Chapter 12
Conclusion

Abstract Research on science learning increases our understanding of the capabilities we want to help students develop, and advances in technology expand the ways we can support and assess their learning. Familiar testing practices offer little guidance, however, for designing valid assessments of more ambitious proficiencies in more complex settings. These *design patterns* can support the development of tasks for assessing model-based reasoning in a variety of contexts, including standards-based assessment, classroom assessment, large-scale accountability testing, and simulation- and game-based assessment.

Model-based reasoning, and inquiry in general, are both increasingly important and difficult to assess (Means & Haertel, 2002). Assessing factual knowledge and isolated procedures is easier and more familiar—and not surprisingly, constitutes the bulk of current science assessment. The *design patterns* developed here can be used as starting points for building assessment tasks that engage more deeply with model-based reasoning. Task developers can determine which aspects of model-based reasoning to address and use the corresponding *design patterns* to make them aware of design choices and support their thinking about how to make them. The *design patterns* are organized around elements of an assessment argument structure as it has emerged from research on assessment design and validity theory. In this way, the *design patterns* leverage both research on model-based reasoning and practical experience in assessment design in this area, in a form that is specifically designed to support task developers.

12.1 Standards-Based Assessment

As part of the standards-based reform movement over the last two decades, states and national organizations have developed content standards outlining what students should know and be able to do in core subjects, including science (e.g., NRC, 1996, 2012). These efforts are an important step toward furthering professional consensus about the knowledge and skills that are important for students to learn at

R.J. Mislevy et al., *Assessing Model-Based Reasoning using Evidence-Centered Design*, SpringerBriefs in Statistics, DOI 10.1007/978-3-319-52246-3_12

various stages of their education. They are the basis of states' large-scale accountability tests, as was the case under the requirements of the 2001 No Child Left Behind (Public Law 107-110, 2002) legislation and is currently advocated by the Next Generation Science Standards (NGSS Core States, 2013).

But standards in their current form are not specifically geared toward guiding assessment design. A single standard for science inquiry will often encompass a broad domain of knowledge and skill, such as "develop descriptions, explanations, predictions, and models using evidence" (NRC, 1996, p. 145) or "communicate and defend a scientific argument" (p. 176). They stop short of laying out the interconnected elements that one must think through to develop a coherent assessment: the competencies that one is interested in assessing, what one would want to see students do as evidence of those competencies, and assessment situations that would elicit such evidence. Even NGSS performance expectations (NGSS Core States, 2013), which sketch illustrative tasks that could elicit processes, disciplinary knowledge, and overarching concepts, provide little guidance for task developers to operationalize the ideas at scale, with reliability and validity.[1]

Design patterns bridge knowledge about aspects of science inquiry that one would want to assess and the structures of a coherent assessment argument, in a format that guides task creation and assessment implementation. The focus at the *design pattern* level is on the substance of the assessment argument rather than on the technical details of operational elements and delivery systems. Thinking through the substance of assessment arguments for capabilities such as model-based reasoning and inquiry promotes the goals of efficiency and validity. It enables test developers to go beyond thinking about individual assessment tasks and to instead see instances of prototypical ways of getting evidence about the acquisition of various aspects of students' capabilities.

Design patterns bring insights from cognitive psychology, science education, and the philosophy of science together in a form that can support designing assessment tasks for both classroom and large-scale assessments. It is a particular advantage of *design patterns* to center on aspects of scientific capabilities, as opposed to task formats or assessment purposes. The essence of the capabilities and building assessment arguments around them is seen as common, with options for tailoring the details of stimulus situations and Work Products to suit the particulars of a given assessment application.

12.2 Classroom Assessment

Design patterns built around national or state science standards constitute a stationary point to connect both classroom and large-scale assessment with developments in science education and educational psychology. There is often a disjuncture

[1]In the terminology of *design patterns*, NGSS Performance Expectation highlight instantiations of Characteristic Features and Variable Features that would tap Focal KSAs.

between classroom assessment and large-scale assessment; *design patterns* help make it clear that it is the same capabilities being addressed in both, although the assessments reflect different design choices about such features as time, interactivity, and Work Products to accommodate the different purposes and constraints of large-scale and instructional tests.

Truly "knowing" models in science is more than echoing concepts and applying procedures in isolation; it is using models to do things in the real world: reasoning about situations through models; selecting, building and critiquing models; working with others and with tools in ways that revolve around the models. Students develop these capabilities by using them, first in supported activities that make explicit the concepts, the processes, and the metacognitive skills for using them. It is no coincidence that most of the examples we have used to illustrate science *assessment* are drawn from projects that focus on science *learning*. These *design patterns* for assessing model-based reasoning can help make the advances in science education more accessible to classroom teachers and curriculum developers as well as to researchers and assessment professionals.

12.3 Large-Scale Accountability Testing

The changing landscape of large-scale accountability assessments places extraordinary demands on state and local education agencies. No Child Left Behind legislation required large-scale testing at the level of the state, with attendant needs for efficient administration, scoring, linking of forms, and cost-effective development of assessment tasks at unprecedented scales. Tasks must address states' content standards. At the same time, educators want tasks that assess higher-level skills and are consistent with both instructional practice and learning science.

It is widely accepted that more complex, multi-part assessment tasks are better suited to measuring higher-level skills. But their cost and incompatibility with conventional test development and implementation practices stand in the way of large-scale use. Many states and their contractors have turned to computer-supported assessment task development and delivery to help them meet these challenges. For large-scale assessments, technology-based tasks such as simulations and investigations to address higher-level skills and support learning have proved difficult and costly to develop, especially when employing procedures that evolved from conventional multiple-choice item development practices.

Traditionally, items for large-scale assessments are developed by item writers who craft each item individually. Often as many as half of the items do not survive review. This low survival rate is tolerable because of the relatively low cost of developing individual multiple-choice items. It is not economical for developing the more complex tasks needed to address higher-level skills. Moreover, the thought and problem solving invested in developing any particular item is tacit in conventional item development procedures. The thinking invested, the design challenges met, and the solutions reached remain undocumented and inaccessible to help item writers develop additional items. This process is untenable in the long run for tasks that require an order of magnitude more time and resources than multiple-choice items.

Design patterns are part of the solution. A *design pattern* specifies a design space of interconnected elements to assemble into an assessment argument. This design space focuses on the science being assessed and guides the design of tasks with different forms and modes for different situations. *Design patterns*, in turn, ground *templates* for authoring more specific families of tasks.

In the context of large scale accountability assessments, *design patterns* thus fill a crucial gap between broad content standards and implemented assessments tasks, in a way that is more generative than test specifications and which addresses alignment through construction rather than retrospective classification. The time and analysis invested in creating *design patterns* eliminates duplicative efforts of re-addressing the same issues task by task, program by program. *Design patterns* can be developed collaboratively and shared across testing programs. Each program can construct tasks which, by virtue of the pattern, address key targets in valid ways, but make design choices that suit their particular constraints and purposes. Thus, *design patterns* add value not just for local development but for accumulating experience and debating standards in the state, national, and international arenas.

12.4 Simulation- and Game-Based Assessment

The ability to create computer-based simulation environments has opened the door to assessing model-based reasoning in complex, interactive environments practically anytime, anywhere. Simulations have a great advantage of making visible and more amenable to the cognitive aspects of modeling phenomena that might be too small, too big, too costly, too distant, or too dangerous. They can provide facsimiles of the tools and representations real scientists and engineers use. They can provide scaffolding, supporting material, just in time information, and provide feedback to students as they work as well as informing teachers or more distant users. And perhaps most importantly, they allow for interactions with situations—a hallmark of model-based reasoning in action.

Furthermore, log files of actions, time-stamps, and plans and products can all be captured automatically as rich, detailed Work Products. An exciting frontier of educational assessment is developing automated methods of evaluating log files, through which cognitively meaningful patterns and features of work are detected and characterized as Observations. There is great potential for many kinds of assessments: From moment-to-moment assessment and feedback in simulations for learning, to simulated investigations embedded in curricula, to technology-enhanced tasks in large-scale assessments.

But it is difficult to design tasks to be valid, comparable, and fair. Design patterns help with some of the thorniest problems—such as identifying alternative explanations for poor performance, and adapting Variable Task Features and Work products to a testing population, and addressing higher-level process skills such as

Model Revision within disciplinary contexts that match students' instructional backgrounds.

Design patterns thus offer support to designers of complex, computer-based tasks, to help make sure that they can produce not only good simulation around valued disciplinary content and processes, but valid assessment evidence as well. Resources on how to use ECD more generally, and *design patterns* in particular, in designing game-based and simulation-based assessments are now beginning to appear (e.g., Clarke-Midura & Dede, 2010; Gobert, Sao Pedro, Baker, Toto, & Montalvo, 2012; Mislevy et al., 2013; Riconscente, Mislevy, & Corrigan, 2015; Shute, Ventura, Bauer, & Zapata-Rivera, 2009).

12.5 Closing Comments

Model-based reasoning is central to science. Research from a sociocognitive perspective on the nature of model-based reasoning and how people become proficient at using it is beginning to revolutionize science education. Assessment is integral to learning, not just for guiding learning but for communicating to students and educators alike just what capabilities are important to develop, and how to know them when we see them. But the interactive, complex, and often technology-based tasks that are needed to assess model-based reasoning in its fullest forms are difficult to develop. The suite of *design patterns* to support the creation of tasks to assess model-based reasoning hold promise to help bring assessment into line with contemporary views of science learning and science assessment.