different cars ▪ Substitution of a variable for car with airplane, train, or boat would change the situation that evokes the desired evidence, but it may or may not affect the difficulty of tasks	
I am a kind of	Design pattern for Number Sense, Properties, and Operations	
These are kinds of me	Design patterns for arithmetic operation	
These are parts of me	Design patterns for understand and apply rate only, for time only, for units (time, distance) conversions	
Educational standards		
Templates		
Exemplar tasks		
Online resources	http://www.nces.ed.gov/nationsreportcard/mathematics/	
References		
I am a part of	Design pattern for NAEP Mathematics	

4.3 DP-3: Scientific Investigation—Establishing Experimental Controls (by Joy Barnes)

This *design pattern* is based on the Maryland State High School Assessment (MDHSA) in Biology. The MDHSA in Biology is an end-of-course assessment designed to evaluate how well students decode the language and evidence of life science as a discipline. It recognizes the interrelatedness of other science content (math, physical science) and seeks to incorporate universal scientific content with life science applications. The MDHSA is a 70-item assessment administered as a field test from 2000–03; it became an operational test starting in 2004. With every administration of the test, approximately 55 items are released via the Internet to the public as a tool for curriculum development (in the case of educators) and for improved public relations (in the case of families). All of the items that have been publicly released are available online at the Web site http://mdk12.org/mspp/high_school/look_like/index.html.

The Maryland State Department of Education (MSDE) has established five science core learning goals: 1) Skills and Processes, 2) Earth/Space Science concepts, 3) Biology concepts, 4) Chemistry concepts, and 5) Physics concepts. The Biology High School Assessment is designed to measure goals 1 and 3. The MSDE format for evaluating the national standards utilizes a system with three convergent descriptors: the goal is the umbrella; the expectations form the ribs that support the goal. The indicators are akin to the stitches that keep the spokes attached to the umbrella. Indicators are the assessment limits from which all test material may be generated. This *design pattern* specifically addresses items that evaluate goal 1, expectation 2, and indicator 6 (1.2.6): "The student will identify appropriate methods for conducting an investigation and affirm the need for proper controls in an experiment."

DP-3: Scientific Investigation—Establishing Experimental Controls

Attribute	Value(s)	Comments
Title	Scientific investigation—establishing experimental controls (MSDE 1.2.6)	
Summary	Identifying realistic parameters for experimentation.	
Rationale	Demonstrate science habits; communicate scientific thinking.	
Focal KSAs	Scientific method	
Additional KSAs	▪ Science equipment identification/usage ▪ Experimental conditions (plant-soil properties, animal behaviors, concentration effects)	
Potential observations	▪ Logical/appropriate use of terminology ▪ Logical/appropriate use of instruments/equipment	
Potential work products	Drawings, tables, calculations, graphs	
Potential rubrics	BCR Scoring Rubric—MDHSA 2000	Rubrics available at http://www.mdk12.org/mspp/high_school/structure/biology/index.html
Characteristic features	▪ Experimental conditions known at all times for one set of data. Known set has predictable result. Variable sets have at least one unique condition apart from the known set. ▪ Instruments/equipment used is appropriate for size, time, resource constraints of the materials being investigated.	

DP-3: Scientific Investigation—Establishing Experimental Controls *(continued)*

Attribute	Value(s)	Comments
Variable features	Description of physiological properties of materials being investigated	
I am a kind of		
These are kinds of me		
These are parts of me		
Educational standards		
Templates		
Exemplar tasks		
Online resources	http://www.mdk12.org/scripts/hsa_practice_scoring.plx?subj=biology&item=12&practice_set=psa	
References		
I am a part of		

4.4 *DP-4: Selected Math Items from the Standardized Achievement Test (SAT) I (by Kia Johnson)*

This *design pattern* is based on the mathematics section of the SAT I. The SAT, a well-known college entrance exam, is among one of the most important and controversial tests of a high school student's career. The test, combined with high school transcripts, extracurricular activities, and recommendations, is given considerable weight in admission to most colleges and universities. Unfortunately, there are concerns about the fairness of this assessment.

By addressing the assessment from a sociocultural perspective, the *design pattern* structure makes visible those aspects of SAT mathematics questions that potentially contribute to an unfair examination. In particular, additional KSAs, Potential Observations, Characteristic Features and Variable Features highlight aspects of the test questions that are not related to mathematics knowledge but that may influence students' performance. Students who do not possess certain additional KSAs, such as vocabulary skills, may not perform well on certain math problems. Scratch work in students' test books, such as the underlining of text, may reveal vocabulary words that were confusing for students. In addition, features of the mathematics assessment items (e.g., number of multiple-choice items) or features of the assessment (time pressure) may favor boys. If these aspects of the assessment are not considered when making inferences about students' mathematics proficiency, it is possible to make an inaccurate judgment about what some students know and can do in mathematics.

DP-4: Selected Math Items from the Standardized Achievement Test (SAT) I

Attribute	Value(s)	Comments
Title	Math section of the SAT I	
Summary	Using a sociocultural perspective, this design pattern underscores considerations for avoiding bias in the math section of the SAT I.	
Rationale	The intended use of this problem was to test a student's ability to comprehend a "real-world" arithmetic word problem and apply ratio/proportion skills to the problem in order to reach the key.	
Focal KSAs	▪ Arithmetic word problem ▪ Ratio/proportion	
Additional KSAs	▪ Vocabulary ▪ Familiarity with structure of SAT (e.g., knowledge about difficulty ordering of test items) ▪ Familiarity with SAT test-taking strategies (e.g., knowledge about process of elimination)	Since lack of vocabulary could lead to inaccurate examinee responses that do not reflect math proficiency, it is important to structure items appropriately and incorporate opportunities to rule out this kind of alternative hypothesis in interpreting responses. Familiarity with the SAT structure and test-taking strategies relate to the impact of coaching and ability to attend SAT prep courses. Lack of familiarity with SAT structure represents another alternative hypothesis in interpreting responses.
Potential observations	▪ Correct answer filled in on bubble sheet ▪ Underlined words in test book ▪ Eliminate wrong answers by crossing them out	Students' scratch work in their test booklets is one source with potential to reveal which vocabulary words were challenging and which strategies students used to solve the problems.
Potential work products	▪ Multiple-choice answer ▪ Short-constructed response ▪ Written explanation of problem situation ▪ Think aloud solution	Written explanations of the problem situation and think aloud solutions are potential work products that could improve the quality of inference about student math proficiency.

DP-4: Selected Math Items from the Standardized Achievement Test (SAT) I *(continued)*

Attribute	Value(s)	Comments
Potential rubrics	Answer key	
Characteristic features	▪ Multiple-choice items ▪ Short-constructed response items ▪ Problems requiring arithmetic skills ▪ Problems requiring geometric skills ▪ Problems requiring combined arithmetic and geometric skills	
Variable features	▪ Length of time allotted for examination ▪ Number of problems requiring arithmetic skills, geometric skills, and combined arithmetic and geometry skills ▪ Paper-pencil vs. computerized ▪ Testing context	Performance on tests also depends on the context in which the task is administered: the classroom environment, the attitude of the test administrator, the cultural climate of the school, and societal messages about the relationship between ethnicity and mathematics ability.
I am a kind of		
These are kinds of me		
These are parts of me		
Educational standards		
Templates		
Exemplar tasks		
Online resources		
References		
I am a part of		

4.5 DPs-5A, 5B, 5C: Portfolios for Performance Assessments (by Patricia Verdines-Arredondo)

The following *design patterns* are based on the features and attributes of two portfolio assessments used in graduate courses at the University of Maryland College Park. DP-5A reflects the course “Introduction to Qualitative Methods in Communication Research,” in which students are required to design and develop a portfolio as a means of assessing their understanding of basic knowledge on qualitative research methods in communication science. DP-5B reflects the course “Collaborative Instructional Design and Evaluation,” in which students design and develop a portfolio that is used to assess their ability to apply their knowledge on instructional design theory and the instructional design process. After analyzing the features of the individual portfolio shown in DP-5A and the collaborative portfolio shown in DP-5B, A Generic Design Pattern for Portfolio Assessment was derived from both portfolio specifications, as shown in DP-5C.

DP-5A: Design Pattern for an Individual Portfolio

Attribute	**Value(s)**	**Comments**
Title	The Qualitative Researcher Portfolio	
Summary	A portfolio is used in this course as a means to assess students' performance regarding the formulation of research questions, the design of a qualitative research project, the interpretation of qualitative data, and as a means for students' ability to reflect on the qualitative research process and their role as qualitative researchers.	
Rationale	A portfolio as a means of performance assessment provides the students not only with an opportunity to understand the epistemology of qualitative research and the nature of qualitative methods but also with an opportunity to reflect about the role of qualitative researchers in a research project.	
Focal KSAs	*Conceptual knowledge about:* ▪ The epistemology of qualitative research ▪ Qualitative research methods ▪ In-depth interviews ▪ Participant observation methods ▪ Qualitative data analysis *Procedural knowledge about:* ▪ In-depth interviews ▪ Participant observation methods ▪ Qualitative data analysis *Self-evaluation skills* *Self-reflection skills*	
Additional KSA(s)	▪ Conceptual knowledge of research topic in a specific domain ▪ Verbal ability	
Potential observations	▪ Quality of data analysis as shown in the transcripts ▪ Depth and breath of self-reflexivity as shown in the memos ▪ Quality and structure of ideas as shown in the portfolio organization	

DP-5A: Design Pattern for an Individual Portfolio *(continued)*

Attribute	Value(s)	Comments
Potential observations *(continued)*	▪ Understanding of in-depth interview design as shown in the interview protocol ▪ Understanding of participatory observation method as shown in the observation field notes ▪ Understanding of the process of open coding as shown in the coding scheme	
Potential work products	*A portfolio with:* ▪ Research questions ▪ Researcher assumptions ▪ Interpretation of research results ▪ Interview protocol ▪ Interview transcripts ▪ Observation field notes ▪ Coding scheme ▪ Self-reflective memos	
Potential rubrics	▪ Accuracy ▪ Consistency ▪ Reflexivity ▪ Completeness ▪ Transparency ▪ Correct use of language ▪ Portfolio organization ▪ Academic integrity	
Characteristic features	▪ Individual portfolio work ▪ Individual practice during class hours ▪ Portfolio topic selected by student ▪ Portfolio work developed over the course of one semester	
Variable features	▪ Research topic selected by student ▪ Conceptual knowledge of research topic in a specific domain	
I am a kind of	▪ Performance assessment tool ▪ Learning tool	
These are kinds of me		
These are parts of me		
Educational standards		
Templates		
Exemplar tasks		
Online resources		

DP-5A: Design Pattern for an Individual Portfolio *(continued)*

Attribute	Value(s)	Comments
References	Syllabus of the course “Introduction to Qualitative Methods in Communication Research” (COMM-714)	
I am a part of		

DP-5B: Design Pattern for a Collaborative Portfolio

Attribute	Value(s)	Comments
Title	The Instructional Design Team Portfolio	
Summary	A collaborative portfolio is used in this course as a means to promote and assess the students' collaboration skills and performance regarding the design of an instructional unit.	
Rationale	A collaborative portfolio as a means of performance assessment provides the students not only with an opportunity to understand the nature of instructional design theory, but also with an opportunity to design an instructional unit for a specific audience, while reflecting about the role of instructional designers in educational environments.	
Focal KSAs	▪ Conceptual knowledge of instructional design theory ▪ Conceptual knowledge of instructional design models ▪ Procedural knowledge of the instructional design process ▪ Self-evaluation skills ▪ Self-reflection skills ▪ Collaboration skills	
Additional KSAs	▪ Conceptual knowledge of information literacy standards ▪ Verbal ability	
Potential observations	▪ Consistency between instructional goals and objectives ▪ Consistency between objectives and instructional activities ▪ Self-reflexivity as shown in self-evaluation forms ▪ Collaboration skills as shown in team-evaluation forms ▪ Quality and structure of ideas as shown in the portfolio organization ▪ Understanding of instructional design theory as shown in the instructional sequence ▪ Understanding of the instructional design process as shown by the content of the portfolio work products	

DP-5B: Design Pattern for a Collaborative Portfolio *(continued)*

Attribute	Value(s)	Comments
Potential work products	▪ Self-evaluation forms ▪ Team evaluation forms ▪ Portfolio with instructional sequence ▪ Collaborative presentation and discussion during class hours	
Potential rubrics	▪ Accuracy ▪ Consistency ▪ Salience ▪ Thoroughness ▪ Correct use of language ▪ Effective organization ▪ Effective use of media ▪ Intellectual integrity	
Characteristic features	▪ Collaborative portfolio work ▪ Team meetings during class hours ▪ Portfolio topic selected by each design team ▪ Portfolio work developed over the course of one semester	
Variable features	▪ Number of team members (2, 3 or 4) ▪ Content of the instructional unit selected by each design team ▪ Conceptual knowledge of the topic selected for each design team	
I am a kind of	▪ Performance assessment tool ▪ Learning tool	
These are kinds of me		
These are parts of me		
Educational standards		
Templates		
Exemplar tasks		
Online resources		
References	Syllabus of the course "Collaborative Instructional Design and Evaluation" (LBSC-742)	
I am a part of		

DP-5C: A Generic Design Pattern for Performance Assessment

Attribute	Value(s)	Comments
Title	The Portfolio Project	
Summary	A portfolio assessment can be used as a means to promote and assess the students' deep understanding of new concepts in a given context. A portfolio can also foster the development of the students' self-evaluation, self-reflection, and collaboration skills.	
Rationale	A portfolio as a means of performance assessment provides the students not only with an opportunity to understand new concepts and methods in a given domain but also with an opportunity to develop their collaborative skills, and to reflect about their role in the portfolio development process.	
Focal KSAs	▪ Conceptual knowledge of the domain ▪ Procedural knowledge of the domain ▪ Self-evaluation skills ▪ Self-reflection skills	
Additional KSAs	▪ Conceptual knowledge in topics related to the domain ▪ Collaboration skills ▪ Verbal ability	
Potential observations	▪ Consistency between portfolio work products and objectives ▪ Self-reflexivity as shown in presentations or written documents ▪ Collaboration skills as shown in team-evaluation forms ▪ Quality and structure of ideas as shown in the portfolio organization ▪ Understanding of concepts shown in the portfolio work products ▪ Understanding of procedures and methods as shown in the portfolio work products	
Potential work products	▪ Portfolio with samples of a student's work ▪ Portfolio with samples of the work of a group of students ▪ Discussion of portfolio content in class ▪ Presentation of portfolios in class ▪ Team evaluation forms ▪ Self-evaluation forms	

DP-5C: A Generic Design Pattern for Performance Assessment *(continued)*

Attribute	Value(s)	Comments
Potential rubrics	▪ Relevance ▪ Accuracy ▪ Originality ▪ Consistency ▪ Reflexivity ▪ Authenticity ▪ Completeness ▪ Correct use of language ▪ Portfolio organization ▪ Academic integrity	
Characteristic features	▪ Portfolio topic selected by student(s) ▪ Portfolio work developed over the course of one semester or a year	
Variable features	▪ Individual or collaborative portfolio ▪ Individual or collaborative instruction and practice ▪ Practice during class hours or after class hours ▪ Conceptual knowledge required for the portfolio work ▪ Procedural knowledge required for the portfolio work ▪ Number of team members (2, 3 or 4) in collaborative portfolios ▪ Content of the portfolio selected by each student or design team	
I am a kind of	▪ Performance assessment tool ▪ Learning tool	
These are kinds of me	▪ The Qualitative Researcher Portfolio ▪ The Instructional Design Team Portfolio	
These are parts of me		
Educational standards		
Templates		
Exemplar tasks		
Online resources		
References		
I am a part of		

4.6 DPs-6A, 6B, 6C, 6D: AP Studio Art Portfolio and Other Hypothetical Design Patterns (by Michelle Riconscente)

The Advanced Placement (AP) Studio Art Program is a national assessment developed and administered by the Educational Testing Service for the College Entrance Examination Board, with the aim of indicating a high school student's level of competency in the visual arts with respect to expected achievement of students completing their first year of post-secondary art instruction (College Entrance Examination Board (CEEB), 2001). In its current implementation, portfolios are accepted in three categories: Drawing, 2-D Design, and 3-D Design. A participating student submits a portfolio of original works consisting of three sections focused respectively on quality, concentration, and breadth. Distinctive of the AP Studio Art program is its success in applying a set of criteria with high interrater reliability to a broad range of works (Myford & Mislevy, 1995). A network of several *design patterns* were created to describe the AP Studio Art assessment. The overarching *design pattern* (DP-6A) considers the entire portfolio and includes descriptions of the individual sections. A finer-grained *design pattern* is considered based on the contents and evaluation criteria for the Quality Section of the AP Studio Art portfolio (DP-6B).

By using the AP Studio Art *design patterns*, it is possible to sketch a hypothetical *design pattern* for a parallel assessment in another domain; for example, writing is considered, and a potential *design pattern* is generated (DP-6C). Although the substantial work involved in attaining agreement around a set of criteria is not represented in the *design pattern*, it shows the potential for such a framework to apply to the assessment of writing. Taking the overall AP Studio Art *design pattern* as a starting point, it is also possible to design "up" a level and consider what a *design pattern* might look like for a domain-free portfolio assessment (DP-6D).

DP-6A: AP Studio Art Drawing Portfolio Design Pattern (derived from College Entrance Examination Board, 2001 and Educational Testing Service, 2002)

Attribute	Value(s)	Comments
Title	AP Studio Art	
Summary	The AP Studio Art assessment indicates the qualifications of a student based on a comparison of a portfolio of original art work to expectations for students completing their first year of college.	
Rationale	"Provides a national standard for performance in the visual arts" (CEEB, 2001, p. 5). "Allows students to earn college-credit and/or advanced placement while still in high school." (CEEB, 2001, p. 5).	
Focal KSAs	▪ Ability to produce high quality works of art (concept, composition, execution) that involve directly making marks on a surface ▪ Expertise in a particular concentration area (particular visual interest or problem ▪ Possession of breadth of ability (formal, technical, expressive)	

DP-6A: AP Studio Art Drawing Portfolio Design Pattern (derived from College Entrance Examination Board, 2001 and Educational Testing Service, 2002) *(continued)*

Attribute	Value(s)	Comments
Additional KSAs	Facility with particular medium in which student concentrates	
Potential observations	Overall quality of each portfolio section	
Potential work products	Three portfolio sections, representing quality, breadth, and concentration	
Potential rubrics	CEEB process for assigning overall score based on individual section scores	
Characteristic features	Three portfolio sections, each containing original works of art according to requirements described in CEEB, 2001	
Variable features	Media in which the work is rendered	
I am a kind of	Studio Art portfolio assessment	
These are kinds of me	2-D design, 3-D design	
These are parts of me	Quality portfolio, concentration portfolio, breadth portfolio, individual piece assessment	
Educational standards		
Templates		
Exemplar tasks		
Online resources	AP Studio Art Web site (http://www.collegeboard.org/ap)	
References		
I am a part of	Portfolio assessment; performance assessment; large-scale assessment	

DP 6B: AP Studio Art Drawing Portfolio Quality Section Design Pattern (derived from College Entrance Examination Board, 2001, 2002)

Attribute	Value(s)	Comments
Name	AP Studio Art: Quality section	
Summary	Student's portfolio of five drawing works is assessed for excellence, focused on concept, composition, and execution.	
Rationale	Demonstrated quality, including the realization of the artist's intentions, is a fundamental aspect of excellence in studio art.	

DP 6B: AP Studio Art Drawing Portfolio Quality Section Design Pattern (derived from College Entrance Examination Board, 2001, 2002) ***(continued)***

Attribute	Value(s)	Comments
Focal KSAs	Possession of sense of excellence in studio art (specifically with regard to concept, composition, and execution)	
Additional KSAs	Technical skill in working with media of choice (e.g., paint, pencil, etc.)	
Potential observations	Degree to which works demonstrate excellence in concept, composition, and execution	
Potential work products	Five actual works of art (flat)	
Potential rubrics	Scoring criteria for the Quality section (CEEB)	
Characteristic features	▪ No time or location restrictions for producing work ▪ Students have access to exemplars (AP Studio Art poster) ▪ Artwork on flat paper, canvas board, or unstretched canvas ▪ Maximum dimensions: 18 in. × 24 in.	
Variable features	▪ Works may or may not be related to each other ▪ Works can take a wide variety of forms (drawing, paintings, prints, diagrams, plans, collages, montages, etc.)	
I am a kind of	Art Portfolio section	
These are kinds of me	Breadth section, concentration section	
These are parts of me	Assessment of individual pieces (drawing, painting, etc.)	
Educational standards		
Templates		
Exemplar tasks		
Online resources	College Board Web site (http://www.collegeboard.org/ap)	
References		
I am a part of	AP Studio Art Portfolio	

DP-6C: A Hypothetical Writing Portfolio Design Pattern (for Breadth section)

Attribute	Value(s)	Comments
Title	A hypothetical writing portfolio: Breadth	
Summary	Student's portfolio of six written works is assessed for quality as well as successful use of a range of styles and structures.	
Rationale	Demonstrated breadth and quality of writing is a fundamental aspect of competent writers.	
Focal KSAs	Ability to produce quality writing across a range of forms	
Additional KSAs	Language proficiency	
Potential observations	Degree to which works demonstrate quality and taken together demonstrate competence across a range of forms	
Potential work products	Six written works	
Potential rubrics	Scoring criteria for the Breadth section	
Characteristic features	Six written works completed	
Variable features		
I am a kind of	Writing Portfolio section	
These are kinds of me	Quality section, Concentration section	
These are parts of me	Assessment of individual pieces (short stories, prose, exposition, plays, comedy sketches)	
Educational standards		
Templates		
Exemplar tasks		
Online resources		
References		
I am a part of	Hypothetical Writing Portfolio	

DP-6D: A Hypothetical Domain-Free Portfolio Design Pattern

Attribute	Value(s)	Comments
Title	A hypothetical domain-free portfolio	
Summary	Student's portfolio within a domain is assessed for overall competence comparable to that of a student completing the first year of college study in that domain.	
Rationale	Demonstrated competence in a domain for a year of college credit and placement.	
Focal KSAs	Ability to perform or produce quality work in a given domain	
Additional KSAs	Abilities and skills germane to specific domains	
Potential observations	Degree to which work products and process demonstrate competence in a given domain	
Potential work products	A collection of works demonstrating quality, expertise, and range (breadth) of ability	
Potential rubrics	Scoring criteria for elements of a given domain	
Characteristic features	Variety of works related to a given domain	
Variable features	Domain-specific characteristics; time and space/resource constraints; feedback processes	
I am a kind of		
These are kinds of me		
These are parts of me		
Educational standards		
Templates		
Exemplar tasks		
Online resources		
References		
I am a part of		

4.7 DP-7: Reasoning to Maximize the Difference Between Two Numbers for NAEP Grade 8 Mathematics (by René Lawless)

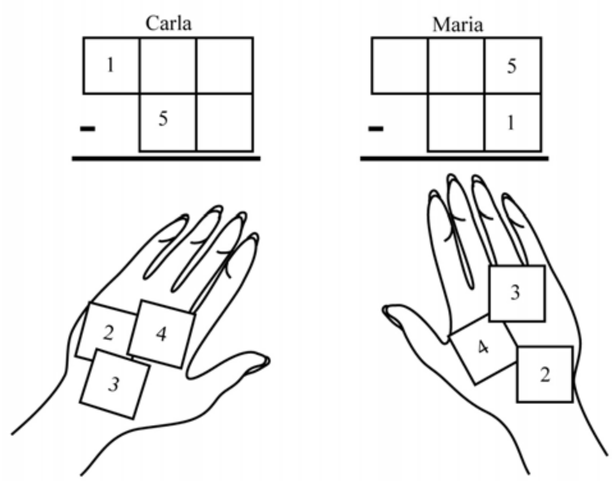
Like DP-2, this *design pattern* focuses on a subset of the NAEP Mathematics content strand, Number Sense, Properties, and Operations. This *design pattern* is based on an item (Figures 1 and 2) that addresses three key features of this content strand: (1) ability to "represent numbers and operations in a variety of equivalent forms using models, diagrams, and symbols"; (2) ability to "compute with numbers (i.e., add, subtract, multiply, divide)"; and (3) ability to "use computation in applications."

Figure 1. NAEP Mathematics Item Used to Develop DP-7

The question requires you to show your work and explain your reasoning. You may use drawings, words, and numbers in your explanation. Your answer should be clear enough so that another person could read it and understand your thinking. It is important that you show <u>all</u> your work.

13. In a game, Carla and Maria are making subtraction problems using tiles numbered 1 to 5. The player whose subtraction problem gives the largest answer wins the game.

Look at where each girl placed two of her tiles.



Who will win the game? ____________________

Explain how you know this person will win.

Note. From U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress [NAEP], 1996 Mathematics Assessment

Figure 2. NAEP Mathematics Scoring Guide Used to Develop DP-7

Scoring Guide

Solution:
Maria will win the game.
The following reasons may be given:

a. The largest possible difference for Carla is less than 100 and the smallest possible difference for Maria is 194.
b. Carla will only get a difference of 91 or less but Maria will get several larger differences.
c. Carla can have only up to 143 as her top number but Maria can have 435 as her largest number.
d. Carla has only 1 hundred but Maria can have 2,3,or 4 hundreds.
e. Maria can never take away as much as Carla.
f.

1	4	3
	5	2

3	4	5
	2	1

Any combination of problems to show that Maria's difference is greater.

Scoring Guide
In this question a student needed to use number skills to understand place value and compare numbers. Since Carla placed her number 1 tile in the hundreds place, the greatest number she could have after subtracting would be less than one hundred. Maria could have used the number 2, 3, or 4 tile in the hundreds place and her difference would always be larger than Carla's. For an extended response, the student needed to answer "Maria" and demonstrate understanding of place value by generalizing a comparison of the possible differences that Carla could obtain to the possible differences that Maria could obtain. ("Generalize" means that the student indicates that since Carla placed her number 1 tile in a place so that she could never win, Maria would always win, no matter how she placed her 2, 3, or 4 tiles.) For a satisfactory response, a student needed to demonstrate understanding that Maria could make a larger top number than Carla, but the response did not generalize Maria's and Carla's possible differences. For a partial response, a student had to provide an explanation that was only partially correct; however, those types of responses did recognize that Maria would have the greater number after determining the difference. A minimal score was earned by responses that indicated that Maria would win, but did not offer an explanation for how Maria would win the game.

Extended
Student answers Maria and gives explanation such as a or b, or an appropriate combination of the other explanations.

Satisfactory
Student answers Maria and gives explanation such as c, d, or e.

Partial
Student answers Maria with partially correct, or incomplete but relevant, explanation.

Minimal
Student answers Maria and gives sample such as in f but no explanation or Maria with an incorrect explanation.

Incorrect
Incorrect response

Note. From: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress [NAEP], 1996 Mathematics Assessment

DP-7: Reasoning to Maximize the Difference Between Two Numbers for NAEP Grade 8 Mathematics

Attribute	Value(s)	Comments
Title	Reasoning to maximize the difference between two numbers	
Summary	Items based on this design pattern seek to measure whether a student knows how to use number skills to understand place value and compare numbers in such a way that when provided with two different scenarios, he/she can determine which combination of numbers will provide the largest difference.	These types of items may lend themselves to automatic item generation.
Rationale	*This design pattern uses an extended-response format (for arithmetic number sense, properties, and operations) that connects with three psychological perspectives:* ▪ The trait perspective—by examining the correctness of the student's mathematical outcomes. ▪ The situative perspective—by examining the student explanations to determine if the student recognizes and understands what the item is asking, applies the correct schema to solve the item, and provides the reasons why his or her solution is correct. Further, this type of item places the numerical issues in social situations that the student can relate to in his/her own experience. ▪ The cognitive perspective—to measure the student's mathematical ability to correctly solve the problem and award partial credit for the student's ability to solve the problem, even if it is partially incorrect. In this case, the student is given credit for using different levels of correct reasoning. *By utilizing an extended response format, the assessor has the opportunity to examine student* learning *through the combination of his/her final solution as well as his/her rationale.*	The extended-response format may also provide more information about the cognitive processes that the student used to reach his/her solution.

DP-7: Reasoning to Maximize the Difference Between Two Numbers for NAEP Grade 8 Mathematics *(continued)*

Attribute	Value(s)	Comments
Focal KSAs	▪ This design pattern is concerned with measuring number sense, properties, and operations. Specifically, it is concerned with measuring students' understanding of numbers, using arithmetic operators correctly, and applying this understanding to a real-world situation. ▪ Students are also expected to demonstrate their ability to generalize from numerical patterns and verify the results that they attain. ▪ Students are expected to read, write, order, and compare numbers. ▪ Students are expected to compute with numbers and describe the effect of operations on size and order of numbers. ▪ Students are also expected to verify solutions and be able to determine the reasonableness of results in the real-world situation presented in the items.	This NAEP mathematics assessment has many KSAs. However, evidence of many of these is only implied and cannot be directly measured. The only KSAs measured in the statistical model are those of a very coarse grain-size, i.e., each of the five categories classified as dimensions in the Content Strands. Thus, in NAEP, only overall mathematical classifications are measured and reported. The student model variables found in the Mathematical Abilities are so highly correlated that they cannot be isolated and measured individually. However, inferences can only be made from student work products as to whether, in fact, the students have the attributes that we are interested in measuring.
Additional KSAs	▪ Student must have an ability to understand written English. ▪ Students must also be able to communicate the reasoning that they used to either construct the problem or justify why their solution is the correct one. ▪ Students may display their ability to construct arithmetic expressions and/or diagrams to communicate their thoughts.	
Potential observations	*Students may provide one of the following written reasons to explain why their solution is correct:* ▪ Reasons indicating the largest possible difference for one scenario as compared with the smallest possible difference for the other.	The reasons that a student might produce could provide clues regarding their targeted thinking.

DP-7: Reasoning to Maximize the Difference Between Two Numbers for NAEP Grade 8 Mathematics *(continued)*

Attribute	Value(s)	Comments
Potential observations *(continued)*	▪ Reasons indicating the smallest possible difference in one scenario as compared with multiple larger solutions in the other. ▪ Reasons indicating the largest possible numbers for each scenario before subtraction. ▪ Reasons comparing the size of the numbers that are being subtracted in each scenario and demonstrating it in the matrices provided in the item. ▪ Reasoning demonstrating the differences through the completion of the matrix. ▪ Any combination of problems to show that the second scenario has a greater difference.	
Potential work products	*Solution in the provided space, answering the first part of the item. In the second part of the item:* ▪ Written explanations by students describe their reasoning behind the answer that they provided to the first part of the item. ▪ Numbers filled into one or both of the matrices provided in the item. ▪ Drawings of the matrices with numbers filled into each cell of each matrix. ▪ Equations showing student work. ▪ Diagrams or arrowed comments used to emphasize the contents of each (or both) matrix.	

DP-7: Reasoning to Maximize the Difference Between Two Numbers for NAEP Grade 8 Mathematics *(continued)*

Attribute	Value(s)	Comments
Potential rubrics	Each student work product is compared to the scoring guide to ascertain at which scoring level it is an exemplar. The first part of the item is compared to the key to determine whether the student arrived at the correct solution. Then, the student's explanation is compared to the potential observations and benchmarks to decide the appropriate scoring level that should be awarded. Each scoring level carries a point value that will be counted as part of the total score. It should be assumed at this stage that the raters have already been calibrated using the benchmarked student solutions and a small, random sample of unscored, student work.	As previously mentioned, there are many student model variables built into the framework. Thus, although the types of evidence sought have been identified, it is difficult to clearly accumulate evidence for many of those variables. This is a result of the implied nature of these variables and their high intercorrelation. Hence, these variables cannot be isolated and measured individually. However, the evaluation rules (evidence rules) used to score the example assessment are consistent and therefore inferences may be made based on the observable work products, particularly in the case of extended open-ended items, in terms of the quality of the student's responses. See Figure 2 for the scoring guides (and rationales) used for the example item. In the case of new items, appropriate variable names would be substituted.
Characteristic features	▪ Instruction: Word problems prefaced with instructions (to students) indicating that the problems have multiple steps. These instructions should indicate that not only should the student show their work but also make sure that their answers are clear enough that another person reading their solution can understand what the student is thinking in the problem. Further, students should be encouraged to use drawings, words, and numbers to help illustrate their reasoning.	

DP-7: Reasoning to Maximize the Difference Between Two Numbers for NAEP Grade 8 Mathematics *(continued)*

Attribute	Value(s)	Comments
Characteristic features *(continued)*	▪ Illustrations: Word problems should have accompanying diagrams and/or graphics to illustrate the initial conditions of the numbers and objects to be manipulated to attain the final solution. ▪ Matrices: The items should contain matrices containing the initial state of numbers and be further clarified using the applicable arithmetic operator and total line. It is intended that through the use of these matrices that the students will be prompted to use the correct schema when solving the presented item.	
Variable features	*To change the focus of the items the following may be altered in order to induce different schemas and/or change the difficulty of the assessment/item:* ▪ The positions of the numbers within each matrix. ▪ The actual numbers within each matrix. ▪ The arithmetic operators of the matrices, i.e., for addition, subtraction, multiplication, or division. ▪ The surface features of the word problems, i.e., the names of the children, games using tiles with numbers, playing cards, dominos, coins. ▪ The objective of the situation is to be manipulated to demonstrate different goals, i.e., a game is being played and you need to identify the biggest difference, the smallest difference, the largest sum, the largest product, or a common divisor; a person is going to a store in a foreign country with new currency and must determine differences, sums, or products, in order to buy something; a farmer is planting a field with different vegetables—each plant yielding a different number of vegetables, etc.	
I am a kind of	Abilities necessary to understand how to perform arithmetic number sense, properties, and operations	

DP-7: Reasoning to Maximize the Difference Between Two Numbers for NAEP Grade 8 Mathematics *(continued)*

Attribute	Value(s)	Comments
These are kinds of me	N/A	
These are parts of me	N/A	
Educational standards		
Templates		
Exemplar tasks	N/A	
Online resources	http://www.nces.ed.gov/nationsreportcard/math/math_about_strand.asp	
References		
I am a part of	NAEP Grade 8 Mathematics	

5.0 Discussion: Benefits of Design Patterns

In this section we discuss three key benefits of using *design patterns* in the assessment development process. *Design patterns* facilitate decision-making about assessment design, explicate the assessment argument, and afford flexibility in usage for assessment design. These benefits of *design patterns* are illustrated with the examples of assessment *design patterns*. In addition, we use these examples to show how *design patterns* can vary in their generality and scale and in the psychological perspective they represent. Table 2 summarizes how these *design pattern* examples reflect these characteristics.

Table 2. Design Pattern Benefits and Examples Matrix

	Reflective Assessment (Morisson)	NAEP Math Word Problems (Yan)	Scientific Investigation (Barnes)	SAT I—Math (Johnson)	Performance Assessment Portfolios (Verdines-Arredondo)	AP Studio Art (Riconscente)	NAEP Math Reasoning (Lawless)
	DP-1	DP-2	DP-3	DP-4	DP-5	DP-6	DP-7
Facilitation of Decision-Making	■		■	■			
Explication of the Assessment Argument	■	■	■	■	■	■	■
Flexibility:							
▪ Psychological Perspectives	■	■	■	■	■	■	■
▪ Generality		■			■	■	
▪ Interdependence	■	■				■	
▪ Scale			■		■	■	

5.1 Facilitation of Decision-Making about Assessment Design

The development of assessments requires a series of decisions which, like a funnel, start from broad possibilities and arrive at specific items or tasks by progressively adding constraints. Given the complexity and range of decisions that may be required to create even a single assessment item, one value afforded by the *design pattern* structure is that of identifying for the assessment designer which decisions already have been made and which still need to be made, thus providing a framework for documenting those decisions. By explicating these decisions within a shared schema, many assessments can subsequently be generated by designers without having to retrace those decision paths.

Although it could be taken for granted that assessments are designed with the purpose of the assessment in mind, in practice, other factors may drive the form or shape of the assessment, resulting in compromised inferences about student understanding or ability. The attributes in *design patterns* are essential in scaffolding thinking and decisions about the purpose of the assessment. In the absence of a framework like *design patterns*, however, it is likely that the decisions explicit in this report may not be considered or intentionally determined. The value of *design patterns* results in strengthening the assessment argument, which in turn maximizes the quality of inferences made about students.

5.2 Explication of the Assessment Argument

In laying out the design decisions, a crucial consideration is the assessment argument, the line of reasoning that will ultimately connect the assessment item or task to the inferences we wish to make about student KSAs. *Design patterns* prompt assessment designers to articulate the assessment argument by eliciting information about the KSAs, Potential Work Products, Rubrics, and Observations, as well as the features of tasks themselves. Rather than immediately constructing a particular assessment item, the *design pattern* structure facilitates a careful consideration of how the item will be measured, which skills the items will assess, and how it will provide evidence of attainment of those skills. That is, the assessment argument is rendered explicit by the decisions expressed in each *design pattern* attribute.

In addition to the value of creating an individual assessment item or task that is based on evidence-centered design (ECD), this explication of the assessment argument is important for generating multiple tasks to inform a related set of inferences. Assessment items derived from a common *design pattern* will consequently share an underlying assessment argument. Therefore, although surface features of two items may be substantially different, these items are appropriate for informing a common set of inferences about student KSAs. All the *design patterns* presented in this paper are useful for illustrating how the assessment argument is rendered explicit within a *design pattern*. For instance, in the DP-1: Reflective Assessment in BioKIDS example, the Summary, Rationale, and KSAs attributes provide a clear description of the focus and goal of assessments to be created from this *design pattern*. A reflective assessment task should provide information about students' metacognitive skills. From the Rationale, we see that these skills are important because they help students to "monitor and improve their learning as well as acquire content knowledge." The Potential Observations, such as "recognizing and resolving contradictions between one's own and a standard Work Product" and "applying generally stated qualities of a rubric to the specifics of one's own or group's work" would provide evidence of metacognitive skills. A Work Product, such as "critiquing a flawed experiment/project," could be used to elicit those skills. The assessment argument is elaborated further as the Characteristic Features and Variable Features of reflective assessment tasks are described.

5.3 Flexibility

The flexibility of *design patterns* is another distinguishing characteristic of PADI's approach to ECD. *Design patterns* can address a range of psychological perspectives on learning that provide the rationale behind the development of different assessments. They also can vary in their generality and scale.

5.3.1 Psychological Perspective

The *design pattern* structure is neutral with respect to psychological perspectives on learning. Although the framework of *design patterns* serves as a guide for important decisions to make in the assessment design process, the content of these decisions remains open to the perspective and purpose of the assessment. Recalling the ECD model, the goal of assessment is to make inferences about student KSAs. However, the concrete meaning of the inferences, as well as the evidence considered as valid support for the assessment argument, will vary as a function of the designer's beliefs about which KSAs are important and how they are acquired (Greeno, Collins, & Resnick, 1997). For example, from a behavioral perspective in which learning is viewed as the connection between ideas, the correct response to a question may be most important. In contrast, from a cognitive perspective in which learning is cast in terms of the development of mental structures, evidence of students' thought processes during a task would be important. Alternatively, socioculturalists may include a focus on peer interactions and familiarity with practices associated with a domain as a Potential Observation. Any of these perspectives can be represented in a *design pattern*. The *design pattern* structure includes the components of a complete assessment argument, but allows the assessment developer to articulate them. Thus, any psychological perspective or framework used to rationalize the development of a particular kind of assessment may be represented within a *design pattern*.

In attempting to ascertain which psychological perspective shapes a given *design pattern*, it is helpful to examine the Focal KSAs and Potential Observations for clues. This is the case in the NAEP Mathematics DP-2 example where we see emphasis on a cognitive view, because schemata—a cognitive construct—are assumed to be underlying students' work and the Potential Observations we might make of them. Most of the *design pattern* examples provided in this paper take a cognitive psychological perspective, in part reflecting the course content from which they came. In DP-4, attention to contextual and cultural differences suggests a sociocultural perspective has been engaged. Some *design patterns* will remain open to multiple psychological perspectives, such as the *design patterns* for portfolios or those reflecting a high level of generality. In these cases, later decisions about assessment components will determine the perspective according to which resulting assessments will inform inferences of interest.

5.3.2 Level of Generality and Interdependence

Design patterns, constrained by the scope of the selected Rationale and KSAs, will vary with respect to their level of generality. This variation in generality is useful for conceptualizing assessments that not only share overarching KSAs but also involve more specific KSAs. Starting from a highly general *design pattern*, "child" *design patterns*, which may present more specific content knowledge or inquiry skills, can be developed. An example is presented in DP-6D, a hypothetical general domain-free portfolio *design pattern*. *Design patterns* that are more specific instantiations of this general *design pattern* will reflect tighter decision constraints and could be considered nested within the more general *design pattern*.

In addition, a *design pattern* can draw on other *design patterns* rather than restating existing assessment arguments. For example, a *design pattern* can be created at a high level of generality in a domain. Later *design patterns* can be generated by increasing constraints and subsequently be considered "kinds of" the more general *design pattern*. From the opposite direction, samples DP-5 and DP-6 show how general *design patterns* could be educed from more specific ones. Sets of relatively specific *design patterns* derived from the same general *design pattern* might address different areas of a given domain, while retaining some common characteristics of the assessment argument.

5.3.3 Scale

In designing assessments and drawing valid inferences, context is a key consideration. Assessments implemented in local, small-scale settings will have more access to information about individual students than those designed to be administered with little knowledge of individual students. *Design patterns* can be used for a range of settings differing in scale, again by providing a framework for explicating the assessment argument according to the scope of the assessment. For example, in small-scale, local assessments, there are more opportunities to rule out alternative hypotheses in interpreting student work. Thus, *design patterns* geared for this context may include additional detail regarding curricular sequence, familiar notation, or materials. In contrast, *design patterns* targeting assessment of students on a large-scale basis may instead take a more general approach in terms of the materials specified. This level of generality also will influence the validity of the inferences that can be made about students' KSAs. In other words, local assessments are more likely to afford rich information about the student and the instructional context, such that inferences can be made at a finer grain size and specificity than those made on larger scales that are based on drop-in-from-the-sky assessments (Braun & Mislevy, 2004).

The *design patterns* presented in this paper reflect a range of scale. NAEP Mathematics is intended for a large-scale implementation, as reflected in the kinds of KSAs and Potential Observations chosen in DP-2 (e.g., overall ability to gather and use mathematical knowledge [KSAs] and correctness of responses to tasks [Potential Observation]). In DP-5 and DP-6, interrater reliability issues are implied as this form of assessment is implemented in large-scale contexts. In DP-1, a smaller-scale assessment is implied given KSAs such as "understand instructional objectives" and "recognize the progress being made towards these objectives."

6.0 Summary

As illustrated through these examples, assessment *design patterns* are a tool to help assessment developers articulate a coherent assessment argument. By making decisions about assessments explicit in a representational form such as the *design pattern*, the relationships among assessment components become accessible to others. The *design pattern* structure promotes thinking about generating suites of related assessments. By varying the components within a *design pattern*, new types of assessment arguments (and ultimately tasks) can be created.

A *design pattern*, even when complete, still does not provide enough detail to create an actual task. *Design patterns* do not include specific information about how materials will be presented to students in a given task or about how student scores will provide evidence about their proficiency in a domain. For specifying technical information such as this, the PADI system provides another representational form called *task templates*. In a *task template*, Student Model Variables, Measurement Models, Evaluation Procedures, and student materials are described. One or more *design patterns* will be used to inform these *task template* components (e.g., KSAs provide information about the range of Student Model Variables to be assessed, and Potential Observations may inform Evaluation Procedures). PADI Technical Report 3, *An Introduction to PADI Task Templates* (Riconscente, Mislevy, Hamel, & PADI Research Group, 2005), describes the role of *task templates* in assessment design and the structure of *task templates* in the PADI system.

7.0 References

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

Bransford, J. D., Brown, A. L., & Cocking, R. R. (Eds.). (2000). *How people learn: Brain, mind, experience, and school*. Washington, DC: National Academy Press.

Braun, H. I., & Mislevy, R. J. (2004). *Intuitive test theory* (CSE Technical Report 631). Los Angeles: National Center for Research on Evaluation, Standards, and Student Testing (CRESST), Center for Studies in Education, UCLA.

College Entrance Examination Board (CEEB) (2001). *Advanced Placement program course description: Studio art*. New York: Author.

Educational Testing Service (ETS) (2002). *Advanced Placement Studio Art 2002 Scoring Guidelines*. New York: College Entrance Examination Board.

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

Haberman, M. (1991). The pedagogy of poverty versus good teaching. *Phi Delta Kappan, 73*(4), 290-294.

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

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

Mislevy, R. J., Hamel, L., Fried, R., Gaffney, T., Haertel, G., Hafter, A., et al. (2003). *Design patterns for assessing science inquiry* (PADI Technical Report 1). Menlo Park, CA: SRI International.

Myford, C. M., & Mislevy, R. J. (1995). *Monitoring and improving a portfolio assessment system* (Center for Performance Assessment Research Report). Princeton, NJ: Educational Testing Service.

Riconscente, M. M., Mislevy, R. M., Hamel, L., & PADI Research Group (2005). *An introduction to PADI task templates* (PADI Technical Report 3). Menlo Park, CA: SRI International.

Songer, N. B. & Wenk, A. (2003, April). *Measuring development of complex reasoning in science*. Paper presented at the American Educational Research Association (AERA) Annual Meeting, Chicago, IL, Session 69.027.

National Center for Education Statistics, National Assessment of Educational Progress (NAEP). *1996 Mathematics Assessment*. Washington, DC: U.S. Department of Education. Retrieved June 6, 2005, from http://nces.ed.gov/nationsreportcard/itmrls/

White, B., & Frederiksen, J. (1998). Inquiry, modeling, and metacognition: Making science accessible to all students. *Cognition and Instruction, 16*(1), 3-117.

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*