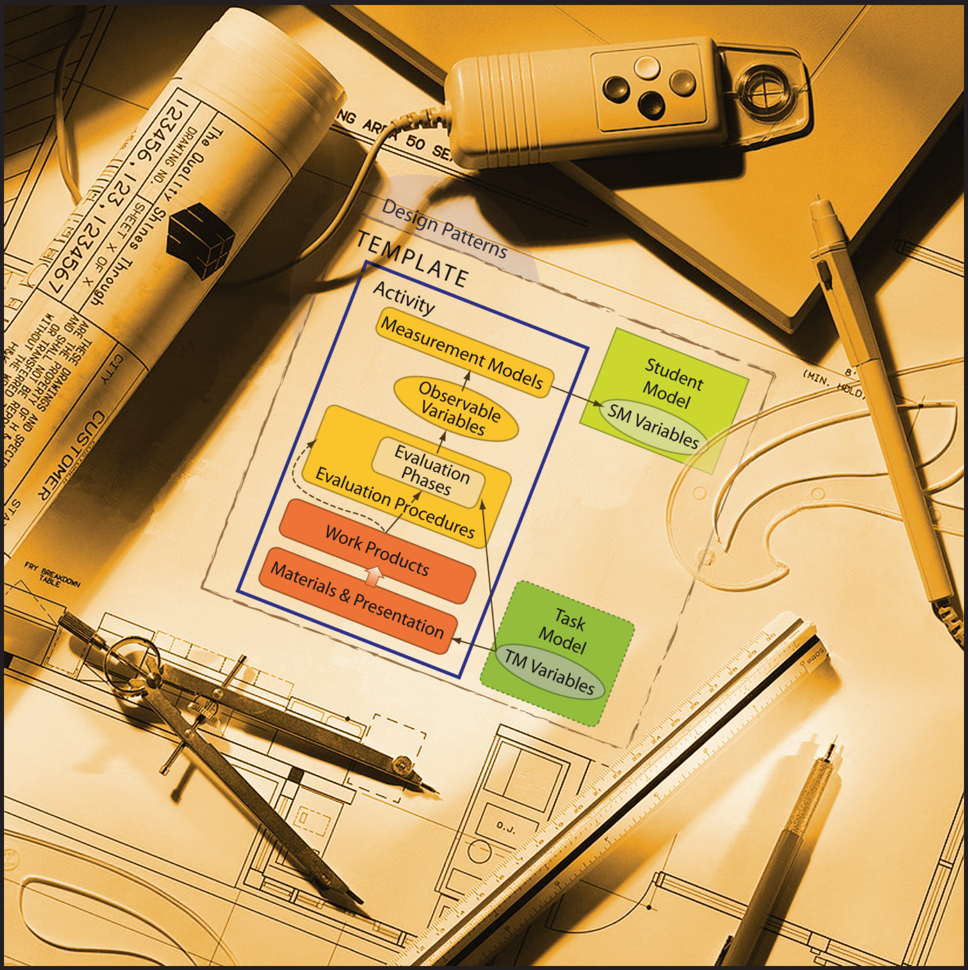


Large-Scale Assessment Technical Report 11| May 2011

Design Patterns: A Tool to Support Assessment Task Authoring

DRAFT

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

Min Liu, University of Maryland
Geneva Haertel, SRI International

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*

Technical Report 11

Design Patterns: A Tool to Support Assessment Task Authoring

Min Liu

University of Maryland

Geneva Haertel

SRI International

Acknowledgments

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

Disclaimer

Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

Design Patterns: A Tool to Support Assessment Task Authoring

Abstract

The ongoing educational reform calls for the design of innovative assessment to validly measure complex content and inquiry skills. This is challenging for task designers. Design pattern, as a tool derived under the evidence-centered design framework, can support their task writing. In this article, we introduced this design pattern tool from aspects of its historical origins, its counterparts in the assessment design literature, and its key attributes/components comprising this narrative form of an assessment argument. After summarizing its benefits to assessment design, we characterize a design pattern resource--PADI Design pattern library and illustrate how to use these design patterns with their technology-enhanced features.

1. Introduction

Assessment is integral to improving education. Sound assessments yield information to inform students' learning, assess individual achievement, and evaluate educational programs. However, designing a valid and fair assessment is challenging. This is particularly true for designing new forms of assessment targeting complex cognitive skills and processes, (Pellegrino, et al., 2001; Gitomer & Bennett, 2002).

Drawing on research under the framework of evidence-centered design (ECD; Mislevy, Steinberg, & Almond, 2003), the NSF-supported Principled Assessment Design for Inquiry (PADI) project developed a new assessment design tool called a design pattern.[1] Design patterns were then used in a follow-on project, called Application of Evidence-Centered Design to State Large-Scale Science Assessment, also supported by NSF.[2] In the latter project, design patterns were to be used by the committee of Minnesota item-writers (mostly current or retired Minnesota science teachers), to support creating tasks that addressed more complex aspects of student science achievement such as inquiry skills, in the context of computer-delivered scenario-based tasks. Design patterns can be used by other assessment developers in operational work as well as by educators and researchers more broadly, not just in science but in other subjects and educational levels, as well as with a variety of task formats. Other projects that have successfully used design patterns include: Domain-Specific Assessment: Bringing the Community College into the Classroom[3], Principled Science Assessment Designs for Students with Disabilities[4]; Alternate Assessment—Mathematics: Applying Evidence-Centered Design to Alternate Assessments in Mathematics for Students with Significant Cognitive Disabilities.[5]; and Alternate Assessment Design Reading (AAD-R): Evidence-Centered Design for Alternate Assessments.[6]

This report describes the rationale and structure of design patterns and provides examples from the projects listed above. First, we introduce its conceptual role in the ECD framework, its historical origins in other disciplines, and its counterparts in the assessment design literature. After describing the key attributes comprising a design pattern and its role as a narrative, non-technical form of an assessment argument, we summarize its benefits to the field of assessment We then characterize the PADI Design Pattern Library, which contains more than 160 design patterns from a wide variety of content areas and grades, and illustrate how to use these design patterns with their technology-enhanced features.

[1] Principled Assessment Design in Inquiry [National Science Foundation, REC-0089122 and REC-0129331].
[2] An Application of Evidence-Centered Design to a State's Large Scale Science Assessment [National Science Foundation, DRK-12 initiative, DRL-0733172].
[3] Domain-Specific Assessment: Bringing the Community College into the Classroom [Institute of Education Sciences, US Department of Education, R305A080109].
[4] Principled Assessment Science Assessment Designs for Students with Disabilities [Institute of Education Sciences, US Department of Education, R324A070035].
[5] Applying Evidence-Centered Design to Alternate Assessments in Mathematics for Students with Significant Cognitive Disabilities [US Department of Education, Contract to State of Utah, 09679]
[6] Alternate Assessment Design Reading (AAD-R): Evidence-Centered Design for Alternate Assessment [US Department of Education, S368A090032].

1.1 An exemplar design pattern and two derivative tasks

Table 1 offers a condensed, static, form of a design pattern. It summarizes the contents of a design pattern created to support writing assessment tasks that center on scientific observational investigations, and was developed in the project with the Minnesota Comprehensive Assessment (MCA) in Science. As will be seen in Section 6, the complete design pattern has additional details, explanations, interactive links, and dynamic views that make it easier for a task designer to use than the static version. The condensed form shown in Table 1 serves to illustrate the essential contents of a design patterns and the kinds of information it provides. Figures 1 and 2 present two assessment tasks that differ in terms of item format and content domain, but were both motivated by this design pattern. These items appear as scenes from multi-stage "storyboard" tasks in the Minnesota Comprehensive Assessment in Science (MCA II[7]). As we will see, the evidentiary arguments they meant to support can be traced back to this design pattern.

Insert Table 1 about here

Insert Figure 1 about here

Insert Figure 2 about here

----

[7] Minnesota Comprehensive Assessments Series II (MCA-II): Test Specifications for Science. http://education.state.mn.us/mdeprod/groups/Assessment/documents/Report/006366.pdf

1.2 Assessment as an Evidentiary Argument

More than half a century ago, Cronbach and Meehl, (1955) noted that an assessment could be structured in terms of argument for the purpose of validation. In fact, although assessments might look quite different or be used for different purposes or in different contexts, they share the property that assessment, by its nature, is always a process of reasoning from limited evidence of what students say, do, and make in particular settings, to claims about what they know and can do more broadly (Messick, 1994).

1.3 Operationalizing the Assessment Argument in ECD

The view of assessment as an evidentiary argument is the foundation of the evidence-centered design (ECD) framework. The goal of ECD is to develop a coordinated and coherent assessment or assessment system by fleshing out an assessment argument across five layers of work that begin with the analysis and organization of the conceptual domain to be assessed and culminate in the delivery, scoring and reporting of the assessment results to stakeholders. The design pattern tool was designed to fit within the ECD framework and to support assessment designers in the second layer of work, referred to as domain modeling. To provide the theoretical grounding of the design pattern tool and its interplay with the other elements in the assessment development process, we briefly introduce ECD and the five layers.

1.3.1 Layers in the Design Process

Evidence-centered assessment design (ECD) was first proposed systematically by Mislevy, Steinberg and Almond in 2003. It provides principles, patterns, examples, common language and knowledge representations for designing, implementing and delivering educational assessment (Mislevy & Haertel, 2006). This structured framework views assessment as an argument and aims to explicate the assessment argument underlying a task, and thus enables designers to more efficiently control the elements and processes of assessment design.

Figure 3 and Table 2 present each of the five layers of ECD. Reading Figure 3 from the top down, we can see the successive refinement and reorganization of knowledge about the content domain and the purpose of the assessment being implemented—from a general substantive argument to an increasingly specific argument that identifies the elements and processes needed in its operation. Different experts or parties may be carried out this work at different stages of the assessment process. The ECD framework provides a common language to help them communicate more efficiently. Table 2 characterizes each layer is in terms of its role in the assessment enterprise, its key concepts and entities, and knowledge representations and tools that assist in achieving each layer's purpose. A brief introduction to each layer is as follows.

Insert Figure 3 about here

Insert Table 2 about here

As the first stage, *domain analysis* is about marshaling substantive information about the content domain. Assessment designers use this substantive information to understand the knowledge, skills, and abilities people use in a domain of interest, the representational forms they use, characteristics of good work, and key features of situations. All of this information has important implications for assessment design, although usually most of the sources were neither originally created to support assessment nor presented in the structure of argument. For example, the National Science Education Standards (NRC, 1996) and some state science standards are good content sources for designing a large-scale science assessment. A thorough analysis of the content domain of interest is prerequisite for generating a design pattern which is the product of the work that is conducted in the next layer of ECD. For more specific examples of the work conducted in the domain analysis to prepare for creating a design pattern, see the discussions of the development of a design pattern for observational investigation (Mislevy et al., 2009)[8] and a design pattern for experimental investigation (Colker et al., 2010)[9].

Domain Modeling. In the *domain modeling* layer, information identified in *domain analysis* is organized along the lines of the assessment argument. Without getting tangled in the technical details of assessment design, this layer directs researchers to clarify what is meant to be assessed, and how and why to do so. *Design patterns* (DPs), as a tool developed as part of the original PADI project (See Mislevy, et al. 2003[10]) to support work in the *domain modeling* level of ECD, helps the assessment designer think through the key elements of an assessment argument in narrative form. Key attributes of *design patterns* are provided in a subsequent section of this report.

The three subsequent layers of the ECD framework concern fleshing out, implementing, and delivering assessment tasks that build on the arguments first sketched in domain modeling. We review them briefly to show the connection between this narrative form of assessment arguments that *design patterns* represent and the more operational details of assessment and ECD . The reader is referred to Almond, Steinberg, and Mislevy (2002) and Mislevy and Riconscente (2006) for further discussion on these layers.

Conceptual Assessment Framework (CAF). The CAF concerns technical specifications for operational elements. An assessment argument laid out in narrative form at the *domain modeling* layer is here expressed in terms of coordinated pieces of machinery such as measurement models, scoring methods, and delivery requirements. The commonality of data structures and reusability of the central CAF models offer opportunities to bring down the costs

[8] http://ecd.sri.com/downloads/ECD_TR2_DesignPattern_for_ObservationalInvestFL.pdf
[9] http://ecd.sri.com/downloads/ECD_TR8_Experimental_Invest_FL.pdf
[10] http://padi.sri.com/downloads/TR1_Design_Patterns.pdf

of task design, which is especially important for computer-based tasks. In the CAF layer, the PADI project produced another ECD associated design tool, *task templates,* to guide the creation of families of tasks at more detailed level than design pattern (Mislevy & Riconscente, 2005[11]; Baxter & Mislevy, 2005[12]). .

Assessment Implementation. The fourth layer, *assessment implementation*, includes activities carried out to prepare for the operational administration for testing examinees, such as authoring tasks, calibrating items into psychometric models, piloting and finalizing scoring rubrics, producing assessment materials and presentation environments, and training interviewers and scorers, all in accordance with the assessment arguments and test specifications created in the CAF.

Assessment Delivery. The final layer, *assessment delivery*, includes activities in presenting tasks to examinees, evaluating performances to assign scores, and reporting the results to provide feedback or support decision making. Development design tools for the last two layers were not been the focus of the original PADI project. Readers can refer to Mislevy and Haertel (2006) for more details about kinds of tools produced by other research projects for these two layers.

1.3.2 A closer look at Domain Modeling

Work in the first layer, *Domain Analysis*, might appear in many forms, and usually is carried out by content experts rather than assessment specialists. The information gathered in this layer is important to assessment design but is not organized in terms of assessment arguments. The third CAF layer mainly concerns how to specify the nuts and bolts of an assessment using the technical language of assessment specialists and psychometricians. There is a vast difference in kind between the work conducted at these two layers--work done by content experts who have expertise in the domain of interest or in teaching and learning, versus the work done by experts in the areas of psychometrics, internet-based delivery systems, database structures, and other technical work to create and deliver assessments. The domain modeling layer facilitates the transition from substantive knowledge about a target construct and the embodied assessment argument in the CAF layer. A design pattern supports this transition by specifying the components of an assessment argument at a high level in a narrative form that is accessible to both the content experts, as it serves to organize their work in terms of assessment arguments, and technical specialists, as it presages the content of the specialized structures they will craft to implement the argument.

Domain analysis is prerequisite step for domain modeling. After information about the assessment domain is collected, possibly including beliefs, theories, research, subject-content expertise, instructional materials, other assessment exemplars, and so on, all of these kinds of information are orchestrated in terms of a substantive assessment argument in narrative form. More specifically, all of this information is reorganized to embody a reasoning process, from which observations about what examinees say, do, or make in a particular circumstance are used to draw inferences about what they know, can do, or have accomplished more broadly

[11] http://padi.sri.com/downloads/TR9_ECD.pdf
[12] http://padi.sri.com/downloads/TR5_IDFramework.pdf

(Messick, 1994; Mislevy, et al., 2003). Clearly, the role of domain modeling is to express the assessment argument in a manner that is readily accessible to those charged with designing the assessment without heavy reliance on psychometric and statistical concepts. Therefore, the work in domain modeling is more of a conceptualization process of a content domain rather than technical implementation of task design.

2. Design Pattern as a Conceptualization Tool in Domain Modeling

Design patterns are proposed as a tool to assist the conceptualization process in the domain modeling layer. The information collected during the Domain Analysis is used to articulate the key components of the assessment argument in a narrative form within the Design Pattern. In addition to the Design Pattern tool, there are other assessment tools and conceptualization methods that can also support aspects of domain modeling processes, such as Wigmore and Toulmin diagrams, Kane's assessment diagrams, "big ideas," shells, item forms and assessment blueprints, and task templates. A brief overview follows.

2.1 Other assessment tools supporting the conceptualization process

- *Toulmin diagrams*, introduced by British philosopher Stephen Toulmin (1958) provide terminology to talk through the structure of a simple argument, which usually is constituted of claims, data, warrants, backing, and alternative explanations. More details will be provided in the context of an assessment argument in section 2.4. To accommodate the complexity of evidentiary reasoning in practice, John Henry Wigmore (1937) developed a system for charting the structure of arguments with multiple propositions, chains of reasoning, dependent claims, and various data in the context of a judicial analysis. This idea sheds some lights on the design issue in the assessment field as well as other disciplines (Bachman, 2005; Kane, 1992, 2006; Messick, 1989).

- *Kane's diagram* of assessment argument can be considered as an assessment tool for domain modeling. It explicates assessment as evidentiary argument by identifying its underlying structure and observable elements and process. Measurement models can be considered as one aspect of a warrant in an assessment argument (Mislevy, 1994)

- *Item forms* were first used by Hively, Patterson, and Page (1968) to write criterion-referenced arithmetic achievement test items, in which they defined a well-specified content domain and general item generation rules for the range of tasks. As Osburn (1968) summarized, an item form has three characteristics: "1) it generates items with a fixed syntactical structure; 2) it contains one or more variables elements; and 3) it defines a class of item sentences by specifying the replacement sets for the variables elements" (p.97). This pioneering idea has been widely accepted and applied in criterion-referenced tests (e.g., Dziuban & Vickery, 1973; Meisner, Luecht, & Reckase, 1993).

- *Item shells* are similar to item forms as tools for developing assessment tasks. The idea of item shells originated from the need for formalizing a procedure of writing multiple-choice and short answer items. Shells can be considered as "hollow" frameworks or templates with syntactic structure, making the generation of a family of similar items more efficient and systematic (Haladyna & Shindoll, 1989).

- An *assessment blueprint* usually is constructed to be aligned with assessment standards and benchmarks and supports the development and use of an entire assessment, not just a single family of items, as is the case with Design Patterns,. An assessment blueprint usually appears in the form of a test matrix or table of specifications (Garden & Orpwood, 1996). Although assessment blueprints vary in complexity, a general one, taking classroom assessment as an example, contains information such as constructs to be tested, the number of questions per construct, cognitive difficulty level associated with each question, and so on (Newble, Hoare, & Elmslie, 1981).

- To ground assessment design, Chung, Delacruz, Dionne, and Bewley (2003) developed an *ontology* consisting of clear expressions of key concepts, the links among these concepts, and the constraints governing these links in the domain of rifle marksmanship. Due to their clear organization of all the relevant entities and their relations, they were able to link assessment and instruction via Bayesian networks. A related method called *big ideas* maps the relationships among the core principle and key concepts in a domain. The Assessment Design and Delivery System (ADDS), as a web-based classroom assessment design tool developed by National Center for Research on Evaluation and Student Testing (CRESST), employed this idea in its on-line formative assessment design system. ADDS has been found to focus more on conceptual knowledge and can be used to create coherent tasks addressing critical ideas (Niemi, et al., 2005; Vendlinski et al., 2005).
- *Task templates* are another assessment design tool developed as part of the original PADI project (Mislevy & Riconscente, 2005[13]). Like Design Patterns, templates are used to support task design, but at the level of the CAF, not the domain modeling level. Task templates articulate the conceptual assessment argument at more technical level (Riconscente et al., 2005).

2.2 Limitation of these assessment tools

The Bachman, Kane, and Messick work building on Toulmin and Wigmore, along with the “big idea” approach, build on the idea of evidentiary argument, while item shells, item forms, and blueprints are patterns with fixed or variable elements or both for generating families of tasks. Task templates utilize both ideas in the detailed technical work so that this tool requires understanding of psychometric and statistical concepts, which usually are not accessible to general task designers. These assessment development tools arose in the setting of familiar tests, however, and have not been focused on the challenges on new forms of assessment. The National Research Council noted in *Knowing what students know* (NRC, 2001) that reform of educational assessment should provide a better understanding of what and how students learn, brought by a synthesis of progress in cognitive psychology, measurement and statistical modeling, and information technologies. With these new advances, innovative assessment can assess more complex skills and abilities, such as multidimensional proficiencies and complex, multi-step performances. Each assessment design tool listed above brings its conceptual or practical advantage in supporting assessment development. What is lacking, however, is a tool that organizes assessment design around the substantive research on learning in content domains in a way that grounds an evidential assessment argument, while providing a friendly structured pattern with both fixed and viable features for generating broad families of tasks. That is, the tool should not be limited to specific forms of assessment, but instead focus on the evidentiary patterns that must be addressed in assessing newly-understood capabilities. We will demonstrate that the assessment design pattern is such a tool.

[13] http://padi.sri.com/downloads/TR9_ECD.pdf

2.3 A Short History of Design Patterns

The idea of design patterns is not brand new. For example, the use of assessment design patterns is similar to that of Georges Polti's (1916) 36 narrative structures for plots in that both design patterns and narrative structures provide basic structure and useful stimuli for the purpose of design. Every one of Polti's dramatic situations sketched out a short situation description with the title of the plot pattern, the main characters, and variants in the plot that storytellers can build around – much as George Lucas did nearly a century later in his *Star Wars* movies.

Formal work on design patterns is marked by the publication of architect Christopher Alexander and his colleagues in *A pattern language: towns, buildings, construction* (Alexander, Ishikawa, & Silverstein, 1977). Alexander described a practical architectural system in a form to empower anyone, not just professionals, to understand design principles at any scale in practical, recurring situations. A design pattern in architecture is a formal way of documenting a general reusable solution to a commonly occurring design problem in a given context. Very often the design problem arises from "conflicting forces," such as the conflict between the desire for a sunny room and the desire for the room not to overheat on summer afternoons. A design pattern won't give a deterministic answer to this problem, but instead it will proposes a range of possible values to guide the designer toward the best solution in their particular situation, after considering other relevant factors, such as how much direct light is needed, how much the materials costs, and so on.

The idea of design patterns has been adapted to other engineering fields, especially software design. The publication of "*Design Pattern: Elements of Reusable Object-Oriented Software*" (Gamma et al., 1994) sparked the use of design pattern in computer science. In this book, each design pattern identifies recurring entities in an object-oriented programming, their roles and collaborations in a particular design problem, and when they can be applied in terms of design constraints. More recently, design patterns were extended to pedagogical use to improve the quality of teaching (Jones, et al., 1999; Frizell & Hubscher, 2002; Mor & Winters, 2007) and to assessment design in different content domains (Mislevy & Haertel, 2006; Mislevy, et al., 2009**;** Colker, et al., 2010).

2.4 Toulmin diagram for assessment arguments

This section sketches out the structure of an assessment argument, which every assessment task reflects, and around which design patterns are structured (see Mislevy, 2003, 2006 for further discussion on assessment design arguments). The attributes of assessment design patterns are specified and their relationship to the elements of an assessment argument is presented.

Toulmin (1958) provides a useful schema for the general structure of arguments. Figure 4 adapts his terminology and representations to educational assessment arguments (Mislevy, 2003, 2006). A series of logically connected claims is supported by data, via warrants, subject to alternative explanations. The claims concern aspects of proficiency that students possess — i.e., what they know or can do in various situations. Data are required to support claims. In the case of assessment, data consist of (1) students' behaviors in particular task situations, (2) the features of task situations, and (3) other relevant information about the relationship between the student and the task situation (e.g., personal or instructional experience).

The arrow going to the claim represents a logically reasoned inference by means of a warrant. The warrant is the logic or reasoning that explains why certain data provide appropriate evidence for the claims. The primary source of the warrants is the underlying psychological conceptualization of knowledge and its acquisition — i.e., a psychological perspective that shapes the nature of claims that assessments aim to make and of the data that are needed to evidence them. *Alternative explanations* for poor performance include deficits in the knowledge or skills that are needed to carry out a task but are not focal to the claims. Figure 4 provides the basic structure of the assessment argument that every assessment must embody, and thus every design pattern must reflect. The key attributes in a design pattern help an assessment designer flesh out this structure to create particular tasks based on a clear argument structure. These design patterns attributes are introduced in following section.

Insert Figure 4 about here

2.5 Key attributes in a design pattern

A design pattern helps task designers think through substantive aspects of the assessment argument. It does not provide deterministic answers for exactly what to include in tasks or how to score them, but rather it offers options on features to include in tasks in the targeted area and rationales for their roles in the assessment argument. Design patterns smooth the transition from content domain analysis to technical implementation of task design in the CAF—for example, helping to fill the gap between academic content standards and specific assessments tasks. The experience and thinking captured in a design pattern provides shared information across applications, such as large-scale and classroom assessment, and instruction, and research. Although creating a design pattern may seem to be a time-consuming job in the beginning, it can save time and energy in the long run by capturing design rationales in a re-usable and generative form, just as design patterns have been shown to function in architecture and software engineering.

A *design pattern* consists of attributes that are associated with components of an assessment argument, as shown in Table 3. They correspond to an assessment argument by identifying the knowledge, skills, or abilities (KSAs) about which assessors want to make a claim, the kinds of data that provide evidence about student acquisition of that KSA, and features of task conditions that can enable students to produce the evidence.

2.5.1 How design pattern support thinking about the assessment argument

Table 3 lists a design pattern's key attributes, their definitions, and associated assessment argument components. An attribute-by-attribute discussion follows describing how design pattern attributes support thinking through the assessment argument.

Insert Table 3 about here

A design pattern is organized around *Focal KSAs*. "KSA" stands for "Knowledge, skill, or ability," a term borrowed from industrial/organization psychology. It is used here simply to stand for whatever aspects of students' proficiency are of relevance, not as a statement about the nature of the proficiency. The structure of a design pattern is meant to be sufficiently general that Focal KSAs can be cast in terms of any view of capabilities, be it behavioral, trait, information-process, or sociocultural (Mislevy, 2003). That said, a given design pattern will exhibit a particular stance on the KSA it is meant to support, in order to help designers create tasks from that perspective. As an example, the Observational Investigation design pattern in Table 1 reflects the "science as inquiry" stance taken in the National Science Education Standards (NRC, 1996).

Focal KSAs will be involved in the claims a task is meant to support, although there may be other KSAs that are included in the target of inference. For example, if proficiency with observational investigations is our target construct, the content knowledge of which scientific models or principles, as applied in what kinds of contexts, might also be part of what we want to assess. We will say a bit more about this shortly in connection with *Additional KSAs*.

Associated with Focal KSAs are *Characteristic Features* of Tasks, which are intended to invoke evidence about the Focal KSAs. For example, in the "Finches on the Galapagos Islands" task presented in Figure 1, the assessment designer presents a table in which data collected about the finches is represented. This table is a characteristic feature of tasks designed to measure observational observation and is used to elicit evidence about the Focal KSA, "Ability to formulate conclusions, create models and appropriately generalize results from observational, non-experimental results,

The *Rationale* provides background into the nature of the Focal KSAs, and the kinds of things that individuals do in given kinds of situations that provide evidence of the Focal KSAs.. It contributes to the Warrant in the assessment argument. The rationale for the Observational Investigation design pattern is drawn from research in science education and the philosophy of science, key references for which are noted as well. The warrant is therefore that "if a student is able to carry out the reasoning required in observational investigation in the given situation, given sufficient familiarity with the science content, she would probably exhibit the kinds of performance noted in the 'potential observations' listed in the design pattern."

Additional KSAs play multiple roles in assessment design. Task designers need to think about which ones are necessary to include as targets of inference (i.e., construct relevant with respect to validity) and which ones are not (i.e., construct irrelevant) and might result in invalid inferences. The Additional KSAs that assessors *do* want to include as targets of inference are part of the claim. For example, if an assessment task intends to test the knowledge of Mendel's laws as well as being able to formulate a model in an investigation, both the content knowledge and the skills of model formation are targets of inferences. The Additional KSAs that assessors *do not* want to include as targets of inference introduce alternative explanations for poor performance, which would blur the claim assessors want to make for students.

This latter role of Additional KSAs is especially important for assessing special populations (Hansen, Mislevy, Steinberg, Lee, & Forer, 2005). To avoid the influence of irrelevant factors such as poor vision or attention deficit disorder, task designers accomplish their work using the principles of universal design for learning and kinds of necessary accommodations for students in need. The Additional KSAs are related to Variable Features of Tasks and Work Products. Variable Features can be manipulated by task designers according to their specific needs for additional KSAs (Haertel, et al., 2010). Specifically, a design pattern can be built to offer specific advice about how to support or circumvent particular construct-irrelevant Additional KSAs by design choices about what information is presented to an examinee and how it is presented, how the examinee interacts with the tasks, and how responses are given.

As noted above, *Characteristic Features* of Tasks help task designers think about critical data concerning the situation by calling to attention to features that should be in the task situation in order to obtain evidence about Focal KSAs. Variable Features of Tasks also help task designer think about data concerning the situation; but they are features of tasks that are variable. For example, some variable features can be used to increase or decrease the difficulty of a task. Others can bring in or reduce demand for Additional KSAs, which is an effective way to avoid alternative explanations as explicated in the Toulmin diagram. Some Variable Features of Tasks help designers be aware of whether and how they might wish to match features of tasks with the characteristics of students such as their interests, familiarity, and previous instruction.

Potential Work Products help designers think about what they want to capture from a performance – product, process, constructed model, written explanation, etc. It can also call attention to demand for Additional KSAs required to produce work, such as a specialized computer program that allows constructed responses but also demands knowledge of how to use it. *Potential Observations* highlight the qualities of Work Products that contain evidence about the Focal KSAs. They look ahead to task scoring, to produce data concerning the performance. *Potential Rubrics* are algorithms/rubrics/rules for evaluating Work Products to get this data concerning the performance.

2.5.2 The relationship of the DP attributes to the models of the CAF

In architecture, having a design patterns ready when anticipating the construction of a new building is important and necessary, because it provides a general plan for constructing a building. In addition to this rough plan, surely more details are required to start this construction work. Similarly, in the field of assessment design, the creation of the design pattern in the domain modeling stage provides a high-level narrative that describes the components of the assessment, but many more technical details must be specified in the subsequent CAF layer, possibly under the support of another assessment design tool, the task template. A task template explicates the connections among assessment arguments for a family of assessment tasks and so it can be viewed as a "pre-blueprint", upon which different sets of blueprints/task specifications can be generated later. The viable options provided by a task template are to be fixed in task specifications, which is parallel to a specific blueprint detailing the construction of a particular building. Likewise, task specifications set forth the technical requirements for the development of particular tasks; task templates set forth the technical requirements for the development of a family of tasks.

The basic structure of a task template includes three ECD design objects: student, evidence, and task models as marked in blue, yellow, and pink respectively in Figure 4. To achieve a coherent and workable assessment, the three basic models must be presented and coordinated in an assessment.

Using the analogy of constructing a building, each object in Figure 5 can be viewed as a building block in a "construction kit" for assessment. In addition, each object has its particular modes of connecting to the other objects as well as to the overall object (i.e., the building or assessment). For example, windows must be placed in walls, which can hold multiple windows, and both of them are necessary component of constructing a building. It is often greatly helpful to make all the explicit and implicit rules and constraints clear in a workable blueprint.

As stated above, a design pattern lays out the conceptual foundation to carry out technical work in the models of the CAF. The connections between design pattern attributes and the three models of the CAF are ticked off in different colors in Table 4, in which the leftmost column lists the key attributes of a design pattern introduced earlier, the top row presents the three CAF models, and the connections between them are checked to clarify how these attributes inform the specification of the models.

Insert Table 4 about here

A student model is a collection of one or more student-model variables and a probability distribution over them. These variables describe some aspects of examinees' knowledge and skills of interest. Starting from an uninformative or population-level prior distribution for these student-model variables, psychometricians update their probability distributions based on students' observed performances on the KSAs – or, in the simpler and more common case, estimate the values of the student model based on a student's responses. Student-model variables mean to support claims about the focal KSAs. In this sense, Focal KSAs in domain modeling layer are precursors to the student-model variable(s) in the CAF layer. Student model variables might be gauged simply by a total score (e.g., a mix of KSAs that a set of tasks intended to assess) or a multivariate psychometric model (e.g., multiple aspects of knowledge or skill are intended to be evidenced in a collection of tasks that may require them jointly in various mixes). Additional KSAs might or might not be part of the student-model variables, depending on whether they are the assessment targets of inference. Assessment designers make these decisions about whether and how the student model variables will be combined. Technically speaking, student model variables are modeled as unobserved variables in a psychometric model so that their values cannot be observed directly but must be inferred from observable variables, which come from scoring students' performance. The mechanism through which this takes place appears in evidence models.

The evidence model concerns the ways in which students' performances constitute evidence to support claims. The Observable Variables appearing in the evidence model characterize features of student work, whether correctness, effectiveness, misconception, or whatever is salient in students' work.

The evidence model consists of two components, evaluation rules and statistical models. The evaluation component identifies the key features of a student's work product and expresses them as values of observable variables. Potential observations in the design pattern offer suggestions to task designers about which and how many observable variables might be identified in work products to constitute evidence for a particular focal KSA. Another attribute of design patterns, potential rubrics, suggests evaluation rules or algorithms for scoring students' performances. The statistical portion of the evidence model explicates how each observable variable depends on one or more student-model variables and then bridges one of the key connections from data to claims in the assessment argument. In an Item Response Theory (IRT) application, for example, the connections between the student model variables and potential observations are expressed by a particular mathematical function for the probability of observed variables given student-model variables (e.g., Embretson & Reise, 2000; Hambleton, Swaminathan, & Rogers, 1991). Variable features of tasks, as attributes controlled by the task designer, can be varied to change the difficulty of test items or to determine which student-model variables are evidenced by a given task.

The task model provides a framework for constructing and describing aspects of situations in which students act. Thus, the task-model variables identify the key features of stimulus materials, tools or other affordance made available to students, or other aspects of the environment, or interactions among them. In the ECD framework, Characteristic Features and Variable Features suggest to the designer ways to specify the task model. Characteristic Features are not variable across tasks but necessary features that must in some form be embedded in the assessment tasks to invoke the evidence regarding a focal KSA. Different from Characteristic Features of tasks, Variable Features of tasks can be varied across the assessment tasks for different assessment purposes. The third design pattern attribute linked to task models is the Potential Work Products. This attribute proposes what assessors want to capture from a performance so that it tells which and how many Work Products are to be captured in tasks.

3. Benefits of Using Design Patterns

The previous section presented the theoretical background of design patterns, their key attributes, and relationships with other ECD design objects. This section summarizes the benefits of using design patterns in assessment.

First, as introduced in section 1.3.2, design patterns facilitate the transition from knowledge about the domain to the objects and processes used in an operational assessment system. By means of this tool, content specialists organize domain information in structured ways to be understood and used by assessment specialists for the purpose of implementing more detailed technical work in subsequent assessment task development. It keeps the designer's focus on the conceptual level rather than moving too quickly to the technical elements – the nuts and bolts and bolts of implementation – although these will eventually be detailed. Understanding and using design pattern does not rely on deep understanding of complex psychometric concepts; rather the plain language of design patterns facilitates communications between content experts and assessment specialists who will carry out the technical aspects of implementation and analysis.

Secondly, design patterns increase the validity of an assessment by explicating a structured assessment argument. When using a design pattern all the key entities /attributes about what complex knowledge, skills or other attributes should be assessed, what performance/behaviors can reveal those constructs, and what tasks or situations can invoke those performances addressing those construct of interest are set forth. The overall design pattern and each of the attributes keeps the assessment designer focused on the proficiency (or construct) of interest. This is particularly useful when assessing hard-to-assess constructs (e.g., inquiry skills) because it helps assessors aim at the right target and avoid getting confused by the complexity of the constructs. Further, because of the discipline imposed in explicating the structured argument, the coherence of the assessment components is increased and thereby the validity of evidentiary reasoning and the assessment design is also increased (Cronbach & Meehl, 1955).

Thirdly, design pattern facilitate decision-making for task designers in the process of assessment design. This design tool clarifies the explicit and implicit constraints and resources that will impact the design, development and delivery of the actual assessment tasks. Instead of providing deterministic answers, design pattern lay out the key components and logical relations among them the key features and variable aspects of situations to help task designer make decisions regarding the key features of tasks to elicit the constructs of interest, and other variable features that could be manipulated for assessment purpose.

Last but not least, design pattern affords flexibility in the design process in several ways:

- Psychological Perspectives. Assessors always want to make an inference about what students know and can do from some perspectives on learning and instruction. Generally speaking, several perspectives are accepted and used in practice: 1) a trait/differential perspective which focuses on a common and stable trait of a person (e.g., Messick, 1989), 2) a behaviorist perspective which focuses on students' learning behaviors (e.g., Krathwohl & Payne, 1971), 3) an information-processing perspective which examines problem-solving in terms of the capabilities and limitations of human cognition and memory (e.g., Newell &

Simon, 1971), and 4) a sociocultural perspective which stresses how knowledge is conditioned and constrained by technologies, information resources, and social situations (e.g., Greeno, Collins, & Resnick, 1997). All of these perspectives can shape an assessment—from the assessment's purpose, students' cognitive patterns and associated actions, to what key features of tasks to use to elicit the evidence for the targeted inference and what students might do or say to constitute the evidence. As a structure, a Design Pattern is able to support work under all these different psychological perspectives because it can model unobserved constructs arising from either one of these psychological perspectives. The contents of design patterns created under the different perspectives might look quite different in terms of the nature of the KSAs, the characteristic and variable features of tasks, and potential observations, for example, but no matter what the perspective, the resulting tasks would support an argument framed in that perspective.

- Generality. Without involving domain-specific knowledge and facts, a design pattern can be built to conceptualize the key elements and processes in complex construct of interest (e.g., inquiry) that often are considered difficult to assess. Thus, design patterns can be generalized to a variety of content domains, grade levels, or populations (e.g., regular education students or those with special needs). For example, the observational design pattern displayed in Table 1 provides a general design space that crosses different disciplines and can be used to generate a family of assessment tasks that focus on reasoning skills in observational investigation. This pattern can be used in conjunction with a variety of content domains at different grade levels. Other examples in the current library of this type include experimental investigation, design under constraints, and model-based reasoning, all of which can be used to support task design in a wide range of domains. Furthermore, although these design patterns were created to support generating assessment tasks and items for large-scale science assessments, the experience and thinking captured in a design pattern can be shared across other applications, such as classroom assessments, instruction, and research.

- Interdependence and Scale. Each design pattern can be used independently; but they could be interdependent and used jointly to fulfill an assessment purpose. Taking observational investigation design pattern as an example, this particular design pattern is related to the design pattern for experimental investigation, because both of them aim to address fundamental methods for scientific inquiry. So task designers could use both design patterns in a complementary way. Another typical example to illustrate this kind of dependent relationship is the suite of design patterns for model-based inquiry, which focuses on six aspects of model-based reasoning skills (Mislevy, Riconscente, & Rutstein, 2009). Each aspect of these reasoning skills could be an independent theme of a design pattern: a design patterns for model formation, model use, model elaboration, model articulation, model evaluation, and model revision. These design patterns may be used separately or conjunctively. This example also demonstrates that design patterns can be generated at different grain sizes. Assessors can decide at what detailed level the KSAs should be addressed in accordance with their assessment purpose.

4. PADI Library of Design Patterns

Current design pattern use is best captured by scanning the PADI Design Pattern Library, which contains about 160 design patterns. These Design Patterns were compiled from some ten assessment-related projects connected with the National Science Foundation, the US Department of Education, and commercial entities. The web-based library of design patterns for each project can be located using one of several URLs presented in Appendix A. These design patterns contained in the library represent proficiencies (constructs of interest) from the content areas of science, mathematics, economics, language arts, management/business, and second language learning; In addition, there are a handful of design patterns not specific to a content area. These design patterns were developed for assessments of students at grades 3-5, 6-8, 9-12 and at the post-secondary levels; and for a range of student populations, including grades 5-16 students in general education classrooms, grades 5-8 students with disabilities who receive instruction in the regular classroom, and grades 3-12 students with significant cognitive disabilities who take alternate assessments. As stated before, the design patterns are organized around themes, models, and cognitive and psychological processes, rather than surface features or formats of tasks. These design patterns have been used to build assessment tasks and items that employ a variety of response formats, including discrete multiple choice items, scenario-based items, portfolio-based assessment tasks and performance tasks. Table 5 presents the current number and types of design patterns available in the PADI web-based library. These design patterns have been classified by content area and type of pattern as well as the project in which they were developed.

The current online design pattern library contains 162 patterns to guide the design and development of hard-to-assess content. Table 5 presents current design patterns in terms of the type of design pattern, the number of design pattern in PADI online library associated with each type and content areas. The library of design patterns could be classified into five broad categories according to their themes: 1) educational standards-based, 2) unifying themes/inquiry, such as those identified in the NSES (1996), 3) big ideas within disciplines (Niemi, 2005), 4) learning progressions that reflect nested levels of KSAs based on disciplinary content that reflects increasingly sophisticated levels of learning, and 5) language proficiency. Definitions for each type of pattern are provided as follows.

Insert Table 5 about here

The education standards-based design patterns are those that were developed to address state or national standards and benchmarks in particular content areas. These standards may include content, processes and practices, and hybrid design patterns that include both content and processes. The next set of design patterns, unifying themes/inquiry were inspired by the National Science Education Standards (NSES) (1996) and include themes, such as model-based reasoning, systems thinking and other inquiry processes. Design patterns inspired by the NSES unifying themes can be applied across many disciplines in the natural and social sciences. Design patterns in the category, big ideas within disciplines, were developed to address propositional knowledge and key conceptual understandings unique to a particular discipline (e.g., supply and demand in economics or photosynthesis within biology). Design patterns that address learning progressions make use of nested Focal KSAs and represent increasingly sophisticated levels of understanding within a discipline; these levels of understanding could be focused on either a disciplinary topic or practice. The design patterns in the category of language proficiency address targeted aspects or situations of language use. Examples include the use of language for special purposes, contextualized listening skills, and content-specific story-retelling. Appendix B contains an example of each type of design pattern.

5. A Detailed Look at the Observational Investigation Design Pattern

Taking the design pattern of observational investigation in Table 1 as an example, this section walks through the content of the key attributes to provide the reader a sense of what they look like in the context of a particular construct. For more details about this design pattern, see “A Design Pattern for Observational Investigation Assessment Tasks” (Mislevy, et al., 2009).

Using Table 1, the Overview and Use attributes of this design pattern explain briefly that it is meant to support the writing of assessment tasks (such as storyboards and items in MCA) that address aspects of reasoning in observational investigation in science. This material corresponds to the warrant in an assessment argument.

The Focal KSAs are the primary attribute of a design pattern and the targets of inferences that assessors aim to make in an assessment, in this case concerning some aspect(s) of proficiencies in observational investigation. The KSAs encompass at a more overarching level the indicated benchmarks (MCA-II and MCA-III) shown in the table. These benchmarks themselves can be found in Mislevy et al. (2009) or the Minnesota Test Specification for Science[14].

The focal KSAs are strongly connected to Characteristic Features of tasks because they make it possible to evoke evidence about the KSA in accordance with an assessment argument. (We will see dynamic forms in the interactive design pattern allow designers to work with these connections.) For example, to let students exhibit their ability to generate a hypothesis, which corresponds to the Focal KSA “Hypothesis generation or evaluation about scientific phenomena that are subject only to observational testing and not to experimental testing,” observational data needs to be presented to students as a stimulus. This requirement is reflected by a corresponding Characteristic Feature, “Collection, presentation, and or representation of observational data.” In the interactive version of this design pattern, these two entries are linked together; that is, if a user highlight either one, the other will be highlighted (e.g., bolded), so that design implications of one aspect of tasks for the other aspect are called to the designer’s attention.

Additional KSAs list categories of other knowledge, skills, and abilities associated with tasks about observational investigation. Whether or not these KSAs are demanded by a task, and to what degree, will be affected by the task designer’s choice of a Variable Task Feature. For example, content knowledge as an additional KSA is necessary for students to manifest their reasoning skills of observational investigation in a concrete context. On one hand, in some assessment applications the designer will be interested in seeing whether a student can use observational-investigation reasoning with a model that is also part of the construct to be assessed. On the other hand, a designer may wish to use science content that is known to be familiar to examinees, so that the focus of evidence will be on reasoning in the observation-investigation context. Task writers can decide what content knowledge and how much to demand in the task by adjusting the associated Variable Feature.

The attributes of Work Products and Potential Observations in this design pattern concern a set of suggested ways one can capture information from students’ performances as evidence about

[14] Minnesota Comprehensive Assessments Series II (MCA-II): Test Specifications for Science. http://education.state.mn.us/mdeprod/groups/Assessment/documents/Report/006366.pdf

reasoning skills in observational investigation. In the online version of the design pattern, there are links between particular Work Products and Potential Observations because certain Work Products support certain types of observations. Taking the Focal KSA “the ability to generate a hypothesis” as an example, a corresponding Work Product might be an explanation or conjecture about the findings that would be observed, or modified/alternative explanation to the original problematic one. Then the assessment designer needs to identify the plausibility/correctness of this explanation associated with given observed phenomena. Associations of this kind among attributes would be highlighted in the online version.

6. Technology-Enhanced Features of the PADI Online Design Patterns.

The PADI online assessment design system is a web-based interactive system that includes a library of design patterns. To make design patterns available in a form that better suits task designers' needs, a set of technology-enhanced features have been developed over the past decade. These technology-based design patterns are illustrated in this section. The focus now shifts from the content of design pattern (i.e., attributes) to the various ways of presenting design patterns using the affordances of information technology. Here we will present an interactive version of the design pattern, which provides more details and examples beyond those provided in the non-technology enhanced version of design pattern in Table 1. The technology-enhanced design patterns highlight associations among attributes using both vertical and horizontal views of the design patterns. In addition, design patterns can be linked to other related design patterns to provide a hierarchical picture of a family of related design patterns. In addition, the PADI design pattern library is supported with a glossary of terms to help users with the language of ECD and the design pattern attributes. All these technology-enhanced features of the online design patterns make it easier for an assessment designer to exploit the connections among attributes and help support the development of assessment tasks.

6.1Vertical view with interactive features

A vertical view of the design pattern is the main presentation format used in the PADI design pattern library. Figures 6 through Figure 8 display different parts of the vertical view of the Observational Investigation design pattern. The left column lists the design pattern attributes while the specific content about observational investigations is filled in the right column. Then we can see the presentation of attributes with some interactive features, which are described below.

6.1.1 Active Pedagogical Content Knowledge and Exemplar Assessment Tasks

The first interactive feature available in the vertical view of design patterns is that more content knowledge and exemplar tasks can be obtained by clicking on the hyperlink associated with the attribute entry. In the electronic form as displayed in Figure 6, "details" indicates a hyperlink that will provide additional information about the Focal KSA of conducting model testing. If the user touches/clicks this link, more information about testing models will pop up—the role of model testing in observational investigation as well as its associated key procedures will be provided. Some of the detail links contain pointers to research literature on the web as well. This information can pop up in a rectangular text box as shown in Figure 6 or in a separate window as the user chooses. If users follow the links provided in the rectangular pop-up box, they can get example items to assess students' ability to evaluate supporting evidence for a hypothesis, which is one aspect of model testing. Thus using hyperlinks can unfold information successively. What is presented as an initial screen in the vertical view can be compact and not overwhelming for users, while the follow-up links provide users with further details only as they want it.

Insert Figure 6 about here

6.1.2 Relates Relevant Standards and Benchmarks

Figure 7 presents a part of the National Educational Science Standards and Minnesota State Science benchmarks associated with the Observational Investigation design pattern. As Mislevy et al. (2009) emphasized, the KSAs in this design pattern encompass the national standards and Minnesota State benchmarks at a more overarching level. Thus, the Focal KSAs can be more closely associated with the unifying themes and inquiry skills demanded by current reform efforts in science education (National Science Education Standard; NRC, 1996).

Insert Figure 7 about here

6.1.3 Links among Associated Assessment Argument Components

As introduced in previous section, the associations among the key design pattern attributes could be highlighted as bolded, just as Figure 8 illustrates. In fact, elements of all the key attributes introduced in Section 5 of this report can be associated, and serve as the integral components of the assessment argument about reasoning skills in observational investigation. Figure 8 provides a more complete version of this design pattern by including further information about these attributes and the associations among them. In Figure 8, the highlighted Focal KSA is the ability to generate a hypothesis with given observations (here labeled FK3), which is also the target of inference. The associated Characteristic Feature (labeled CF2) indicates that observational data must be presented to evoke the evidence about this Focal KSA. Usually this aspect of reasoning skills in observational investigation is assessed in a particular context with particular scientific content and models, which may or may not be included in the construct of interest. This depends on the design choices made by the assessment task designers, according to what the task is intended to assess. The association of FK3 with PO4 points out that the task designer will evaluate the plausibility of students' explanations for observed phenomena to provide evidence for FK3. The highlighted information in Figure 8 can be used by the task designers to think about how to produce tasks that address FK3 from the perspective of the task, students' potential performances, and salient features associated with this Focal KSA.

Insert Figure 8 about here

6.1.4 Links among Design Patterns

Figure 9 presents some additional attributes for the Model Formation design pattern in the model-based reasoning suite of design patterns (Mislevy, Riconscente, & Rutstein, 2009). The three attributes enclosed within rectangles indicate the design pattern's relationship to other design patterns. "I am kind of " in the first rectangle indicates that the key inquiry skills this design pattern aims to embody is one kind of scientific reasoning, a bigger theme that also include other kinds of reasoning skills such as experimentation. The second rectangle of "there are parts of me" shows one possible elemental design pattern for conducting investigations, which is part of reasoning skills that model formation includes. The third rectangle indicates "I am a part of " which shows that the inquiry skills that this design pattern addresses are an integral part of the model-based inquiry skills that a suite of design pattern aims to address. These links indicate that a hierarchical relationship exists among the design patterns and that task designers can use more of them conjunctively, for example in a multiple stage investigation.

Insert Figure 9 about here

6.2 Horizontal Views

The vertical view of design pattern is the default presentation format in the PADI design pattern library. However, this format with all its details can provide too much information at once as task designers proceed through particular steps of designing items. For example, Figure 8 presents only the key attributes, but there is still a large amount of information presented. Task users might need to scroll up and down several times to get the information about elemental attributes and might lose track of key associations among attributes.

Nichols (2008) and Nichols and Fulkerson (2010) examined item writers' work processes and found they engaged in three phases of problem solving (initial representation, exploration, and solution phases) when using storyboard and benchmarks as design stimuli. Under the user-centered principle, they suggested design pattern could be structured to support item writers' work in these three stages. The use of a horizontal view is one way to improve task writers' work efficiency when using design pattern. Two particularly important groupings of attributes are illustrated within horizontal views (Figures 10 and 11.)

Insert Figure 10 about here

Insert Figure 11 about here

Using the same example in Section 5, we can see this group of attributes in the horizontal view can present the associations among attributes in a more compact way. Usually we suggest users pull down the first "Anchor" box to as they decide which Focal KSAs to target, but they could also choose other attributes rather than Focal KSAs to suit their needs. Based on this selection, users can choose some associated attributes to appear in the other two pull-down boxes as both Figures 10 and 11 exemplify. Potential Work Products and Potential Observations are closely related since both concern students' performances. Figure 10 lists possible three kinds of work products, identify or generate a hypothesis, generate or select an explanation, and fill-in a representational form to support a hypothesis, all of which are produced by students in response to the FK3. These three attributes serve as "stepping stones" that a task designer can look across when deciding what kinds of observations are needed to provide evidence of a student's performance on a particular Focal KSA and what kinds of Work Products support the collection of evidence that is needed.

Figure 11 shows another way of grouping attributes for creating tasks. This alignment indicates possible Additional KSAs that could be tested jointly with Focal KSAs in tasks as well as possible Variable Features that the task designer can fine tune to suit their purpose. For example, if they want to increase the difficulty of items, they could provide observations with rich content demanding profound understanding of the context knowledge (i.e., VF1); they could let students analyze more kinds of observational data simultaneously (i.e., VF8); or they could present an incomplete model as observations to students and thus more inferential work is required from the students. When building assessments for use with students with disabilities, it is possible to use the horizontal view presented in Figure 11 another way. If Additional KSAs represented categories of perceptive, expressive and cognitive skills that students needed in order to perform successfully on the assessment tasks, but they are not the target of the assessment, then the task designer could use this horizontal view to identify Variable Features that could be designed into the tasks in order to support the Additional KSAs and mitigate their influence on the students' performance on the Focal KSAs.

Here we just introduced two groupings of design pattern attributes that our work with actual task designers has proved useful. Other different groupings can be used to provide insights in the task development process, as the user chooses.

7. Discussion

Design patterns are knowledge representations that structure an assessment as an argument based on evidence. This tool helps task designers organize their knowledge and thinking on the construct to be assessed. Due to this ideal property, this design tool focuses the task designer on the relationship among the knowledge, skills and abilities to be assessed, the evidence needed to evaluate performance and the features of tasks that will elicit the evidence. Thus, design patterns are an epistemic form, similar to those catalogued and described by Collins and Ferguson (1993) and further illustrate the value of such tools in addressing complex design tasks. Collins and Ferguson chose the term "epistemic form" to underscore how a representation that builds around important principles can be a powerful cognitive tool, to help people organize work, coordinate their activities, and even construct new knowledge. Their examples range from simple lists (which are so familiar that we fail to appreciate their importance in our thinking!) to more complex forms such as blueprints and financial reports. We aim to make design patterns such a tool, where the domain is assessment design, the underlying ideas are the implicit but essential structure of assessment arguments, and the task at hand is writing assessment tasks. We have built around both the theoretical work in evidence-centered design and the practical lessons from the ways good task designers do their work.

Collins and Ferguson put equal weight on the "epistemic games" that one must learn to play, so to speak, to take advantage of the affordances of epistemic forms. Using "games" in the sense of Wittgenstein (1953) would be systematic, purposeful, ways of interacting with some conceptual tool, which one must learn. In the case of assessment design, the games one must learn to play with design patterns concern how to use the support they provide for relating aspects of task features and scoring with validity argumentation into the larger design process—which includes deep knowledge of the content area, the students to be assessed, and the constraints and the resources that characterize the assessment project at hand. To this end, the interactive version of design patterns has grown from our research in the projects discussed above, along with usability studies and talk-alouds with task designers, and cognitive research on task design and on design under constraints more generally. As shown in Section 6, technology has helped us enhance the capabilities and user-friendliness of design patterns, to make them more accessible and shareable across the user community.

Design patterns are particularly useful in guiding the development of innovative assessments, including those used in the MCA-scenario based tasks, performance assessments, and games and simulations. The design pattern has attributes that can be used to guard the validity of these complex assessments. The design pattern is seen to be a construct-oriented support tool, rather than simply just an organizational or procedural support.

Ongoing educational reforms in the United States call for the design of innovative assessments that can validly measure complex content and inquiry skills. This demand challenges expert assessment designers, and is even more challenging for novice designers. Design patterns are a support tool, derived under the framework of ECD to help meet the new needs in assessment. In this way, we hope to make explicit and make available more widely some of the tacit knowledge that characterizes the work of the best task developers.

REFERENCE LIST:

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

Almond, R.G., Steinberg, L.S., & Mislevy, R.J. (2002). A four-process architecture for assessment delivery, with connections to assessment design. *Journal of Technology, Learning, and Assessment*, *1*(5). Online journal; article at http://www.bc.edu/research/intasc/jtla/journal/v1n5.shtml

Baxter, G., & Mislevy, R. J. (2005). *The case for an integrated design framework for assessing science inquiry* (PADI Technical Report 5). Menlo Park, CA: SRI International.

Bachman, L. F. (2005). Building and supporting a case for assessment use. *Language Assessment Quarterly, 2*, 1-34.

Chung, G. K., Delacruz, W. K., Dionne, G. B., & Bewley, W. L. (2003, December). Linking assessment and instruction using ontologies. Proceedings of the I/ITSEC, Orlando, FL.

Colker, A. M., Liu, M., Mislevy, R., Haertel, G., Fried, R., & Zalles, D. (2010). *A Design Pattern for Experimental Investigation (Large-Scale Assessment Technical Report 8).* Menlo Park, CA: SRI International.

Collins, A., & Ferguson, W. (1993). Epistemic forms and epistemic games: Structures and strategies to guide inquiry. *Educational Psychologist, 28*, 25-42.

Cronbach, L. J., & Meehl, P.E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, 52, 281-302

Dziuban, C. D., & Vickery, K. V.(1973). Criterion-Referenced Measurements: Some Recent Developments. *Educational Leadership, 30*, 483 - 486.

Embretson, S. E. & Reise, S. (2000). Item response theory for psychologists. Mahwah, NJ: Erlbaum Publishers.

Jones, D., Stewart, S., & Power, L. (1999). Patterns: Using Proven Experience to Develop Online Learning. Paper presented at the Proceedings of ASCILITE'99, Brisbane.

Frizell, S. S., & Hubscher, R. (2002). *Aligning Theory and Web-based Instructional Design Practice with Design Patterns*. Proceedings of E-Learning 2002: World Conference on E-Learning in Corporate, Government, HealthCare, & Higher Education.

Gamma, E., Helm, R., Johnson, R., & Vlissides, J.(1995). *Design Patterns: Elements of Reusable Objective-Oriented Software*. Addison-Wesley, Reading, MA.

Garden R.A., & Orpwood, G. (1996). Development of the timss achievement tests. In M. O. Martin & D. L. Kelly (Eds.), *Third international mathematics and science study (timss) technical report, volume i: Design and development.* Chestnut Hill: MA:Boston College.

Greeno, J.G. (1976). Cognitive objectives of instruction: Theory of knowledge for solving problems and answering questions. In D. Klahr (Ed.), *Cognition and instruction* (pp. 123-159). Hillsdale, NJ: Erlbaum

Greeno, J.G., Collins, A.M., & Resnick, L.B. (1997). Cognition and learning. In D. Berliner & R. Calfee (Eds.), *Handbook of educational psychology* (pp. 15-47). New York: Simon & Schuster Macmillan.

Gitomer, D. H. & Bennett, R. E. (2002). Unmasking constructs through new technology, measurement theory, and cognitive science. Research Memorandum, February 2002, RM-02-01, Educational Testing Service, Princeton, NJ. [verified 2 Jul 2007] http://www.nap.edu/openbook/0309083206/html/1.html

Haertel, G., Haydel DeBarger, A., Cheng, B., Blackorby, J., Javitz, H., Ructtinger, L., Snow, E., Mislevy, R. J., Zhang, T., Murray, E., Gravel, J., Rose, D., Mitman Colker, A., & Hansen, E. G. (2010). Using Evidence-Centered Design and Univeral Design for Learning to Design Science Assessment Tasks for Students with Disabilities *(Assessment for Students with Disabilities Technical Report 1).* Menlo Park, CA: SRI International.

Haladyna, T.M. & Shindoll, R. T. (1989). Shells: A method for writing effective multiple-choice test items. *Evaluation and the Health Professions, 12*, 97-104.

Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory*. Newbury Park, CA: Sage.

Hansen, E.G., Mislevy, R.J., Steinberg, L.S., Lee, M.J., & Forer, D.C. (2005). Accessibility of tests within a validity framework. *System: An International Journal of Educational Technology and Applied Linguistics, 33,* 107-133.

Hively, W., Patterson, H. L., & Page, S. H. (1968). A “universe-defined” system of arithmetic achievement tests. *Journal of Educational Measurement, 5,* 275-290.

Kane, M. (1992). An argument-based approach to validation. *Psychological Bulletin*, 112, 527-535.

Kane, M. (2006). Validation. In R. J. Brennan (Ed.), *Educational measurement* (4th ed.) (pp. 18-64). Westport, CT: Praeger.

Krathwohl, D. R., & Payne, D. A. (1971). Defining and Assessing Educational Objectives. In R. L. Thorndike (Ed.), *Educational Measurement*. (Second Edition.). Washington, D.C.: American Council on Education.

Meisner, R.; Luecht, R. M.; & Reckase, M.R. (1993). *Statistical characteristics of items generated by computer algorithms, ACT Research Report Series, RR-93-9*. Iowa City, IA: ACT, Inc.

Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational Measurement*. (pp. 13-103). Macmillan, New York.

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

Mislevy, R. J. (1994). Evidence and inference in educational assessment. *Psychometrika*, *59*, 439–483.

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

Mislevy, R., & Riconscente, M. (2005). *Evidence-centered assessment design: Layers, structures, and terminology (PADI Technical Report 9).* Menlo Park, CA: SRI International. http://padi.sri.com/downloads/TR9_ECD.pdf

Mislevy, R.J. (2006). Cognitive psychology and educational assessment. In R.L. Brennan (Ed.), *Educational Measurement* (Fourth Edition) (pp. 257-305). Phoenix, AZ: Greenwood.

Mislevy, R.J., Behrens, J.T., Bennett, R.E., Demark, S.F., Frezzo, D.C., Levy, R., Robinson, D.H., Rutstein, D.W., Shute, V.J., Stanley, K., & Winters, F.I. (2010). On the roles of external knowledge representations in assessment design. *Journal of Technology, Learning, and Assessment*, *8*(2). http://escholarship.bc.edu/jtla/vol8/2

Mislevy, R., Liu, M., Cho, Y., Fulkerson, D., Nichols, P., Zalles, D., Fried, R., Haertel, G., Cheng, B., DeBarger, A., Villalba, S., Colker, A., Haynie, K., & Hamel, L., (2009). *A Design Pattern for Observational Investigation Assessment Tasks (Large-Scale Assessment Technical Report 2).* Menlo Park, CA: SRI International. http://ecd.sri.com/downloads/ECD_TR2_DesignPattern_for_ObservationalInvestFL.pdf

Mislevy, R.J., Riconscente, M.M., & Rutstein, D.W. (2009). Design patterns for assessing model-based reasoning *(PADI-Large Systems Technical Report 6).* Menlo Park, CA: SRI International.

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

Mor, Y., & Winters, N. (2007) Design approaches in technology enhanced learning, *Interactive Learning Environments* 15(1), 61-75.

National Research Council (1996). *National Science Education Standards.* National Committee on Science Education Standards and Assessment. Washington, DC: National Academy Press.

Niemi, D. (2005, April). Assessment objects for domain-independent and domain specific assessment. In F. C. Sloane & J. W. Pellegrino (co-Chairs), *Moving technology up-design requirements for valid, effective classroom and large-scale assessment*. Paper presented at the annual meeting of the American Educational Research Association, Montreal.

Newble, D.I., Hoare, J., & Elmslie, R. G. (1981). *The validity and reliability of a new examination of the clinical competence of medical students.* Medical Education.

Newell, A., & Simon, H. A.(1972). *Human problem solving.* Englewood Cliffs, N.J.: Prentice Hall.

Nichols, P., Mittelholtz, D., Adams, J., & van Deusen, R. (2008, March). *User-Centered Assessment Design.* Paper presented at the annual meeting of the American Educational Research Association, New York.

Nichols, P., & Fulkerson, D. (2010). *Informing Design Patterns Using Research on Item Writing Expertise (Large-Scale Assessment Technical Report 9).* Menlo Park, CA: SRI International.

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

Polti, G. (1916). *The Thirty-Six Dramatic Situations.* Boston: The Writer, Inc.

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

Toulmin, S. E. (1958). *The Use of Argument*. Cambridge University Press, Cambridge, UK.

Vendlinski, T., Niemi, D., and Wang, J. (2005). Learning assessment by designing assessments: an on-line formative assessment design tool. In *Proceedings of the Society for Information Technology and Teacher Education International Conference 2005*, edited by C. Crawford, R. Carlsen, I. Gibson, K. McFerrin, J. Price, and R. Weber, pp. 228–240. Norfolk, VA: Association for the Advancement of Computing in Education.

Wigmore, J. H. (1937). *The science of judicial proof* (3rd ed.). Boston: Little, Brown, & Co.

Wittgenstein, L. (1953). *Philosophical Investigations*. Translated by G.E.M. Anscombe. New York: MacMillan.

APPENDIX

TABLES and FIGURES:

Table 1**:** *Condensed, Static Version of Observational Investigations Design Pattern*

Overview	This design pattern supports the writing of tasks that address scientific reasoning and process skills in the context of observational (non-experimental) investigations. This design pattern can be used in conjunction with any science content strand.
Use	This design pattern informs the writing of tasks that evoke evidence about reasoning in the context observational investigations.
Focal Knowledge, Skills and Abilities (KSAs)	▪Understanding why some scientific ideas need to be investigated through observational methods ▪Ability to analyze situations in which observational methods are more appropriate than experimental methods ▪Ability to distinguish between observational and experimental methodology ▪Hypothesis generation or evaluation about scientific phenomena that are subject only to observational testing and not to experimental testing ▪Hypothesis testing through observational methods ▪Ability to formulate conclusions, create models, and appropriately generalize results from observational, non-experimental research
Supported Benchmarks in MCA-II and MCA-III (latest version)	This design pattern can be used to support writing tasks for the following benchmarks associated with the Minnesota Comprehensive Assessment II and MCA-III.. ▪MCA II: 6.I.A.2, 7.I.A.2, 6,I.B.1, 7.I.B.1, 6.I.B.2, 6.I.B.4, 8.I.B.1, 8.I.B.2; High School: 9-12.I.A.3, 9-12.I.B.1, 9-12.I.B.6 ▪MCA III: 7.1.1.1.1, 7.1.1.1.2, 7.1.1.2.1, 7.1.1.2.3, 7.1.1.2.4, 7.1.3.4.1, 8.1.1.1.1, 8.1.1.2.1, 8.1.3.4.1.
Additional KSAs	▪Content knowledge (may be construct relevant) ▪Prerequisite knowledge from earlier grades ▪Ability to collect data ▪Ability to analyze data ▪Knowledge of representational forms (e.g., graphs, maps)
Characteristic Features	Tasks written using this design pattern will exhibit one or more of the following features: ▪Focus on HNS (Strand I) benchmarks relating to observational investigations at the appropriate grade level ▪Collection, presentation, and/or representation of observational data ▪Analysis and explanation of data ▪Conclusion generation, given observational data ▪Hypothesis generation, explanation, and/or modeling ▪Model development, analysis, and testing

Variable Features	The following features can be varied: ▪ Content (strand) context ▪ Qualitative vs. quantitative investigations ▪ Number of variables and complexity of their interrelationships ▪ Simple or complex investigations ▪ Data representation (e.g., patterns in geographically distributed phenomena via geospatial visualizations; patterns in data; similarities in specialized representations appropriate to the scientific phenomenon)
Potential Work Products	▪ Generate or identify an explanation for observed findings ▪ Modify or criticize problematic explanations. ▪ Identify or generate different observational settings that would help confirm or disconfirm hypotheses ▪ Identify or suggest other data that confirm or disconfirm a hypothesis for which evidence has already been identified from a different data source ▪ Identify or suggest potentially disconfirming observations that are stronger in being disconfirming than confirming ▪ Identify or suggest a process that may be occurring over time or across locations to produce observations (connected with a content-area) ▪ Create or fill in representation form (such as a graph, chart, or map) to express a hypothesis about what would be expected to happen under that hypothesis ▪ Critiques of peers (hypothetical in a standard assessment, real in classroom work) on their evaluations, explanations, or confirmation/disconfirmation procedures.
Potential Observations	▪ Plausibility / correctness of explanation for observed findings ▪ Appropriateness of other potential observations for confirming or disconfirming hypothesis ▪ Accuracy in identifying the effects of an observed active phenomenon and how they may be a sign of a cause and effect relationship ▪ Strength of evidence of a suggested or identified situation where observation could help confirm or disconfirm a hypothesis ▪ Correctness or aptness of recognized patterns that ground a hypothesis ▪ Accuracy in critiques of others (hypothetical in a standard assessment, real in classroom work) on the accuracy of what they identify in any of the above potential observations
Narrative Structures	▪ Investigation ▪ Specific to general ▪ Parts to whole ▪ Topic with examples ▪ Change over time ▪ Cause and effect

Table 2. *Layers of Evidence-Centered Design for Educational Assessments*

Layer	Role	Key Entities	Selected Knowledge Representations
Domain Analysis	Gather substantive information about the domain of interest that has implications for assessment; how knowledge is constructed, acquired, used, communicated.	Domain concepts, terminology, tools, knowledge representations, analyses, situations of use, patterns of interaction.	Representational forms and symbol systems used in domain (e.g., algebraic notation, Punnett squares, maps, computer program interfaces, content standards, concept maps).
Domain Modeling	Express assessment argument in narrative form based on information from Domain Analysis.	Specifications of knowledge, skills, or other attributes to be assessed; features of situations that can evoke evidence; kinds of performances that convey evidence.	Design patterns; “big ideas”, Toulmin and Wigmore diagrams for assessment arguments; assessment blueprints, ontologies, generic rubrics.
Conceptual Assessment Framework	Express assessment argument in structures and specifications for tasks and tests, evaluation procedures, measurement models.	Student, evidence, and task models; student, observable, and task variables; rubrics; measurement models; test assembly specifications; task templates and task specifications.	Algebraic and graphical representations of measurement models; task templates and task specifications; item generation models; generic rubrics; algorithms for automated scoring.
Assessment Implementation	Implement assessment, including presentation-ready tasks and calibrated measurement models	Task materials (including all materials, tools, affordances); pilot test data to hone evaluation procedures and fit measurement models.	Coded algorithms for rendering tasks, interacting with examinees and evaluating work products; tasks as displayed; IMS/QTI representation of materials; ASCII files of item parameters.
Assessment Delivery	Coordinate interactions of students and tasks: task-and test-level scoring; reporting.	Tasks as presented; work products as created; scores as evaluated.	Renderings of materials; numerical and graphical summaries for individual and groups; specifications for results files.

Table 3: *Key Attributes of a Design Pattern*

Attribute	Definition	Assessment Argument Component
Focal KSA	The primary knowledge/skill/abilities targeted by this design pattern	Claim
Rationale	Nature of the KSA of interest and how it is manifest	Warrant
Additional KSA	Other knowledge/skills/abilities that may be required by tasks motivated by this design pattern.	Claim, if relevant; Alternative Explanation, if irrelevant
Potential Work Products	Things students say, do, or make that can provide evidence about the focal knowledge/skills/abilities.	Data concerning students' actions
Potential Observations	Features of work products that encapsulate evidence about focal KSA	Data concerning students' actions
Characteristic Features	Aspects of assessment situations likely to evoke the desired evidence.	Data concerning situation
Variable Features	Aspects of assessment situations that can be varied in order to control difficulty or target emphasis on various aspects of KSA.	Data concerning situation
Potential Rubrics	Ways of evaluating work products to produce values of observations.	Warrant

Table 4. How Design Pattern Attributes Inform the Specification of ECD Models

	ECD Models in the CAF			
Design Pattern Attribute	Student Model	Evidence Model--Measurement Component	Evidence Model--Evaluation Component	Task Model
Focal KSA(s)	√			
Additional KSA(s)				
Potential observations		√	√	
Potential work products				√
Potential rubrics			√	
Characteristic Features of tasks				√
Variable Features of tasks		√		√

Table 5. *PADI Online Design Pattern Library: Number and Types of Design Pattern*

Subject Areas	Project	Education Standards	Unifying Themes/ Inquiry	Big Ideas within Disciplines	Learning Progressions	Language Proficiency	Total
Science	*DSA*	0		3	0	0	3
	LS	2	11	0	2	0	15
	B	0	0	0	0	0	0
	OP	2	46	0	0	0	48
	PBL	0	0	1	0	0	1
	SE	13	0	0	0	0	13
	Total	**17**	**57**	**4**	**2**	**0**	**80**
Mathematics	*OP*	0	2	0	0	0	2
	PBL	0	0	3	0	0	3
	SWSCD	30	0	0	0	0	30
	Total	**30**	**2**	**3**	**0**	**0**	**35**
Economics	*DSA*	0	0	3	0	0	3
	Total	**0**	**0**	**3**	**0**	**0**	**3**
Language Arts	*SWSCD*	30	0	0	0	0	30
	SE	0	0	0	0	1	1
	Total	**30**	**0**	**0**	**0**	**1**	**31**
Management/ Business	*UP*	0	5	0	0	0	5
	B	0	2	0	0	0	2
	PBL	0	0	3	0	0	3
	Total	**0**	**7**	**3**	**0**	**0**	**10**
Second Language Learning	*OP*	**0**	**0**	**0**	**0**	3	3
	Total	**0**	**0**	**0**	**0**	**3**	**3**
Grand Total		**77**	**66**	**13**	**2**	**4**	**162**

Note: Legend of Project Names: SE = PADI Special Ed; LS = PADI Large Scale; OP = Original PADI; DSA = Domain Specific Assessment; SWSCD = Students with Significant Cognitive Disabilities; PBL = Problem Based Learning; B = Benesse; UP = Urban Planning.

Figure 1: *Multiple-choice task motivated by the Observational Investigation design pattern*

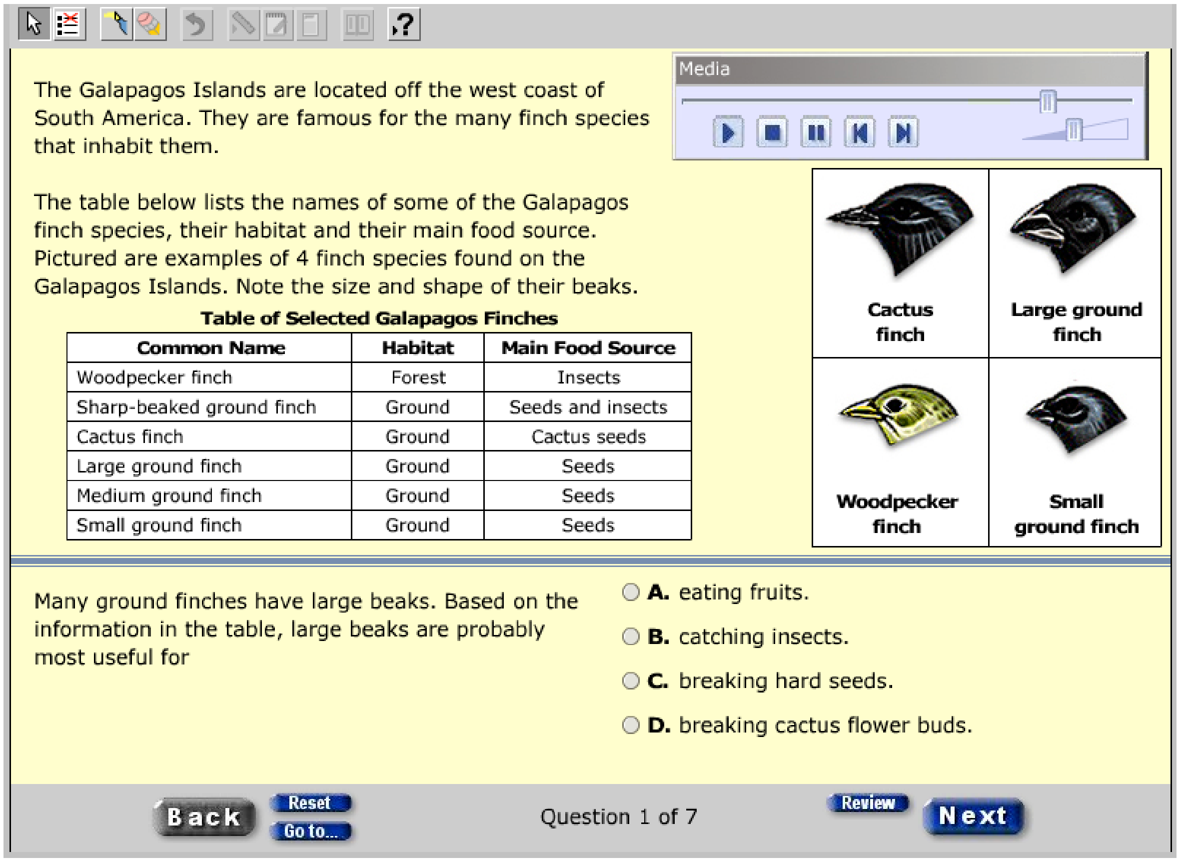


Figure 2: *Constructed response task motivated by the Observational Investigation design pattern*

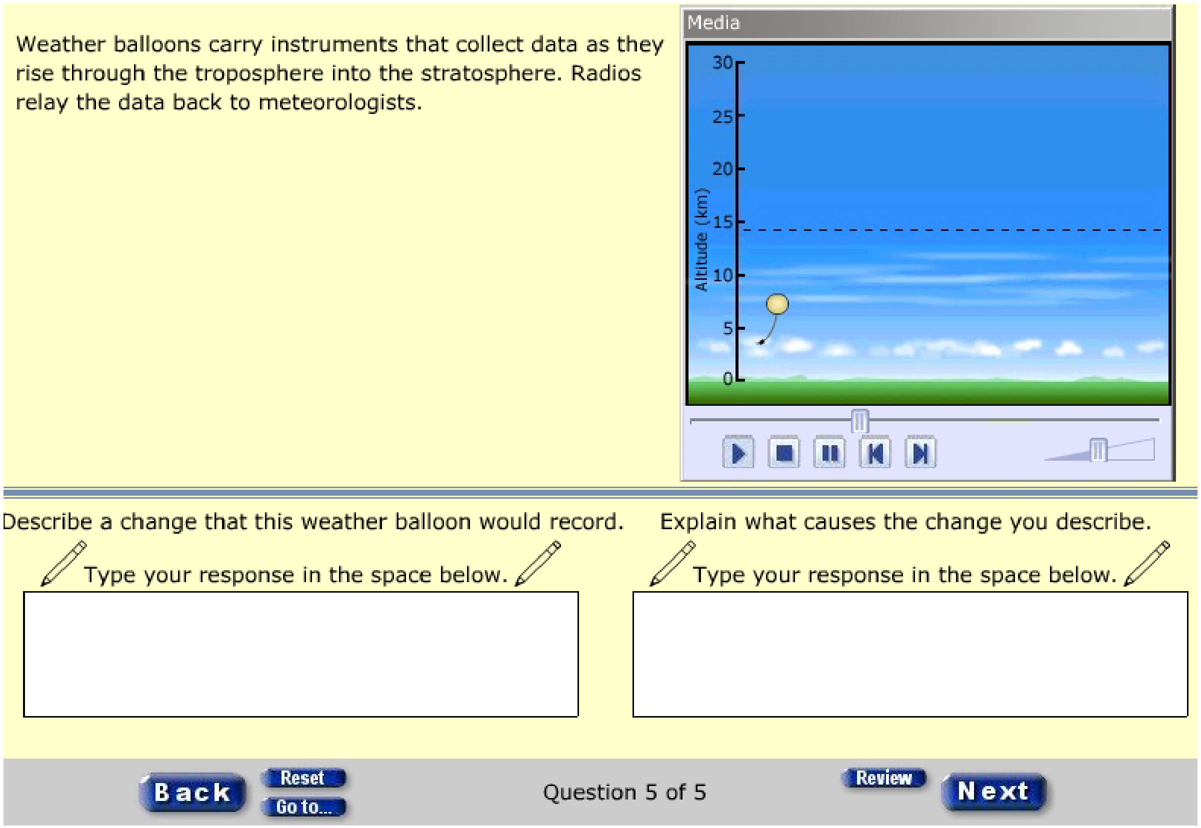


Figure 3. *Layers of Evidence-Centered Design for Educational Assessment*

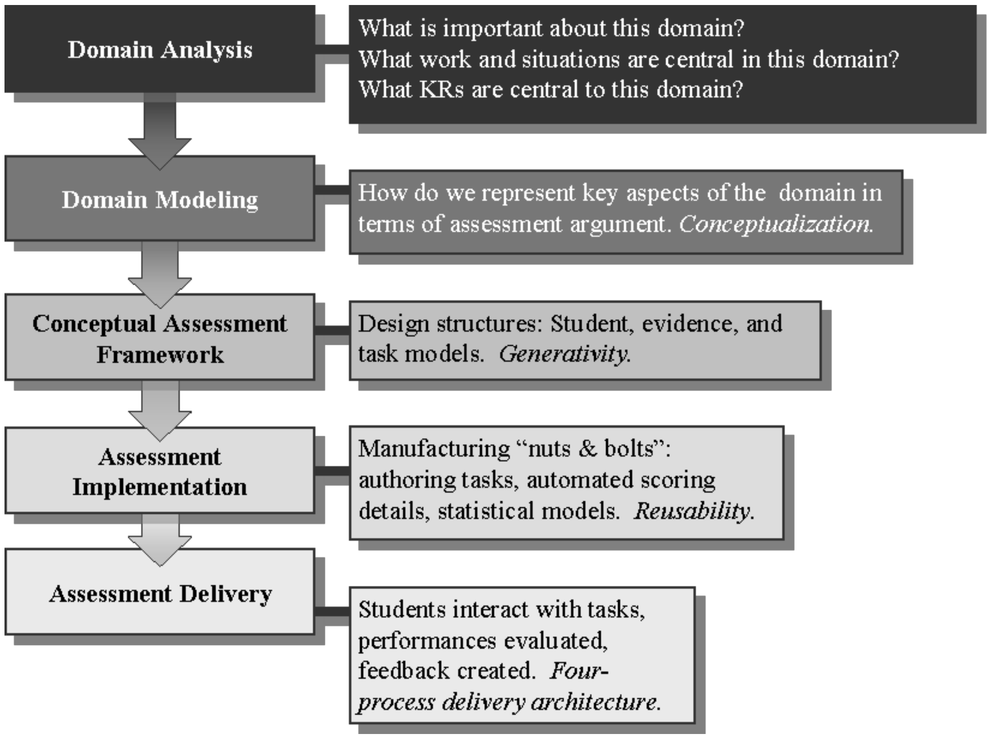


Figure 4. *A Toulmin Argument Diagram for Assessment Arguments*

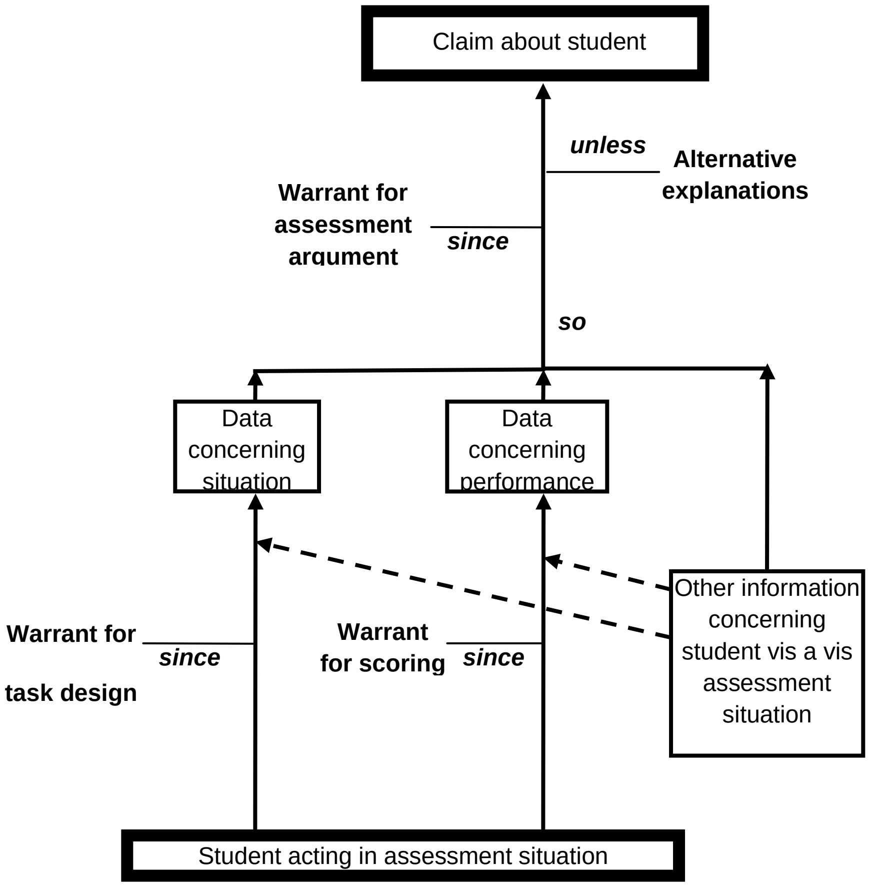


Figure 5. *Basic (Generic) Template Structure*

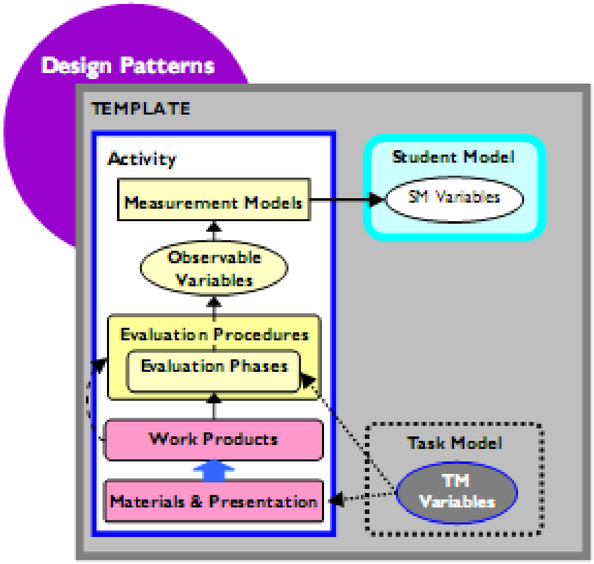


Figure 6: *Links for Details and Examples about Attributes in the Vertical View of Design Pattern for Observational Investigation*

Figure 7. *View of Selected National Educational Standards and Minnesota State Science Benchmarks Addressed by the Design Pattern of Observational Investigation.*

National educational standards	[Edit]	NSES 8ASI1.1. Identify questions that can be answered through scientific investigations. Students should develop the ability to refine and refocus broad and ill-defined questions. An important aspect of this ability consists of students' ability to clarify questions and inquiries and direct them toward objects and phenomena that can be described, explained, or predicted by scientific investigations. Students should develop the ability to identify their questions with scientific ideas, concepts, and quantitative relationships that guide investigation. NSES 8ASI1.2. Design and conduct a scientific investigation. Students should develop general abilities, such as systematic observation, making accurate measurements, and identifying and controlling variables. They should also develop the ability to clarify their ideas that are influencing and guiding the inquiry, and to understand how those ideas compare with current scientific knowledge. Students can learn to formulate questions, design investigations, execute investigations, interpret data, use evidence to generate explanations, propose alternative explanations, and critique explanations and procedures. NSES 8ASI1.3. Use appropriate tools and techniques to gather, analyze, and interpret data. The use of tools and techniques, including mathematics, will be guided by the question asked and the investigations students design. The use of computers for the collection, summary, and display of evidence is part of this standard. Students should be able to access, gather, store, retrieve, and organize data, using hardware and software designed for these purposes.
State benchmarks	[Edit]	MCA II: 6.I.A.2. The student will explain why scientists often repeat investigations to be sure of the results. MCA II: 6.I.B.1. The student will identify questions that can be answered through scientific investigation and those that cannot. MCA II: 6.I.B.2. The student will distinguish among observation, prediction and inference. MCA II: 6.I.B.4. The student will present and explain data and findings from controlled experiments using multiple representations including tables, graphs, physical models and demonstrations. MCA II: 7.I.A.2. The student will explain natural phenomena by using appropriate physical, conceptual and mathematical models. MCA II: 7.I.B.1. The student will formulate a testable hypothesis based on prior knowledge. MCA II: 8.I.B.1. The student will know that scientific investigations involve the common elements of systematic observations, the careful collection of relevant evidence, logical reasoning and innovation in developing hypotheses and explanations

Figure 8. *Highlighted Associations of Attributes in Observational Investigation Design Pattern*

Focal knowledge, skills, and abilities	[Edit]	Fk1. Ability to analyze why observational investigation methods are more appropriate than experimental methods for some phenomena/situations details
		Fk2. Ability to distinguish between observational and experimental methodology details
		Fk3. **Ability to generate or evaluate predictions or hypotheses about scientific phenomena that are appropriate for observational investigation** details
		Fk4. Ability to formulate conclusions, create models, and appropriately generalize results from observational investigations details
		Fk5. Ability to test predictions or hypotheses using observational methods details
		Fk6. Ability to plan a systematic collection of observational data based on a predicted relationship
		Fk7. Ability to collect, analyze, and interpret observational data with appropriate tools
Additional knowledge, skills, and abilities	[Edit]	Ak1. **Content knowledge (may be construct relevant)** details
		Ak2. Prerequisite knowledge from earlier grades details
		Ak3. Data collection and analysis details
		Ak4. Representational forms (e.g., graphs, maps) details
Potential observations	[Edit]	Po1. Appropriateness/strength of observational evidence to help confirm or disconfirm a prediction or hypothesis details
		Po2. Accuracy in identifying the effects of an observed active phenomenon and how these effects are consistent with a posited cause and effect relationship details
		Po3. Correctness of recognized pattern in data to support a prediction or hypothesis details
		Po4. **Plausibility/correctness of explanation for observed findings** details
		Po5. Accuracy in critiquing the observational investigation methods, evidence, and conclusions of others details
		Po6. Plausibility and systematicity of the data collection plan
		Po7. Correctness of selected tools and procedures for data collection
		Po8. Systematicity and appropriateness of collected data
		Po9. Appropriateness of measurement precision
Potential work products	[Edit]	Pw1. **Identification or generation of a prediction or hypothesis that is appropriate to an observational investigation situation** details
		Pw2. Identification of observational settings where data could be collected to confirm or disconfirm a prediction or hypothesis details
		Pw3. Identification of additional source of data that could confirm or disconfirm a prediction or hypothesis supported by existing data details
		Pw4. Identification or generation of a replicable data collection process (e.g., repeated sampling over time or at several locations) details
		Pw5. Identification of potentially disconfirming observations
		Pw6. **Filling in of a representational form (e.g., a graph, chart, or map) to show the relationship among variables relevant to a prediction or hypothesis** details
		Pw7. **Generation or selection of an explanation for observed findings** details
		Pw8. Critique of flawed explanation based on observations
		Pw9. Peer critique (hypothetical in a standard assessment, real in classroom work) of the observational investigation methods, evidence, and conclusions details

Characteristic features [Edit]

Cf1. **Focus on Nature of Science (Strand I) benchmarks relating to observational investigations at the appropriate grade level** details

Cf2. **Presentation of a real-world situation with patterns suggesting the relationship between at least two variables that can be observed systematically (but are not amenable to experimental investigation).** details

Variable features [Edit]

Vf1. Content (strand) context details

Vf2. Qualitative or quantitative investigations details

Vf3. Number of variables and the complexity of their relationships details

Vf4. Simple or complex investigations details

Vf5. Type of data representation (e.g., patterns in geographically distributed phenomena via geospatial visualizations; patterns in data; similarities in specialized representations appropriate to the scientific phenomenon) details

Vf6. Sufficient or insufficient data about an already established relationship details

Vf7. Amount of scaffolding given to student to guide the presentation or representation of data collected details

Vf8. **Amount of observational data from which an analysis, explanation, or conclusion is to be drawn** details

Vf9. **Completeness of model given from which predictions or hypotheses can be generated** details

Figure 9. *Example of Links among Design Patterns*

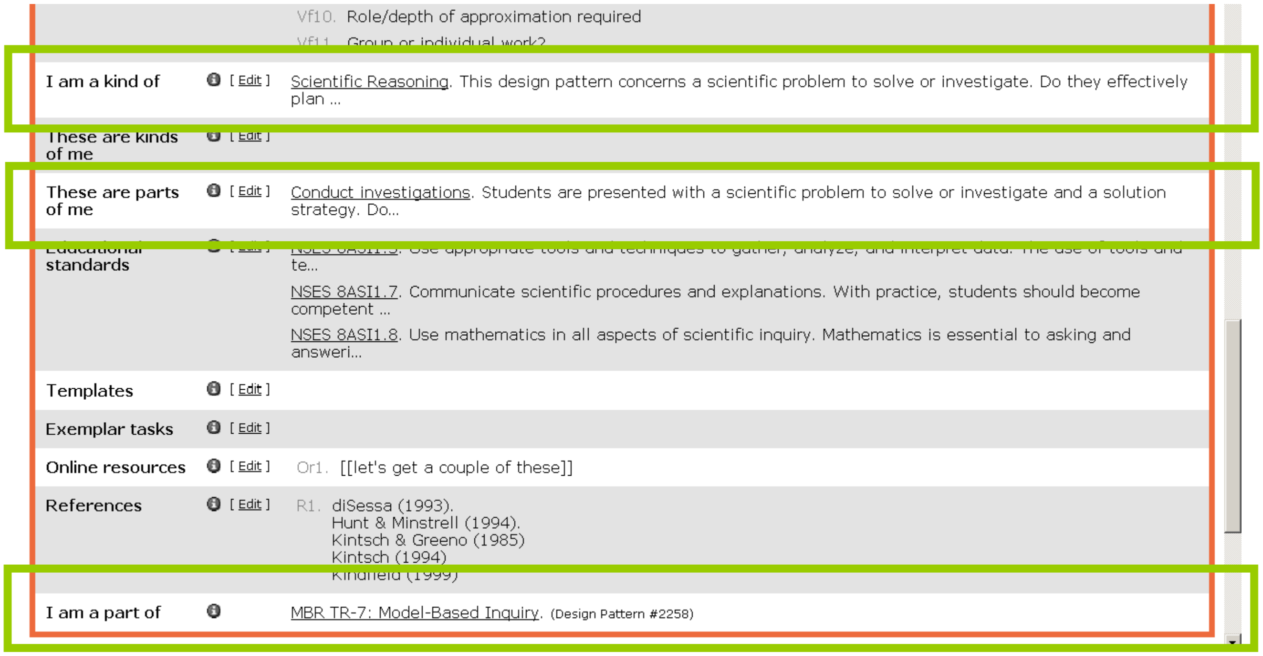


Figure 10. *Horizontal View of Focal KSAs, Potential Observations, and Work Products*

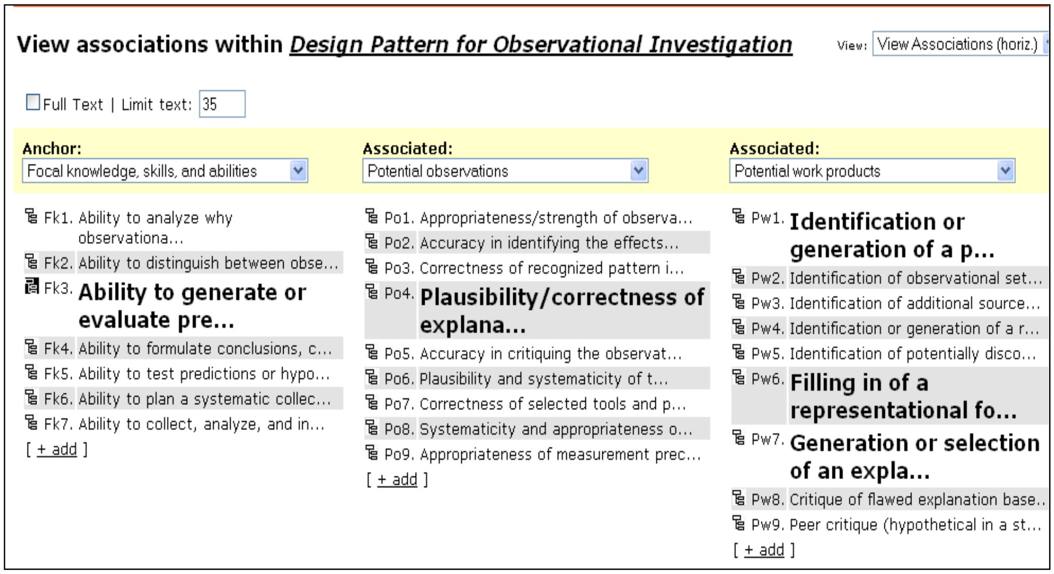
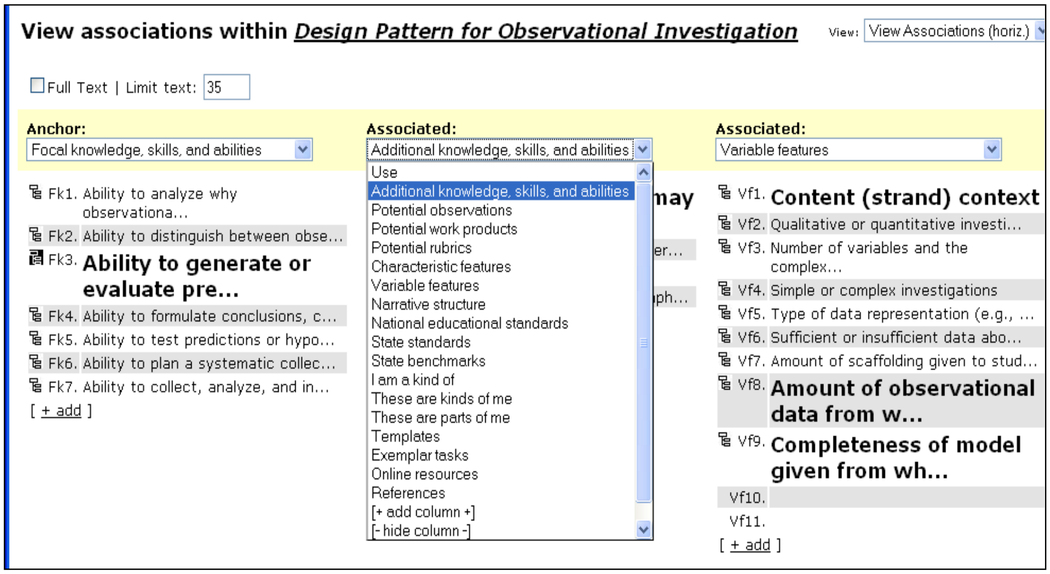


Figure 11. *Horizontal View of Focal KSAs, Additional KSAs, and Variable Features*



!

Sponsor
The National Science Foundation, Grant DRL - 0733172

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

Subgrantees
University of Maryland
Pearson