Chapter 11
Model-Based Inquiry

Abstract Model-based inquiry highlights the metacognitive aspects of managing and moving effectively through cycles of inquiry. The Focal KSAs in this *design pattern* are students' capabilities to manage their reasoning across in inquiry cycles. A key Variable Task Feature to consider is the degree of scaffolding to provide students as they move from one aspect of an inquiry to another. All the considerations, design choices, work products, and observations addressed in the preceding design patterns can be involved in a model-based inquiry task.

Distinguishing aspects of reasoning is useful in instruction and assessment, but it is their coordinated use that marks model-based reasoning in practice. We would like to help students learn to move among these aspects of reasoning, often without clear demarcation, to understand systems and act through models of them. The general design pattern for model based inquiry subsumes the *design patterns* for each of the aspects and calls attention to the coordination among them. More than any of the individual aspects, model-based inquiry highlights the importance of metacognition in moving effectively through cycles of inquiry.

This section draws on the model-based inquiry framework in White and Frederiksen (1998) and White, Shimoda, and Frederiksen (1999). More recently these ideas have been used in simulation environments to support students to carry out investigations, work through inquiry cycles, and build and test models (Clarke-Midura, Code, Zap, & Dede, 2012; Shute et al., 2010; Quellmalz et al., 2012). Providing students with considerable flexibility to choose what to do, when and where, in a simulated microworld, be it in a laboratory, out in the field, under the sea, or on an alien planet, makes it possible to assess their information management and interactive, iterative, reasoning. Capturing log files of actions as rich Work Products makes it possible to evaluate many Observables automatically. This design pattern provides support to designers wishing to assess this overarching aspect of model-based reasoning.

11.1   Rationale, Focal KSAs, and Characteristic Task Features

The philosophy of science, Giere (1994) argues, assumes that the language of science has a syntax, a semantics, and, finally, a pragmatics. He continues,

> While syntax is deemed important, semantics, which includes the basic notions of reference and truth, has received the most attention. Much of the debate regarding scientific realism, for example, has been conducted in terms of the reference of theoretical terms and the truth of theoretical hypotheses. Pragmatics has been largely a catchall for whatever is left over, but seldom systematically investigated. I now think that this way of conceiving representation in science has things upside down (p. 742).

Model-based reasoning is all about pragmatics. A philosophy of science is not sufficient for either understanding how scientists use models in practice or for how to help students learn to use them; a cognitive psychology of science is required as well. While the preceding sections on aspects of model-based reasoning illuminate important cognitive activities in model-based scientific inquiry, it is the heuristics, the strategies, the procedures, and the self-regulating tools that people need to use models effectively in real-world situations. It is this higher-level, coordinating, or executive level of cognition that the Model-Based Inquiry design pattern addresses.

The Focal KSAs in this design pattern are students' capabilities to manage their reasoning in inquiry cycles. The specific aspects of model-based reasoning discussed in the preceding sections are brought to bear, but is their use coordinated, efficient, coherent, and effective—or is movement through the investigation disjointed, unsystematic, inefficient, and aimless? Are students bringing to bear self-monitoring skills to understand whether model evaluation is needed, or whether a provisional model need to be revised or elaborated?

Any task developed for an overall assessment of model-based reasoning must contain more than one characteristic feature-set from the more specific *design patterns*. As with all of these *design patterns*, there must be a real-world problem being addressed. This problem must require the use of models and/or a modification of models in order to develop an explanation or prediction of some phenomena. The Model-Based Inquiry design pattern goes beyond the specific *design patterns* by addressing information and reasoning across the aspects.

Many of the examples mentioned in the previous sections can be expanded to include multiple aspects of model-based reasoning, and would therefore be instances from the overall design pattern. Stewart and Hafner's genetics curriculum can be thought of as one large assessment task, or it can be broken down into several distinct assessments. In this case, the assessment would start out where the students are applying the simple dominance model to a given situation (as seen in model use). The students then are presented a situation where it does not fit—say, three possible traits instead of two. The students must identify the inadequacies of the simple dominance model (model evaluation) and modify their model (model elaboration.) Students are given further information to lead to more complicated models. At points, they must revise or further elaborate their model in light of new

data. Work Products for this overarching task would include the explanations for the models and how they fit the situations, the overall outcomes of using the model to explain or predict behavior, and representations of the models. These Work Products can then be used to evaluate a student's model-based reasoning in the context of modes of inheritance.

Box Robotics-6. Model-Based Inquiry While the preceding discussions of the robotics task have focused on particular aspects of model-based reasoning, it will be clear by now that cycles of design, construction, testing, evaluating, and revising the rover are at the heart of the task. In each phase, reasoning through the underlying gear model and circuit model are required. But the task is structured so as to help the students become aware of the reasoning aspects and the rhythms of such investigations.

The Focal KSA is managing one's work through such cycles, here in the context of generously scaffolded disciplinary content through the MOOC. Additional KSAs are the disciplinary models, the specifics of the circuits, motors, gearboxes, and wheels through which the rovers are constructed, and the proficiency with the necessary tools, representations, and manipulations in a given phase of the investigation. In the simulation phase, these are the tools, affordances, and representations of the simulation environment. In the physical phase, they are proficiencies for the manual planning, assembly of, and operation of the components (plus proficiency of using the laser cutter, if a student is making custom wheels).

We have defined Model-Based Inquiry as an organizing framework for organizing the more specific aspects of model-based reasoning: awareness of those aspects, knowing how they are related, and how to move from one another effectively. The Characteristic Feature for a situation to provide evidence about these capabilities is that it must require two or more aspects of reasoning, and a student must move among them.

An important Variable Feature is the nature and amount of scaffolding that is provided for moving among aspects. The simulation phase in the robotics task provides a good deal of support, in two ways. First, the MOOC materials walk the student through the required background information on the models and the simulation tools and affordances, then structure the initial work in building the first simulation model (Model Formation) and running it (Model Use). Second, the Learning Companion (Fig. R5) provides more specific advice for examining the results of a hill-climbing attempt (Model Evaluation) and offers suggestions on what to try next (Model Revision). As seen in the flowchart, after three unsuccessful tries, it suggests getting help from the outside—an instructor or a friend perhaps—because the inquiry cycles are not converging within the amount of scaffolding the Learning Companion can offer. Note that providing its advice, the Learning Companion is carrying out assessment itself, using the log file Work Product, and counting attempts and comparing attempt results and students' revisions in response to them.

The physical phase offers much less explicit support. The rationale is that after successful completion of the analogous task in the simulation world, a student will have acquired some understanding of the build-run-evaluate-revise inquiry cycle. With less scaffolding, this may or may not happen. Potential Work Products that can provide evidence could include a video capture of the work, an after-the-fact explanation of the work, and a student's running record of models, results, interpretations, and revisions. Note that asking for students to keep a running record with these categories is itself is a mild form of scaffolding. Potential Observations of such Work Products could include the following:

- The degree to which a student organized their activity around such organized cycles.
- Instances of skipping necessary aspects of reasoning, or missing cues as to what actions should be taken next.
- "Churning" activity, with lots of building and running models but no real systematic learning from results and acting to improve on them.

11.2   Additional KSAs

As with the other *design patterns*, the Additional KSAs in the design pattern for assessing model-based inquiry include knowledge of the models, context, and scientific content involved. The mix of these Additional KSAs, if any, that is jointly a target of inference with inquiry itself must be determined in light of the purpose of the assessment and test population. Additional KSAs that are not part of the target of the assessment should be avoided or supported, or the assessor should ascertain that the students are sufficiently familiar with them so that they are not significant sources of difficulty.

11.3   Variable Task Features

Because inquiry tasks encompass the aspects of model-based reasoning addressed so far, all of the Variable Task Features for relevant aspects are open for consid-eration. This includes the identification and complexity of the model and which tools and representational forms are used. Some design choices can cut across aspects of the larger task (such as the models and content area that are involved) while others (such as scaffolding) can differ from one aspect to another (e.g., a checklist just for model evaluation). Time frame is an important Variable Feature

for investigations. Non-trivial investigations can easily take an hour or more, and learning tasks can extend to days or weeks.

Choices regarding the content area will be shaped by the intended purpose of the task. In the classroom or as part of a curriculum, the content is likely based on the models that are the focus of instruction, so the task can pose high demands for this knowledge. The students in the Baxter et al. Mystery Boxes study had just completed a unit in electrical circuits. In a high-stakes accountability test where both the models and the inquiry processes are addressed in the standards, demands for both may be imposed and the Additional KSAs regarding the model and scientific content can be construct-relevant. In a large-scale task that is meant to focus on the inquiry process and not be confounded with content, the models and content can be chosen to be familiar enough to students to minimize poor performance for these reasons. For example, models from middle school standards could be used in a secondary-level task in order to focus its evidentiary value on inquiry.

An important Variable Task Feature in designing inquiry tasks is the degree of scaffolding to provide students as they move from one aspect of an inquiry to another, for managing information, evaluating progress, and deciding what to do next. This self-monitoring is central to inquiry and one of the hardest aspects for students to learn (and for educators to assess). Research on scaffolding students' learning about inquiry holds insights for task designers. In inquiry assessment, more scaffolding is appropriate for earlier learners; it helps them engage meaningfully with the task and ensure that evidence will be obtained for aspects of the investigation. On the other hand, scaffolding the processes means less evidence is available about students' capability to manage their activity in the investigation.

White and Frederiksen (1998) describe a sequence of seven instructional tasks that constitute a middle-school course on mechanics, implemented in the ThinkerTools software. Scaffolding was progressively decreased as students became familiar with inquiry processes and expectations. Associated with each task context is a task document in which students carry out their work. They include a Project Journal, a Project Report, a Project Evaluation, and a System Modification Journal for recording their system modifications and the reasons for them. The documents are organized around a sequence of subtasks (or subgoals) for that task. For example, the Project Journal is organized around the inquiry cycle. The White et al. (1999) simulation environment SCI-WISE additionally provides interactive support in the form of personified "agents":

> In addition to Task Documents, each Task Context has a set of advisors associated with it, including a Head Advisor and a set of Task Specialists. There is a Head Advisor for each Task Context; namely, the Inquirer for doing research projects, the Presenter for creating presentations, the Assessor for evaluating projects, and the Modifier for making changes to the SCI-WISE system. The Head Advisor gives advice regarding how to manage its associated task, suggests possible goal structures for that task, and puts together an appropriate team of advisors. For example, our version of the Inquirer follows the Inquiry Cycle shown in [Fig. 2.2 of this paper]. It suggests pursuing a sequence of subgoals, and each such subgoal has a Task Specialist associated with it, namely, a Questioner, Hypothesizer, Investigator, Analyzer, Modeler, and Evaluator (p. 164).

In computer-based tasks, a developer could choose which agents to make available to examinees and what degree of support they could provide, in order to tailor scaffolding within and between aspects of model-based reasoning during an inquiry task. As always, however, providing tools that support inquiry-related KSAs introduces at the same time a demand for the Additional KSAs to use them effectively.

11.4   Potential Work Products and Potential Observations

Model-based inquiry tasks can be designed to produce Work Products that provide evidence about specific aspects of model-based reasoning within the investigation and/or evidence about managing reasoning across aspects over the course of the investigation. Since aspect-specific Work Products and Potential Observations were discussed previously, after a brief comment, this section focuses on Work Products and Potential Observations that address the encompassing inquiry process.

As mentioned above, all of the potential Work Products that contain evidence about aspects of model-based reasoning can be considered in a fuller inquiry task, and all of the Potential Observations that could be evaluated for these aspects can be considered. In a more detailed scoring scheme, the Observable Variables from the specific aspects can be evaluated and reported separately. This is useful for providing feedback to students in instructional settings: What did they do well in this task, where did they have trouble, and what experiences will help them improve?

Work Products that directly evidence the larger inquiry process must provide information beyond specific aspects of model-based reasoning. This means evidence about the way a student moves through the investigation. One class of Work Products provides some form of trace of the steps a student has taken, such as a video recording, a think-aloud protocol, or a log of actions captured in a computer–based task. The National Board of Medical Examiners' Primum® computer-based diagnostic tests, which are now required for licensure in the United States, capture each step in a solution in a "transaction list." Automated scoring algorithms (more about this below) extract information from the transaction list about both the final solution and selected aspects of the process. In general, less comprehensive Work Products include notebooks, explicit reports of inquiry phases, and written or oral explanations along the way of why certain actions were taken. Oral explanations can be prompted or unprompted. We will say more below about responses to "metacognitive" questions.

Final and intermediate products in an inquiry task are Work Products that can provide indirect evidence about inquiry procedures. A correct solution presumably is more likely to have occurred from effective model-based reasoning, although the efficiency of that reasoning is not available to evaluate from this Work Product alone. The qualities of a final solution to a problem, such as a model proposed for a situation after multiple iterations through the inquiry cycle, can be of interest in and of themselves. Only qualities of the final product may be addressed when the

Table 11.1 Quality of cognitive activity in mystery box solutions (Baxter, Elder, & Glaser, 1996)

Cognitive activity	Range of variation	
	Low	High
Explanation	Single statement of fact or descriptions of superficial features	Principled, coherent
Plan	Single hypothesis	Procedures and outcomes
Strategy	Trial and error	Efficient, informative, goal-oriented
Monitoring	Minimal and sporadic	Frequent and flexible

purpose of an assessment is licensure, for example. But when the purpose is learning, the evaluation of successive provisional models offers clues about the efficiency and appropriateness of successive cycles of model evaluation and revision.

The choice of Work Products to capture is linked to the choice of scaffolding to provide. The task documents White et al. (1999) provided students to record, evaluate, and explain their progress through an investigation not only serve as Work Products, but they support metacognition to manage their activity through the investigation.

What Observable Variables that hold evidence about model-based inquiry can be evaluated from Work Products? Baxter et al. used the Mystery Boxes tasks to study "expertise" in middle school students' inquiry capabilities in a domain known to be familiar to them. Table 11.1 summarizes dimensions of variation they found in a think-aloud protocol and solution-trace Work Products. They are the basis of generic Observable Variables that can be applied more generally in inquiry assessments, as tailored to the processes in the specific investigation.

Baxter et al. evaluated students' investigation procedures by painstakingly parsing "thick" Work Products such as explanations, solution paths, and conversations of thirty-one students. In more complex investigations at larger scales, the amount of rater time and expertise required to carry out such evaluations for these Observable Variables renders them impractical.

An alternative that is available when the investigations are carried out in a computer-based form is automated scoring of solution traces (Bejar, Mislevy, Rupp, & Zhang 2016). In Primum® tasks, low-level features of solutions are identified, combined into higher-level features through logical rules (such as whether efforts to stabilize an emergency patient were carried out first rather than later in the investigation), and evaluated using a regression function that compares them to the high-level features of experts' solutions (Margolis & Clauser, 2006).

More generally, Gobert, Sao Pedro, Baker, Toto, and Montalvo (2012) provide both an overview of approaches to automated scoring of performances on inquiry tasks in simulation environments and examples from their work with *Science Assistments*. The first category they discuss is knowledge engineering/cognitive task analysis approaches, in which rules are defined a priori to encapsulate specific behaviors or differing levels of systematic experimentation skill. The second

category is educational data mining/machine learning approaches, in which student inquiry behaviors are discovered from data. Their own examples blend ideas from the two. Leveraging Gobert's previous research on inquiry (including the model-based reasoning research cited above), they designed a simulated laboratory and affordances that minimized construct-irrelevant demands and maximized the evidentiary value of students' actions for how they were managing the inquiry process. For example, they provided a tool using drop-down menus for students to build hypotheses they would then test. The general structure was

> When the [*independent variable*] is [*increased/decreased*], the [*dependent variable*] [*increases/decreases/doesn't change*].

The Work Product produced by filling out the hypothesis is a filled in hypothesis statement–captured in a manner that the system knows exactly what the student has specified. Then, the trace of students' more open-ended actions through the environment of setting up tests, monitoring (or not monitoring) results, and setting up subsequent tests based on previous results (or seemingly not) could be detected by patterns discovered in data mining, based on a subset of actions tagged by expert reviewers. Further, an explanation tool similar to the hypothesis tools was used to capture students' interpretations of what they had done:

> When I changed the [*independent variable*] so that it [*increased/decreased*], the [*dependent variable*] *increased/decreased/didn't change*]. I am basing this on: Data from trial [*trial number from table*] compared to data from trial: [*trial number from table*] this statement [*does support/does not support/is not related to*] my hypothesis.

Together, these Work Products and consequent Observable Variables captured consistencies and inconsistencies, efficiencies, and appropriate stepping through inquiry actions, even though the investigation phases could be accomplished in any numbers of ways.

A class of paired Potential Work Products and Potential Observables that is particularly well-suited to instructional tasks is based on responses to metacognitive questions. These are the questions that students should be learning to ask themselves as they develop their inquiry capabilities. For earlier learners, the answers to these questions provide evidence about the degree to which they are thinking about the appropriate features of their work as it proceeds. Their very presence helps the students learn that these are questions that are important in inquiry, and they come to internalize them as they gain experience. For example, White and Frederiksen (1998) acquaint students with a concept they called "Being Systematic": "Students are careful, organized, and logical in planning and carrying out their work. When problems come up, they are thoughtful in examining their progress and deciding whether to alter their approach or strategy." As a Work Product, students rate their own solutions with respect to how systematic they were, on a 1-to-5 scale from "not adequate" to "exceptional."

11.5   Some Connections with Other Design Patterns

Model-based inquiry is an encompassing activity that draws repeatedly and cyclically on more specific aspects of model-based reasoning. When designing an inquiry task, a test developer can use this design pattern to consider the characteristics of Task Features and Work Products that will provide evidence about the movement in the larger space, and the specific *design patterns* to ensure that evidence is elicited as needed about details of the investigation.

The iterative testing and repairing that characterizes troubleshooting can be viewed as a special case of model-based inquiry. Steinberg and Gitomer's (1996) troubleshooting tasks in the hydraulic system of the F-15 aircraft, for example, required iterative cycles of model use, model evaluation, and model revision, with the efficiency of diagnostic tests at the crux of evaluation. The efficiency of tests for evaluating a model becomes particularly important in these more complex tasks. Efficiency is intimately related to understanding both the system in question and the tests that can be carried out, both Additional KSAs that are required jointly for effective troubleshooting. Frezzo, Behrens, and Mislevy (2009) showed how *design patterns* for creating troubleshooting tasks in network engineering are used in the Cisco Networking academy. Seibert, Hamel, Haynie, Mislevy, and Bao (2006) presented a more general design pattern that encompasses troubleshooting, called "Hypothetico-Deductive Problem Solving in a Finite Space."