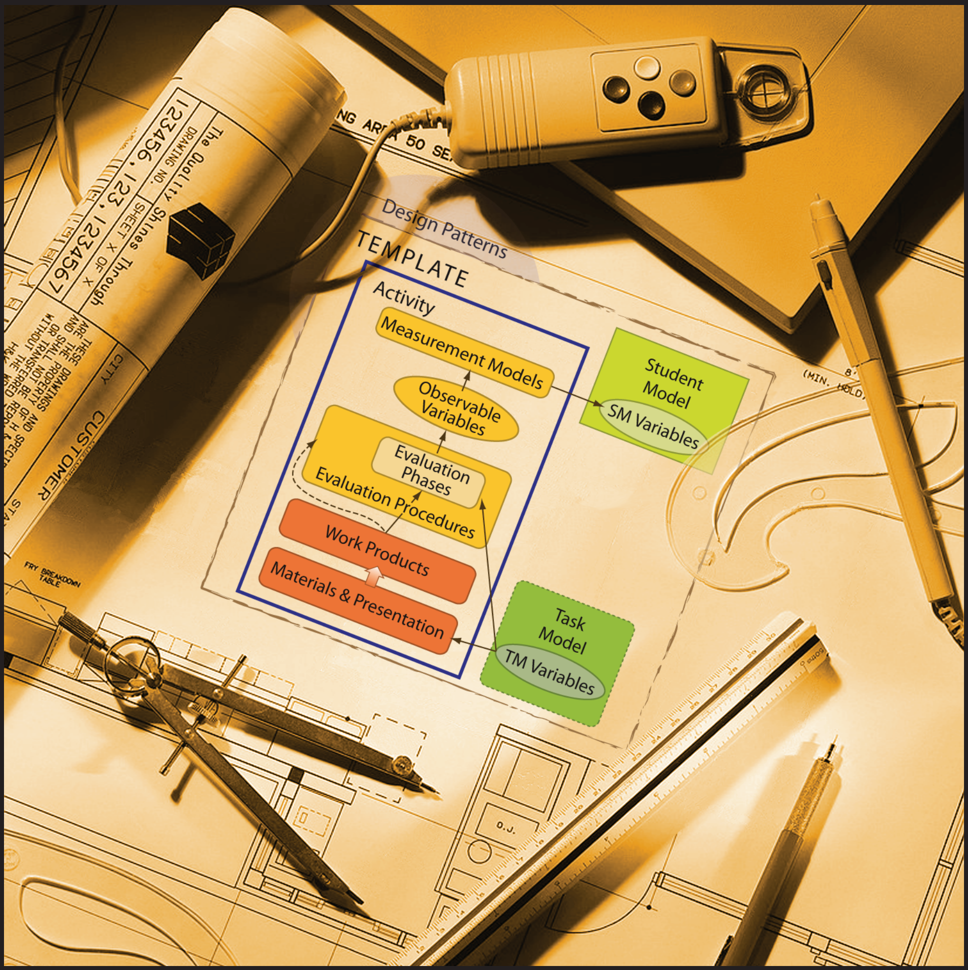


Large-Scale Assessment Technical Report 8|
November 2010

A Design Pattern for Experimental Investigation

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

Alexis Mitman Colker, SRI International Consultant
Min Liu and Robert Mislevy, University of Maryland
Geneva Haertel, Ron Fried, and Daniel Zalles, 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

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

TECHNICAL REPORT 8

A Design Pattern for Experimental Investigation

November 2010

Prepared by:

Alexis Mitman Colker, SRI International Consultant
Min Liu and Robert J. Mislevy, University of Maryland
Geneva Haertel, Ron Fried, and Daniel Zalles, SRI International

Acknowledgments
This material is based on work supported by the National Science Foundation under grant DRL-0733172 (An Application of Evidence-Centered Design to State Large-Scale Science Assessment).

CONTENTS

FIGURES

List Of Tables

ABSTRACT

The significance of inquiry skills is widely acknowledged in science practice across many areas. Carrying out experimental investigations is an indispensable element of scientific inquiry and, therefore, an important capability to assess. Drawing on research development in assessment design, this report provides a *design pattern* to help assessment designers create tasks assessing students' reasoning skills in experimental investigation. The *design pattern* lays out considerations regarding targeted knowledge and skills in this inquiry process, characteristics of situations in which student can evidence that skill, and ways of evaluating their work with specific examples.

1.0 Introduction

The National Science Education Standards (NSES) (National Research Council, 1996, 2000), in addressing reform of American science education, calls for an increased emphasis on inquiry. As the Chinese proverb states, “Tell me and I will forget; show me and I may remember; involve me and I will understand.” Through the activities of inquiry, students can come to more deeply understand scientific principles and concepts, and develop the reasoning and procedure skills that scientists use. More importantly, as mentioned in NSES, a mastery of inquiry can enable students to acquire new knowledge and tackle hard problems not only during their schools years but throughout their lives.

Among the diverse aspects of scientific inquiry, experimental investigation is the most rigorous and has been emphasized in K-12 education. Much science education research has been done to address experimental investigation in curriculum and instruction. How to validly and efficiently assess students’ capabilities with such complex inquiry processes in large-scale state-level testing remains a challenging issue.

This report addresses the design of tasks to provide evidence about students’ capabilities in experimental investigation in a way that supports such efforts across different science areas and levels of education. It draws on research on assessment design carried out under the evidence-centered design approach (ECD; Mislevy, Steinberg, & Almond, 2003; Mislevy & Haertel, 2006). Specifically, it presents a *design pattern* (Mislevy, et al., 2003) for assessing scientific reasoning skills in experimental investigation. *Design patterns* are a tool developed in the Principled Assessment Design for Inquiry (PADI) project, supported by the National Science Foundation (NSF). The particular *design pattern* for this report was constructed for the project, “Application of Evidence-Centered Design to State Large-Scale Science Assessment,” also supported by NSF to apply ECD in the context of a state-level large-scale accountability assessment in science. It is being applied in operational work by the committees of Minnesota item-writers (mostly current or retired Minnesota science teachers) who, with coordination, training, and support from Pearson and the staff of the Minnesota Department of Education, create the Minnesota Comprehensive Assessment (MCA-II) in science.

The following section sets the stage for the Experimental Investigation Design Pattern with background on ECD and PADI and then on the nature of experimental investigation. Next, attributes of the *design pattern* are discussed in detail and illustrated with example tasks.

2.0 Evidence Centered Design and Assessment Arguments

In order to show how *design patterns* support the authoring of tasks that assess students' capabilities in experimental investigation, we briefly review the ECD framework and Toulmin's (1958) structure for evidentiary argument and its relation to the *design pattern*.

2.1 Evidence–Centered Assessment Design

Evidence-centered assessment design (ECD) provides principles, patterns, and examples to guide task designers through articulating the theoretical foundation to the operational work of assessment development (e.g., item writing, directions, test administration, scoring procedures) (Mislevy, Almond, & Lukas, 2004). This structured framework explicates the assessment argument that underlies a task and, thus, enables designers to more efficiently manage the elements and underlying processes of assessment design.

ECD lays out the structure/process of an assessment design in terms of five layers that conceptualize different work being carried out by different experts or parties at different stages of design process (although in simple assessments, all may be done, usually implicitly, by the same person). Figure 1 summarizes the ECD layers in a way that reflects successive refinement and reorganization of knowledge about the content domain and the purpose of the assessment, from a substantive argument to the specific elements and processes needed in its operation.

As the first stage, *domain analysis* is about marshaling substantive information about the domain. It helps us 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, but most of the sources for analyzing a domain such as experimental investigation are neither originally created to support assessment nor presented in the structure of an argument. The cognitive research on experimental investigation discussed below and the identification of relevant Minnesota Academic Standards in Science[1] are examples of work in *domain analysis* to prepare for creating a *design pattern* to support task design for assessing these capabilities.

In the *domain modeling* layer, information identified in *domain analysis* is organized along the lines of assessment arguments. Without getting tangled in the technical details of assessment design and psychometric models, this layer directs researchers to clarify what is meant to be

1 http://education.state.mn.us/mdeprod/idcplg?IdcService=GET_FILE&dDocName=000282&RevisionSelectionMethod=latestReleased&Rendition=primary

assessed and how and why to do so. A tool for supporting work in *domain modeling*, *design patterns* (DPs) help the assessment designer think through the key elements of an assessment argument in narrative form. Details of *design patterns* will be given later in a section that reviews the attributes of a *design pattern* and in a section that discusses in detail the contents of the *design pattern* for assessing students' proficiencies with regard to experimental investigations.

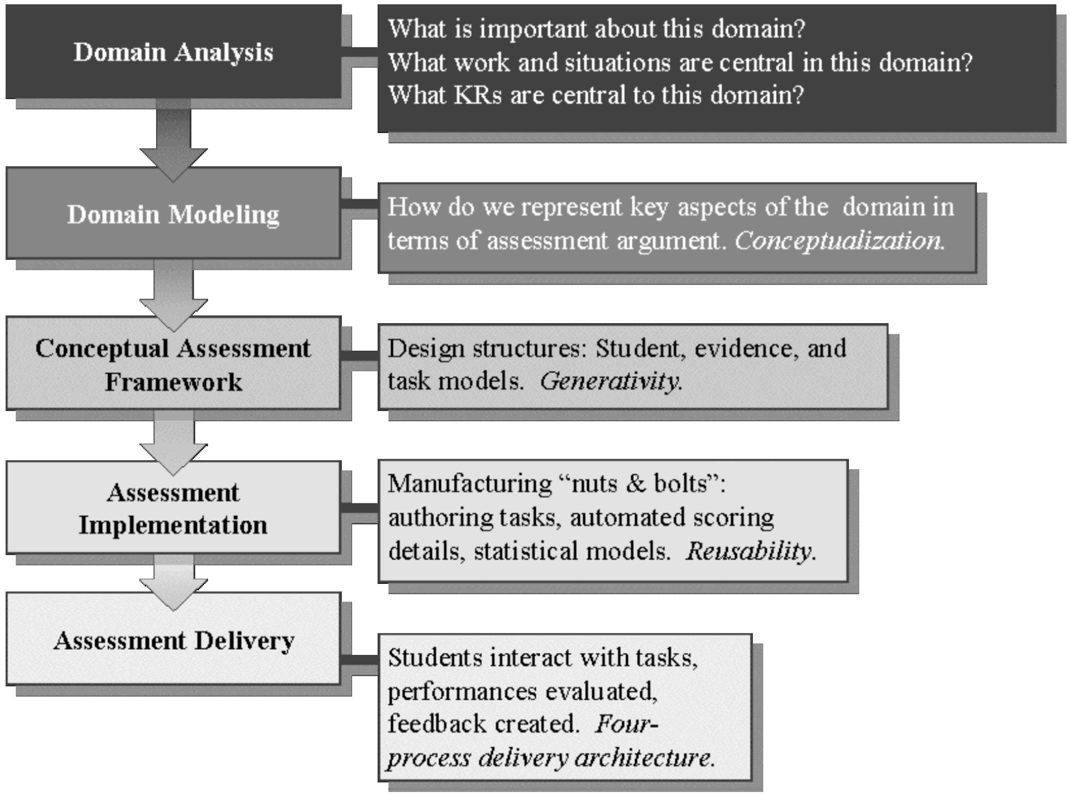
While the other three remaining layers of the ECD framework are less directly related to the creation of this *design pattern*, they are introduced for the sake of completeness. The reader is referred to Almond, Steinberg, and Mislevy (2002) and Mislevy and Riconscente (2006) for further discussion on these layers.

The *conceptual assessment framework (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 of task design, which is especially important for computer-based tasks.

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 previous stages.

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.

Figure 1. Layers of Evidence-Centered Design for Educational Assessment

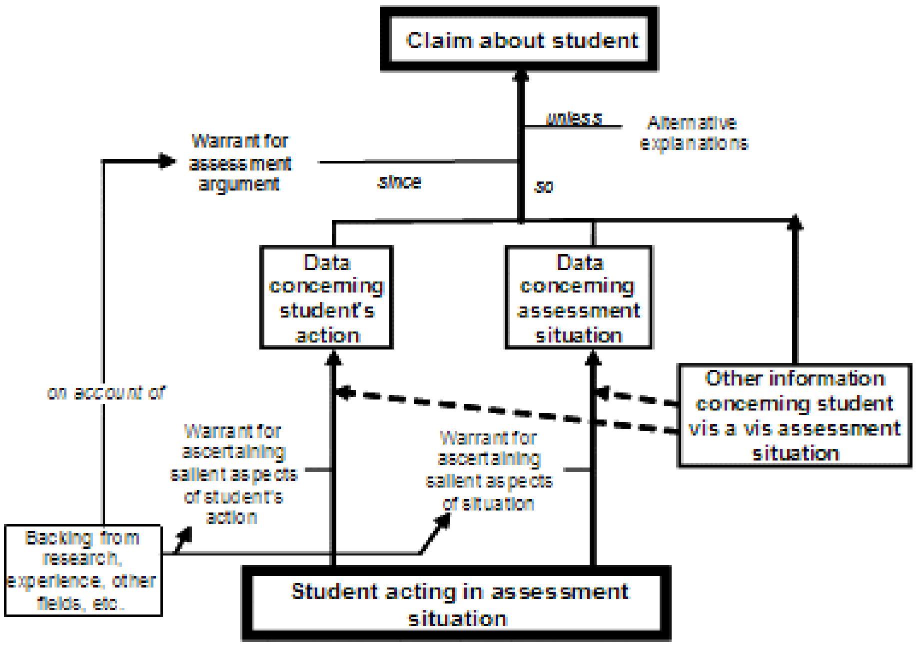


2.2 Assessment Arguments

An educational assessment can be viewed as an evidentiary argument that draws inferences from what students say, do, or make in task settings, to claims about what they can know, can do, or accomplish more generally (Messick, 1994). Toulmin (1958) provides a useful schema for the general structure of argument. Figure 2 adapts his terminology and representations to educational assessment arguments (Mislevy et al., 2003, 2006). In this diagram, a series of logically connected claims are 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; in the case of the MCA-II, presuming knowledge of science content from benchmarks at grade levels lower than the assessment at hand). The arrow going to the claim represents a logically reasoned inference by means of a *warrant*. The *warrant* posits how responses in situations with the noted features depend on proficiency. 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 are deficits in the knowledge or skills that are needed to carry out a task but are not focal to the claims.

Figure 2. A Toulmin Argument Diagram for Assessment Arguments



2.3 Attributes of a Design Pattern

Figure 2 indicates the structure of an assessment argument but not its content. A *design pattern* makes content or skill specific suggestions to guide task designers in categories that are related to the elements of an assessment argument. It thus can help task designers think through substantive aspects of the assessment argument. *Design patterns* can be used to fill the gap between academic content standards and specific assessments tasks. Although creating a *design pattern* may seem to be a time-consuming job, it can save time and energy in the long run by capturing design rationales in a re-usable and generative form. A *design pattern* also can smooth the transition to more technical work in the next ECD layers, by serving as a foundation for many tasks that must address key knowledge and skills in the domain. Furthermore, the experience and thinking captured in a *design pattern* provides shared information across applications, such as large-scale and classroom assessment, and assessment, instruction, and research. For these reasons, Mislevy and Haertel (2006) identified *design patterns* as a primary leverage point to improve design efficiency and validity in large-scale assessments.

A *design pattern* consists of attributes that can be associated with components of an assessment argument, as shown in Table 1. 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 those KSAs, and features of task conditions that can enable students to produce the evidence.

Table 1: Attributes of a Design Pattern[2]

Attribute	Definition	Assessment Argument Component
Name	Short name for the design pattern	
Summary	Brief description of the family of tasks implied by the design pattern	
Rationale	Nature of the KSA of interest and how it is manifest	Warrant
Focal KSAs	The primary knowledge/skill/abilities targeted by this design pattern	Claim
Supported Benchmarks	Benchmarks in the MCA-II test specifications corresponding to Minnesota Standards that this design pattern supports (specific to the MCA-II context)	Claim
Additional KSAs	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
Narrative Structures	Aspects of assessment situations that can be varied or combined to construct a storyboard outline.	Data concerning situation

[2] The *design pattern* attributes shown here differ slightly from the presentation in Mislevy et al. (2003). First, this table omits some less central attributes that appear in the extended version of the *design pattern* structure. Second, two additional attributes have been added specifically to support task design for the MCA-II: Science Assessment.

- *Supported Benchmarks* indicates those benchmarks from the MCA-II test specifications that correspond to standards that the design pattern supports.
- *Narrative Structures* has been added in view of the particular structure of tasks in the MCA-II: Science Assessments. MCA-II science tasks are multi-item clusters that revolve around an incident, investigation, data set, or some other unifying context. These are called storyboards. Narrative Structures are story frameworks that are analogous to basic plotlines for movies. Examples are "General to Specific" and "Cause and Effect."

3.0 Experimental Investigation

This section reviews the importance of experimental investigation in science and briefly defines the process of experimental investigation. The outcomes of an alignment study are noted, which indicate that a *design pattern* to support authoring tasks on experimental investigation can help bridge Minnesota and national content standards with the creation of tasks to assess these capabilities.

3.1 Experimental Investigation in the Inquiry Process

Experimental investigation is the inquiry approach favored in the sciences such as biology, chemistry, medicine, psychology, and physics. At its simplest, an experiment involves testing a proposed causal relationship (i.e. hypothesis) by manipulating one or more so-called independent variables to determine the effects on another, so-called dependent variable. This contrasts with the observational investigation methods commonly used, say, in astronomy and geology, where manipulation of variables typically is not possible. (For a *design pattern* focused on observational investigation, see Mislevy, et al., 2009).

No one individual or set of individuals clearly can be credited with the development of experimental investigation. We know that during the Renaissance, advances in instrumentation and recognition of serious pitfalls stemming from a sole reliance on inductive reasoning led to what we would today recognize as forays into conducting experiments. Furthermore,

> Despite the creative use of experimental design features from the seventeenth century onward, it was not until the past century or so that experimental design notions became systematized. This systematization at first emphasized physical control of conditions—isolation, insulation, sterilization, strong steel chamber walls, soundproofing, lead shielding against Hertzian waves, and so forth. Much more recently, as biological research moved from the laboratory to the open field, the modern theory of experimental control through randomized assignment to treatment emerged. (Cook & Campbell, 1979, p. 4).

To facilitate students learning science as a way of knowing instead of memorizing scientific facts, many science educators suggest that students should be introduced to the reasoning and methods of experimental investigation as early as possible (e.g., Wagner, 1983; Hammrich, 1997; American Association for the Advancement of Science, 1993). Both the NSES (National Research Council, 1996) and the Minnesota Science benchmarks indicate that experimentation should be included as a major approach to scientific inquiry and thus emphasized in standards from the elementary grades through high school. More related details will be described in the next section.

An experiment usually consists of three key elements (Cox & Reid, 2000): the experimental units (e.g., plants), the manipulation of the independent variable or treatment (e.g., amount of fertilizer), and the measured difference in the dependent variable or response (e.g., growth of plants). Manipulating the independent variable involves procedures that are to be applied to each experimental unit, and they can be varied qualitatively or quantitatively. The dependent variable specifies the criterion that is supposed to be causally affected by the independent variable(s). The simplest form of an experiment manipulates one independent variable at only two levels so that there is, a treatment and a control (e.g., a fixed amount of fertilizer versus no fertilizer). More complex experiments might include a collection of different treatments to form combination of levels of different treatments. It is worth noting that in the MCA-II, students at the eighth grade and beyond are required to address investigations with up to two independent variables and know the importance of manipulating one variable at a time to capture its influence on the dependent variable.

For students who conduct experimental investigations, the biggest challenge is how to carry out the initial reasoning processes, including breaking down an area of informed scientific speculation into one or more testable hypotheses, operationalizing the hypothesis through a procedure in a particular type of experiment, collecting observations, and relating it back to experimental hypothesis in order to obtain a valid causal conclusion. A complete cycle in experimental investigation includes the following:

- Identify aspects of phenomena to be investigated experimentally.
- Propose a testable hypothesis.
- Design and conduct an experiment to ensure a valid test of the cause-effect relationship of interest.
- Collect, analyze, and interpret data with appropriate tools.
- Develop explanations or models using logic and evidence.

In this cycle, if the experimental results fail to support the proposed hypothesis, the hypothesis is typically revised, with a new experimental cycle resumed from the second step. These key steps of experimentation are shared by the different content areas that employ the experimental method although the nature of the variables, measurements, and the procedures will vary accordingly.

Besides the basic cycle of experimental investigation and the notion of testing the relationship between independent and dependent variables, there are several other concepts critical to well-designed experiments—concepts that one should expect students to gain increasing familiarity with as they progress through school. These include the ideas of controlling variables, repeatability, and random assignment.

The notion of controlling variables is extremely important because in many experiments, it is often difficult to isolate or neutralize the effect of variables other than the one being manipulated and thus claim that this independent variable is the only cause for the change in the dependent variable. When one can identify a number of non-treatment variables in an experimental situation that might possibly influence the dependent variable (so-called extraneous or confounding variables), one should attempt to control these variables by holding them constant for all treatment levels in the experiment. In the example of testing the impact of fertilizer (the independent variable) on plant growth (the dependent variable), possible confounding variables would be amount of light and water. One would control these variables by making sure that every plant in the experiment received the same amount of light and water, letting only the amount of fertilizer vary. Of course it is also possible to identify variables that should be inconsequential in the experiment (in our example, say, the color of the watering can or the particular faucet in the room from which the water is drawn). Being able to distinguish between inconsequential variables and confounding variables is important not only for being able to maximize the design and execution of an experiment but also to guide the interpretation of experimental results.

The concept of repeatability is important in experiments because for a relationship to be truly causal, one should be able to reproduce this relationship consistently. This means that a cause (the manipulation of the independent variable) should always lead to an effect (on the dependent variable) of the same direction and similar magnitude. If the effect is inconsistent when researchers attempt to replicate the study, then the proposed causal relationship is in jeopardy. Repeatability is often achieved by having more than one experimental unit receive the treatment (e.g. ten plants receive the fertilizer and ten do not) and by conducting the same experiment at different points in time.

Randomized assignment to treatment conditions (resulting from the manipulation of the independent variable) is a common, systemized technique in modern experimental investigation. The notion behind randomization is that it is a primary way to remove bias or systematic error that may result from possible confounding variables that cannot be controlled or from error resulting from unanticipated sources of variation. In the example of the plant and fertilizer experiment, say, if 20 plants are going to be used—half for the treatment group and half for the non-treatment group—it makes sense to randomly assign the plants to the two groups because even among the same species of plant, there may be variations causing some plants to grow more vigorously than others. In the non-physical sciences such as psychology, where people often are the treatment unit, it is common practice to assign individuals randomly to treatment groups in order to try to "equalize" groups on a number variables (e.g., an individual characteristic such as years of schooling) that may be important but cannot be practically controlled or fully anticipated. Thus, randomization is an important technique for ensuring the validity of causal relationships tested in experiments.

Experimental investigation is always carried out in a particular context. Sufficient knowledge about the content domain at issue is requisite for students conducting an investigation in a given domain. Familiarity with the relevant scientific knowledge can facilitate a student raising an appropriate testable question to be investigated, conducting the experiment appropriately, identifying relevant features of data from irrelevant ones, collecting and displaying supporting or refuting data with techniques and tools, and formulating an appropriate explanation. None of these things can be done without knowledge of the content domain. However, it also should be noted that content knowledge can sometimes lead an experimenter to operate from biased assumptions (see Fleck, 1935; Mayo, 1996), perhaps leading them to overlook competing hypotheses, interpretations, and explanations.

3.2 Minnesota Science Standards and National Science Standards

The present project aims to illustrate the use of ECD in the MCA-II science assessments in ways that not only benefit the MCA-II but hold value in the larger science education and assessment communities. Ideally, the *design pattern* we create can support aspects of science learning reflected both in national standards and in Minnesota standards and benchmarks. To this end, a systematic alignment study was carried out among Minnesota middle school science benchmarks and the NSES (NRC, 1996). Science inquiry skills are highlighted in both sets of science education standards.

The NSES (NRC, 1996) emphasizes the importance of unifying concepts and processes being shared by different scientific disciplines because they provide schemas that help students understand natural phenomena both within and across areas (p. 105). This unifying theme is also implied in Strand I of the Minnesota Academic Standards for Science, "Nature and History of Science." More specifically, NSES clearly stresses the ability to conduct scientific inquiry and understand it across content areas and grades (p. 106). In the earliest grades (K-4), students are required to design and conduct simple investigations (but not formal experiments) to answer the questions they raised and to use simple instruments to collect data, from which they formulate reasonable explanations (pp. 122-123). In middle grades (5-8), students should be able to identify testable questions with quantitative relationships, design and execute investigations through their general abilities, use tools and techniques including computers to collect, summarize, and display data, and finally develop evidence-based explanations (p. 145). In high school, students are expected to be able to formulate a testable hypothesis, design and conduct an investigation with clarification of the method, controls, and variables; collect and analyze data by means of advanced technologies and mathematics; and then formulate and revise scientific explanations and models based on evidence and logic (p.175). When these standards refer to scientific investigations, it is clear that these investigations can include both those that are observational and those that are experimental, with a shift towards the more advanced understandings and techniques of experimental design occurring at the higher grades.

Our alignment study showed that the Minnesota Academic Standards for Science also emphasize inquiry skills by specifying a sub-strand titled “Scientific Inquiry” separately. As Table 2 indicates, the Minnesota benchmarks also indicate that students be capable of various investigation skills at varying difficulty levels for different grades. Many of these benchmarks refer to investigation more broadly, but a subset of these refers to experimental investigation specifically. For the first block of benchmarks (Grades 3-5), students in Grade 4 are expected with guidance to be able to collect, organize, analyze, and present data from an experiment. By Grade 5, they can perform an experiment with step-by-step support and guidance (see 4.I.B.2, 5.I.B.1). In grades 6-8, students’ understanding of experimental investigation advances to include the presentation of experimental data through multiple representations, manipulating one variable at a time, and differentiating among variables to be changed, controlled, and measured (see 6.I.B.4, 7.I.B.2, 8.I.B.3). The last block for high school requires students to design and complete an experiment using scientific methods (see 9-12.I.B.1). The Minnesota requirements become increasingly technical and demanding in knowledge and skills as the grade levels increase, and this includes increasing sophistication about experimental investigation.

Table 2. Minnesota Benchmarks Related to Scientific Investigations

Grades 3-5:

- 3.I.B.1: The student will ask questions about the natural world that can be investigated scientifically.
- 3.I.B.2: The students will participate in a scientific investigation using appropriate tools.
- 3.I.B.3: The student will know that scientists use different kinds of investigations depending on the questions they are trying to answer.
- 4.I.B.1: The students will recognize when comparisons might not be fair because some conditions are not kept the same.
- 4.I.B.2: The student will collect, organize, analyze and present data from a controlled experiment.

- 4.I.B.3: The students will recognize that evidence and logic are necessary to support scientific understandings.
- 5.I.B.1: The students will perform a controlled experiment using a specific step-by-step procedure and present conclusions supported by the evidence.
- 5.I.B.2: The student will observe that when a science investigation or experiment is repeated, a similar result is expected.

Grades 6-8

- 6.I.A.2: The student will explain why scientists often repeat investigations to be sure of the results.
- 6.I.B.1: The student will identify questions that can be answered through scientific investigation and those that cannot.
- 6.I.B.3: The student will use appropriate tools and international system units for measuring length, time, mass, volume and temperature with suitable precision and accuracy.
- 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.
- 7.I.A.2: The student will explain natural phenomena by using appropriate physical, conceptual and mathematical models.
- 7.I.B.1: The student will formulate a testable hypothesis based on prior knowledge.
- 7.I.B.2: The students will recognize that a variable is a condition that may influence the outcome of an investigation and know the importance of manipulating one variable at a time.
- 7.I.B.3: The student will write a specific step-by-step procedure for a scientific investigation.
- 7.I.B.4: The student will explain how classroom scientific investigations relate to established scientific investigation.
- 8.I.A.2: The student will explain the development, usefulness and limitations of scientific models in the explanation and prediction of natural phenomena.
- 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.
- 8.I.B.2: The student will describe how scientists can conduct investigations in a simple system and make generalizations to more complex systems
- 8.I.B.3: The student will specify variables to be changed, controlled and measured.
- 8.I.B.4: The student will use sufficient trials and adequate sample size to ensure reliable data.

High School:

- 9-12.I.B.1: The student will design and complete a scientific experiment using scientific methods by determining a testable question, making a hypothesis, designing a scientific investigation with appropriate controls, analyzing data, making conclusions based evidence and comparing conclusions to the original hypothesis and prior knowledge.
- 9-12.I.B.2: The student will distinguish between qualitative and quantitative data.
- 9-12.I.B.3: The student will apply mathematics and models to analyze data and support conclusions.
- 9. -12.I.B.4: The student will identify possible sources of error and their effects on results.
- 9-12.I.B.6: The student will give examples of how different domains of science use different bodies of scientific knowledge and employ different methods to investigate questions.

4.0 Design Pattern for Experimental Investigation

This section presents the Experimental Investigation *design pattern* and illustrates some of its attributes with exemplar tasks. This *design pattern* is related to a *design pattern* for observational investigation (Mislevy, et al., 2009), another form of science inquiry. For users who are interested in the latter or in using the two inquiry *design patterns* in a complementary way, an online version of the observational investigation *design pattern* can provide more details.[3]

In this project, the target users of the *design pattern* are the storyboard and item writers who create the MCA-II and would directly use the *design pattern* to support their work. Other users within the MCA-II context will be the Pearson professional test developers who structure the authoring and assembly of the MCA-II, train the storyboard and item writers, and edit and refine their products as needed. Minnesota expert review panels also are potential users because they examine storyboards and items for content and appropriateness. Secondary user groups in Minnesota would be classroom teachers and curriculum developers who would be able to use *design patterns* to create classroom tasks and curriculum-embedded tasks that address the same standards that the MCA-II addresses, but in less constrained contexts. These uses lie outside the current project, but they constitute an opportunity to improve the alignment of instruction and large-scale accountability assessment at the level of the targeted science standards rather than at the level of specific test items. More broadly, educators and researchers in the science education community can also use these products as a reference to understand how a *design pattern* helps writers produce tasks validly and efficiently for assessing students' scientific reasoning skills, such as those in experimental investigations.

An assessment writer needs to know that not all experimental investigations have all phases of an experiment, and phases can appear iteratively. A given task or storyboard can focus on just one phase, a transition between phases, or working through multiple phases. This is a design decision that is up to the storyboard writer. Again, the *design pattern* does not make the decision, but it makes clear that there is a decision to be made and provides information to inform it.

Table 3 presents a print-based "writer-friendly" version of the *design pattern*. It is intended to be brief so that storyboard and item writers can easily reference the central ideas. In its other electronic (i.e. online) form, many of these summarized points are accompanied by links to examples or more detailed discussion. In the table, "*D"* denotes the availability of further detail for an entry, and "*E"* indicates a hyperlink to an example. In this way, the user has a brief form of the *design pattern* immediately at hand but also has access to further detail should he or she want it.

[3] http://design-drk.padi.sri.com/padi/do/AddNodeAction?NODE_ID=2167&state=viewNode

This feature is meant to offer further support particularly to storyboard and item writers who are new to the MCA. For more experienced writers, these links may serve to prompt and confirm previously learned information and provide examples for inspiration.[4] The following discussion includes some of the additional material in those links.

Table 3. Writer-Friendly Version of Design Pattern for Experimental Investigation

Overview	This design pattern supports the writing of storyboards and items that address scientific reasoning and process skills in experimental investigations. In experimental investigations, it is necessary to manipulate one or more of the variables of interest and to control others while testing a prediction or hypothesis. This contrasts with observational investigations, where variables typically cannot be manipulated. This design pattern may be used to generate groups of tasks for science content strands amenable to experimentation.
Use	This design pattern supports the construction of tasks that address experimental investigations - that is, investigations where experimental methods are appropriate (as compared with investigations where only observations of phenomena are possible). In order for students to have a well-rounded understanding of the scientific method, they need to be familiar with the context and methods of experimental investigations.
Focal KSAs	1. Ability to distinguish between experimental and observational methodology *D* 2. Ability to recognize that when a situation of scientific interest includes aspects that can be altered or manipulated practically, it is suitable for experimental investigation *D* 3. Ability to recognize that the purpose of an experiment is to test a prediction/hypothesis about a causal relationship *D* 4. Ability to identify, generate or evaluate a prediction/hypothesis that is testable with a simple experiment *D* 5. Ability to plan and conduct a simple experiment step-by-step given a prediction or hypothesis *D* 6. Ability to recognize that at a basic level, an experiment involves manipulating one variable at a time and measuring the effect on (or value of) another variable D 7. Ability to identify variables of the scientific situation (other than the ones being manipulated or treated as an outcome) that should be controlled (i.e. kept the same) in order to prevent misleading information about the nature of the causal relationship *D* 8. Ability to recognize variables that are inconsequential in the design of an experiment 9. Ability to recognize that steps in an experiment must be repeatable to reliably predict future results 10. Ability to recognize that random assignment to treatment conditions (i.e. levels of the independent variable) is a primary way to rule out alternative explanations for a causal relationship 11. Ability to interpret or appropriately generalize the results of a simple experiment or to formulate conclusions or create models from the results

[4] An "in progress" version is currently available on line at http://design-drk.padi.sri.com/padi/do/AddNodeAction?NODE_ID=2245&state=viewNode

Activated Benchmarks	This design pattern can be used to support writing storyboards and items for the following benchmarks. ▪ Grades 3-5: 3 4.I.B.2, 5.I.B.1 ▪ Grades 6-8: 6.I.B.4, 7.I.B.2, 8.1.B.3 ▪ Grades 9-12: 9-12.I.B.1
Additional KSAs	1. Content knowledge (may be construct relevant) *D* 2. Prerequisite knowledge from earlier grades *D* 3. Prerequisite experience assessing or conducting component steps of an investigation *D* 4. Ability to collect, organize, analyze, and present data 5. Familiarity with representational forms (e.g., graphs, maps)
Characteristic Features	Storyboards and items written using this design pattern will exhibit one or more of the following features: 1. Focus on Nature of Science (Strand I in MCA) benchmarks that relate to experimental investigations at the appropriate grade level. 2. Presentation of situation of scientific interest where variables can be (or have been) practically altered to address a causal prediction. *D* 3. Presentation of situation requiring the design or conduct of a controlled experiment D 4. Presentation or representation of an experimental design 5. Presentation of observed result from an experiment requiring the development of explanations, conclusions, or models *D*
Variable Features	The following features are variable depending on the storyboard and items: 1. Content (strand) context *D* 2. Which one or multiple phases of experimental investigation will be addressed 3. Qualitative or quantitative investigation or a combination 4. Ease or difficulty with which the treatment (independent) variable can be manipulated 5. Are manipulated variables given or to be determined? 6. The number of variables investigated and the complexity of their interrelationships (up to 2 independent variables in Grade 8 MCA-II) 7. Number of variables that need to be controlled to unambiguously study the relationship between the manipulated variable and the outcome variable *D* 8. Length of time over which the experiment much be conducted in order to study the potential impact of the treatment variable 9. Data representations *D*
Potential Work Products (MC questions, open-ended responses, figural responses)	1. Select, identify, or evaluate a measurable investigable question D 2. Identify or differentiate independent and dependent variables in a given scientific situation. 3. Identify or differentiate variables that do and do not need to be controlled in a given scientific situation 4. Complete some phases of experimentation with given information, such as selecting levels or determining steps. 5. Generate or identify data pattern from results in a simple experiment 6. Generate an interpretation/explanation/conclusion from a set of experimental results 7. Critiques of peers on their choice of experimental procedures or explanations of experimental results D 8. Given an experiment with unexpected or confusing results, identify possible reasons D

Potential Observations	1. Accuracy in identifying situation suitable for experimental investigation 2. Plausibility of a measurable research question being raised 3. Plausibility of hypothesis as being testable by a simple experiment 4. Plausibility/correctness of design for a simple experiment 5. Correct identification of independent and dependent variables 6. Accuracy in identifying variables (other than the treatment variables of interest) that should be controlled (held constant) or made equivalent (e.g., through random assignment) 7. Plausibility/correctness of steps to take in the conduct of an experiment 8. Plausibility of plan for repeating an experiment 9. Correctness of recognized data patterns from experimental data 10. Plausibility/correctness of interpretation/explanation of experimental results 11. Accuracy in critiques of others (hypothetical in a standard assessment, real in classroom work) re the above potential observations.
Narrative Structures	1. Investigation 2. Cause and effect 3. Change over time

Note: *D* indicates a hyperlink to extended detail or discussion of the entry, and *E* indicates a hyperlink to an example task that illustrates the point.

4.1 Overview and Use

The Overview and Use attributes of this *design pattern* explain briefly that it is meant to support writing assessment storyboards and items that address aspects of reasoning in experimental investigation in science. The distinction between experimental and observational investigation is specifically noted.

The entries in the Overview and Use attributes makes clear that this is a content-neutral *design pattern* in the sense that it can be used in conjunction with any science content for which such investigations can be carried out. Therefore, it supports writing tasks that address the nature and concepts of this aspect of scientific reasoning or assessing the skills in the context of a particular content-specific investigation. In the MCA II, for example, a storyboard could be built around experimental investigation as motivated by a Nature of Science benchmark, as it applies to particular models and processes from, say, Physical Sciences or Natural Sciences benchmarks.

4.2 Focal KSAs, Supported Benchmarks, and Characteristic Task Features

The primary attribute of a *design pattern* is the Focal KSAs. The Focal KSAs are the targets of inferences that assessors aim to make in an assessment, concerning some aspect(s) of proficiencies. A *design pattern* sometimes designates a group of related KSAs, as does this one. A task designer needs to decide whether to test all of these KSA entries as a composite or to emphasize smaller groups of them.

Based on the analysis for experimental investigation and motivated by the benchmarks in the MCA-II Test Specifications, this *design pattern* specifies as Focal KSAs a broad set of cross-disciplinary knowledge, skills, and abilities that students need to exercise when pursuing an experimental investigation. As noted earlier, however, delineating these Focal KSAs is *not* meant to imply that these are skills that students possess in isolation of actual scientific content. Rather, they are aspects of the scientific activity that are pursued and are integral to the content being investigated. The KSAs encompass the indicated benchmarks shown in the *design pattern* at a more overarching level than the benchmarks or standards themselves (see Table 2), so that they are in accord more closely with the unifying themes and inquiry skills emphasized in NSES, and connect more strongly to the research base of experimental investigation.

The *design pattern* lists eleven entries under the Focal KSAs. The first two entries concern whether a student can recognize a situation in which experimental investigation is suitable, particularly in contrast to situations suitable only for observational investigation or not suitable for any systematic investigation. The third and fourth Focal KSAs have to do with the student understanding that an experiment requires a prediction or hypothesis about a causal relationship such that the student should be able to identify, generate, or evaluate such a prediction/hypothesis. The fifth Focal KSA concerns the ability to plan the steps necessary to carry out a simple experiment. The next three Focal KSAs (6-8) have to do with knowledge about the types of variables that can be involved in an experiment. Specifically, there is the expectation that the student be able to identify and distinguish between independent (the manipulated) and dependent (the outcome) variables. Also, there is the expectation that students be able to think about variables that may need to be controlled or that have no bearing on a possible experimental result. The ninth Focal KSA concerns the knowledge that an experiment needs to be repeatable in order for its results to be considered reliable. The tenth Focal KSA has to do with knowledge about the technique of random assignment to levels of an independent variable. Last, Focal KSA 11 has to do with the ability to interpret and draw conclusions (or create models) from experimental results. Note that several of the Focal KSAs (1, 4, 5, and 11) are directly linked to a student's ability to engage in the cycle of experimental investigation described earlier (Section 3.1). The remaining Focal KSAs have to do with knowledge of and reasoning about constructs involved in the successful design, implementation, and interpretation of an experiment.

The *design pattern* attribute of Supported Benchmarks is next, and it is directly linked to the Focal KSAs. We already have discussed the relevant NSES standards and Minnesota benchmarks (see Section 3.2 and Table 2) and have noted that while most of these speak to scientific investigations more broadly, specific Minnesota benchmarks clearly identify working with experimental methods. These are the benchmarks listed in the writer-friendly version of the *design pattern* in Table 3. (In the online version of this design pattern, more broadly relevant NSES standards and Minnesota benchmarks for the middle grade levels are listed.) In general, it is worth noting that the Focal KSAs are written at a more detailed level than the benchmarks as is appropriate for guiding assessment designers toward specific choices for families of tasks.

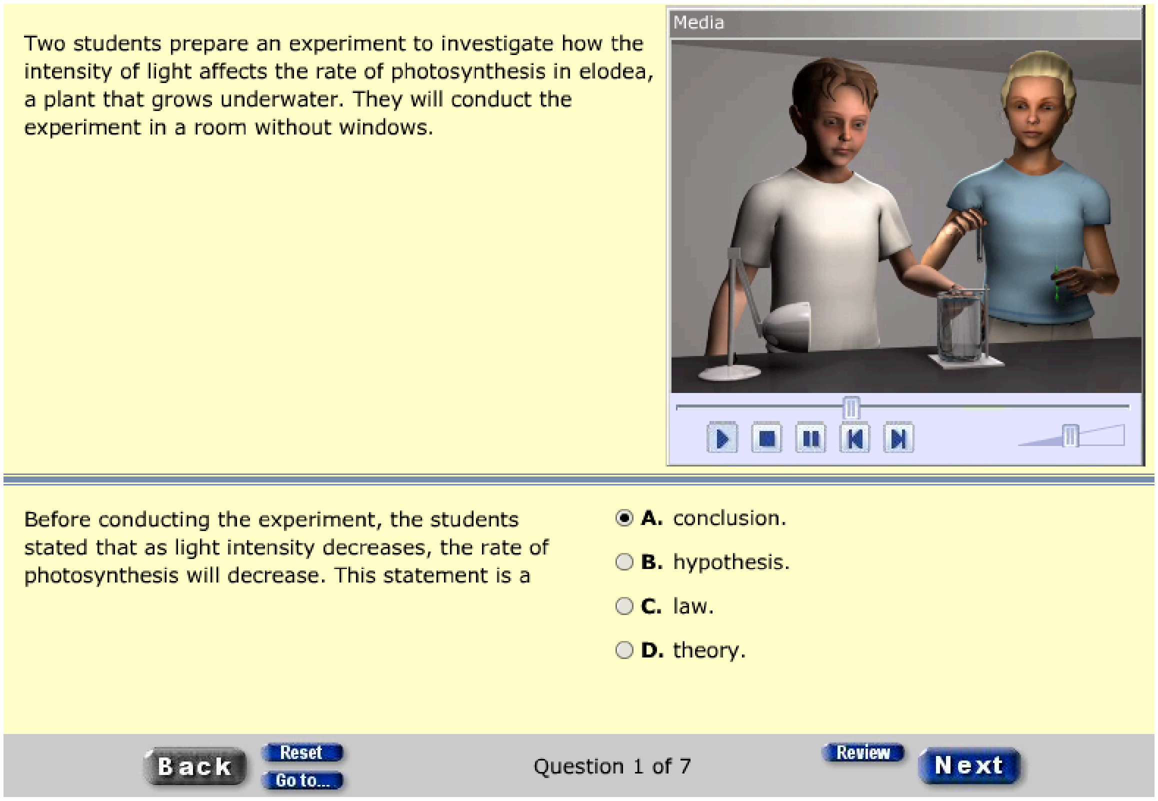
After Focal KSAs are identified, we can think of what characteristics of tasks are necessary to evoke students' demonstration of such knowledge, skills, and abilities. Hence, we need to turn to another key attribute of *design patterns*, Characteristic Features. All assessment tasks motivated by this *design pattern* need to incorporate some aspects of these features in order to evoke evidence about the Focal KSAs.

Five entries are listed under Characteristic Features. The first entry specifically references the alignment to the Minnesota benchmarks concerning experimental investigation. The remaining four Characteristic Features all reflect the fact that tasks for this design pattern would have to include a situation suitable for experimental investigation. Beyond this, the features reference various stages in the experimental investigation process with the expectation that not all tasks would necessarily involve all phases of the process. The second and third Characteristic Features concern the presentation of a situation that can address a causal prediction/hypothesis and lead to the design or conduct of a controlled experiment. The fourth Characteristic Feature refers to a task that would include (or begin) with the presentation of an experimental design. The last Characteristic Feature involves the presentation of an experimental result that students would have to interpret in some way.

As an illustration of how the Focal KSAs and Characteristic Features from the design pattern can play out in an actual assessment, consider a Minnesota sample task on "Photosynthesis Investigation" (see Figure 3). Here, experimental investigation is applied to a topic in life science. Note that there are 7 scenes for this task, and that every scene is accompanied by an item except for the third scene, which illustrates how students would observe the outcome variable to be measured. Furthermore, items for Scenes 4 and 5 focus on topic content knowledge (plant cells, the importance of photosynthesis), while the item for Scene 6 taps knowledge about graphing. The attention to other kinds of Focal KSAs (which would appear in other types of design patterns) is not unusual in a multistep complex task because each task typically has to address a number of different benchmarks that are not necessarily thematically linked. In any case, it is Scenes 1, 2, and 7 that have items relevant to experimental investigation.

Figure 3. Minnesota Task Sample "Photosynthesis Investigation" (For an interactive version of this task and many others, see http://www.pearsonaccess.com/cs/Satellite?c=Page&childpagename=Minnesota%2FmnPALPLayout&cid=1205461255328&p=1205461255328&pagename=mnPALPWrapper&resourcecategory=Item+Samplers

Scene1:



Scene2:

A sprig of elodea is placed upside down in a test tube filled with a solution. The test tube is placed in a large beaker full of water, which serves as a water bath. Photosynthesis will produce bubbles of gas. The bubbles will eventually come out of the stem and rise to the top of the test tube.

The students will use a light source during their experiment. The light source will be moved closer to and farther away from the plant to change the intensity of light it receives. The rate of photosynthesis will be measured by counting how many bubbles come out of the stem in 1 minute.

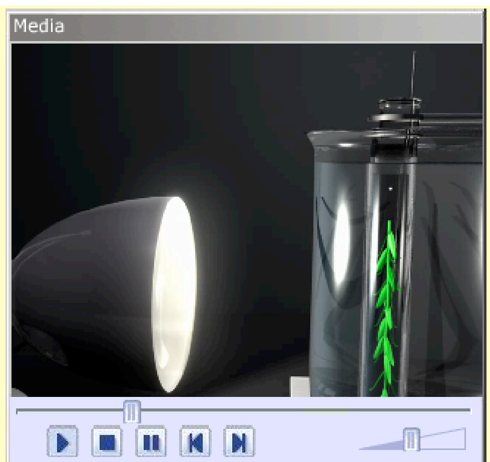
Which of these variables would be most important to keep constant during the experiment?

- (●) **A.** Number of bubbles produced by the elodea
- ( ) **B.** Distance of the light from the test tube
- ( ) **C.** Temperature of the solution in the test tube
- ( ) **D.** Time of day the experiment is conducted

Back | Reset | Go to... | Question 2 of 7 | Review | Next

Scene 3:

After the students dimmed the room's lights and turned on the light source, they began their experiment. This is what they saw when the light source was placed 15 centimeters from the plant. They counted the number of bubbles coming from the plant in 1 minute.



Click the Next button to go on.

Back | Reset | Go to... | Question 3 of 7 | Review | Next

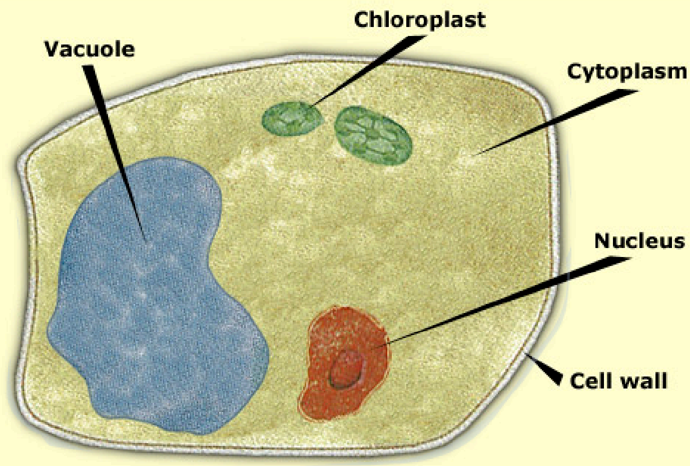
Scene 4:

Next, the students viewed a prepared slide of elodea cells. The figure represents what the students saw when they viewed the cells under a microscope.

Media

Choose the organelle whose primary function is photosynthesis.

Click on the diagram to put a "+" on an organelle.



Back | Reset | Go to... | Question 4 of 7 | Review | Next

Scene 5:

The bubbles produced by the elodea are the result of photosynthesis.

Media

Explain the importance of photosynthesis in food chains.

Type your response in the space below.

Back | Reset | Go to... | Question 5 of 7 | Review | Next

Scene 6:

Each trial was replicated 3 times and the average of the results was calculated. The results of the experiment are shown in the table.

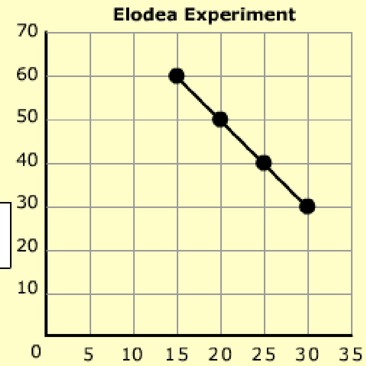
Media

The figure shows the data in a graph. In the areas indicated, insert a label for each axis of the graph. Be sure to include units of measure.

Bubble Production at Various Distances

Distance of Light from Plant (cm)	Bubbles Produced per Minute (Average Number)
15	60
20	50
25	40
30	30

Y Axis Label



X Axis Label

Back Reset Go to... Question 6 of 7 Review Next

Scene 7:

Each trial was replicated 3 times and the average of the results was calculated. The results of the experiment are shown in the table.

Media

Based on the data, describe the effect of light intensity on photosynthesis.

Type your response in the space below.

Bubble Production at Various Distances

Distance of Light from Plant (cm)	Bubbles Produced per Minute (Average Number)
15	60
20	50
25	40
30	30

Back Reset Go to... Question 7 of 7 Review Next

Scene 1 sets up an experimental situation (thus, the situation is given). The item for Scene 1 (asking student to recognize that a statement is a hypothesis) is related to Focal KSAs 3, 4, an 6 (i.e., knowing than an experiment should involve a hypothesis, being able to identify a hypothesis, and knowing that a hypothesis will include talking about how a manipulated independent variable — light intensity — impacts a dependent variable — rate of photosynthesis). The item for Scene 2 asks students to reason about what variable (from among a list) should be controlled in the experiment. Focal KSAs 6, 7, and 8 are involved here because the student needs to understand all the types of variables that can be noted in an experiment (i.e., independent, dependent, control, and inconsequential). Scene 7 presents students with a table of results for the study (showing the levels of the independent variable in one column and the measured outcome variable in the second column) and asks them to construct a response to interpreting these results in terms of a possible causal relationship. This item clearly involves Focal KSA 11 (interpreting the results of an experiment), but it also involves Focal KSAs 3 and 4 (linking back to the hypothesis of the study) and FKSA 6 (knowing that it is the independent and dependent variables being represented in the table).

Looking at the "Photosynthesis Investigation" from the standpoint of Characteristic Features, this task encompasses all of the Characteristic Features from the Experimental Investigation Design Pattern. Scenes 1 through 3 present the experimental situation and describe in some detail how the experiment is going to be carried out (Characteristic Features 2, 3, and 4). Scene 7 presents the results from the experiment, thus linking to Characteristic Feature 5. Last, this assessment storyboard aligns with the broader MCA-II middle school benchmarks about experimental investigation — i.e. Characteristic Feature 1. In sum, most of the scenes from this storyboard have the kinds of features one would expect of a task focused on eliciting from students their knowledge about experimental investigation – in this case, in the specific context of the topic of photosynthesis.

4.3 Additional KSAs and Variable Task Features

Additional KSAs of a *design pattern* are the other knowledge, skills, and abilities that may be involved in a task that assesses the Focal KSAs. Categories of Additional KSAs include things like prerequisite knowledge, content knowledge, familiarity with task tools, task expectations and formats, and the cognitive and physical capabilities needed to apprehend, interact with, and respond to a task. Additional KSAs can thus have positive, negative, or evidence-conditional effects on the validity of the assessment argument. They are included in the *design pattern* to alert the task developer to possible validity threats. Whether or not these KSAs are demanded by a task, and to what degree, will be affected by the writer's choice of Variable Task Features. Variable Features of tasks are a primary tool for refining a task. These features can take different values in order to adjust the difficulty of tasks, to shift the task emphasis on different aspects of Focal KSAs, or to incorporate different Additional KSAs. In the online version of the *design pattern*, there are a couple ways to emphasize the relationship among particular Additional KSAs and Variable Task Features (see Section 5.0).

Looking at the design pattern in Table 3, note that there are five entries for Additional KSAs. The first is content knowledge, which is necessary for students to conduct experimental investigations (but is not a Focal KSA for this particular design pattern). Potential content areas include physics, biology, chemistry, psychology and other areas that might be addressed in the MCA-II. The presence of content knowledge as an Additional KSA emphasizes that the task designer must decide *what* content knowledge is involved and *how much* demand to place on it in the task. To assess inquiry capabilities, a task designer can choose to have low demand for content knowledge by using a familiar everyday context or a scientific context from earlier grade levels. Alternatively, if the objective is to jointly obtain evidence about content (as specified by a targeted benchmark), then the demand for content knowledge can be appropriately high, and both the inquiry process and content knowledge will be tested. In this approach, evidence about inquiry capabilities is conditional on content knowledge.

The second Additional KSA is prerequisite knowledge from earlier grades. In the case of the MCA-II, storyboards and items that are written by using this *design pattern* can require content-related KSAs that students should have gained in prior grades before they entered the grade for which a task is designed. For example, Grade 5 benchmarks in Minnesota Assessment Science Standards can be considered as Additional KSAs that are appropriate to include when task writers generate storyboards or items for students in middle or high school. The presumption is that this prerequisite knowledge is not a primary source of difficulty for students who actually are proficient in the Focal KSAs.

The third Additional KSA is prerequisite experience assessing or conducting component steps of an investigation. This has to do with the idea that any type of investigation (not just an experiment) has to proceed with a set of steps planned ahead of time and a method for systematically recording these steps and what transpires. Also implied is the notion that students would have some familiarity with using measurement and/or instrumentation as part of the process. In short, this Additional KSA makes clear that if students are going to be assessed for their ability to plan and conduct an experiment (the fifth Focal KSA), this demand builds on some experience in conducting investigations more generally. If they do not have this Additional KSA, they may perform badly for reasons that have little to do with their understanding of the concepts behind an experiment.

The fourth Additional KSA concerns the ability to collect, organize, analyze, and present data. Much like the prior Additional KSA, this speaks more broadly to a set of investigation-related skills including the use of measurement or data-collection methodology (e.g., instruments, and procedures related to experimentation in a particular domain) and even proficiency with quantitative tools and how to label and record the resulting data.

Familiarity with representational forms is the fifth Additional KSA. Here we are referring to the conventions of data display, such as tables, graphs, maps, and some domain specific illustrations (e.g., food webs). Again, this is an Additional KSAs that, if a student does not have it, can also affect the validity of the claims made for Focal KSAs. Taking the “photosynthesis investigation” as an example, the 7th scene requires students’ to interpret the results of an experiment (Focal KSA 11) but the item also necessarily involves familiarity with the presentation of data in a table form. Because this item was written to address an MCA benchmark addressing students’ proficiency with tables and graphs, the joint dependence on experimental investigation and representational skills is construct relevant.

The Experimental Investigation Design Pattern lists nine Variable Features. As mentioned earlier, these Variable Features are aspects of the task that can be manipulated to influence the level of demand from the KSAs, particularly the Additional KSAs. We briefly summarize each Variable Feature below:

- Content for the context of the investigation. As discussed in connection with Additional KSAs above, all experimental investigations involve some content. What content will be involved — everyday knowledge, content from earlier grades’ standards presumed to be familiar, or content that is also at issue at the grade level being assessed? Different content domains can be varied or combined as a context for tasks. In the MCA-II, content can be described in terms of the Minnesota Standards content classifications or test specification benchmarks.
- Which one (or more) of multiple phases of experimentation is addressed. A storyboard or task can focus on one phase, the transition between phases, or multiple phases of experimental investigation. Note that the “photosynthesis investigation” example touches on many phases – hypothesis generation, the conduct of the experiment, and the formulation of a conclusion from data. The designer needs to decide upon how many and which phases of an experiment will be addressed in a task.
- Qualitative or quantitative investigations. Qualitative investigation refers to recording information that is not immediately translatable into numerical data (e.g., the use of running records, interview responses, photos, videotaping). These data typically can be analyzed and then coded in quantitative terms that allow for comparison (say, between the experimental

and control group). Quantitative recording refers to using methods of counting or instrumentation that immediately yield numerical data (e.g., a ruler to measure distance, a scale to measure weight). The demands of either qualitative or quantitative methods can vary widely, depending on which particular approach is selected. These demands should be minimized if one only wants to draw inferences about KSAs directly linked to experimental investigation (and not methodology more broadly).

- Ease or difficulty with which treatment variable can be manipulated. A successful experiment depends on the systematic manipulation of the independent variable, and some manipulations are much easier (in terms of technical and time demands) than others. The manipulation in the "photosynthesis investigation," for example, is quite straightforward — the moving of a constant light source (a lamp) in 5cm increments from the plant. Other manipulations can be considerably more involved (e.g., exposing an organism to a substance that must be carefully made from different compounds, having to build a special piece of equipment to test the effects of magnetism on an object).
- Whether manipulated variables are given or to be determined. A storyboard or task can tell the student what the independent variable is or it can ask the student to make this determination on his or her own. In the "photosynthesis investigation," the entire experimental situation is given to the student – hence, it is understood that the distance of the light source is the independent variable. If the student had been presented with the much broader scientific problem (e.g., "determine why the photosynthesis rate in some elodea might vary") and then asked to generate a hypothesis along with specifying the variables, the task difficulty would be increased significantly. An on-demand accountability test such as the MCA-II usually uses a clearly specified problem with variables given. More open-ended problem definition is better suited to learning assessments in the classroom.
- The number of variables investigated and the complexity of their interrelationships. Complexity features in a situation include the number and variety of potential elements in a situation and the complexity of their interrelations. In thinking about the general problem involved in the example storyboard, one can imagine that in the real world, a number of factors might influence the photosynthesis rate of elodea – e.g., the size of the plant, the temperature of the water, the substances in the water, and the amount of light. To make the situation practical for the assessment context, the writers decided to make only one of these factors manipulatible (i.e. the amount of light) while presumably holding the other variables constant (i.e. by using only one plant in one water beaker). If the writers had allowed one more variable to be manipulated in the situation – say, the temperature of the water – this would have made the experiment much more involved (e.g., probably requiring at least two plants in separate beakers) with the possibility of variable interaction (e.g., the relationship of light distance might not have the same impact in the two different temperature conditions).

We need to note one more time that students in Minnesota middle school are not required to address experimentations with more than two independent variables, which is specified in the content limit of a benchmark (8.I.B.4) in MCA-II.

- The number of variables that need to be controlled. Experimental situations vary in the number of factors that may impinge on the causal relationship being studied. Indeed, one reason that experiments often are conducted in laboratory settings is because it is easier to control such variables in the indoors using specialized pieces of equipment. In the "photosynthesis investigation," students are given a relatively simple situation where one plant serves as its own control (i.e. the same plant is used for the different light distance manipulations) and students are asked to identify one thing that should be held constant (Scene 2). If one were to ask students to work on an experiment involving testing the impact of different fertilizers on tomato plants, students would need to think about many more variables that needed to be controlled (because they would be dealing with multiple plants and having to be sure that these plants were all alike and growing in nearly identical conditions save for the fertilizer they were receiving). The demands of the task increase according to the number of such control variables.
- Length of time over which the experiment is conducted. Another feature of an experimental task that directly affects level of complexity is the length of time one allots to measuring the experimental effect. In the "photosynthesis investigation," the effect of the distanced light source on the plant is assumed to be immediate (i.e., the number of bubbles coming off the plant will change right after the light is moved). For the fertilizer-tomato plant experiment, in contrast, it may take weeks to discern the impact of the treatment. This increase in time means not only that multiple measurements get taken, but it also introduces the possibility that something unexpected can happen in the time interval to alter the experimental situation (thus, requiring more front-end precaution).
- The representational forms being used for data. Task writers can vary this feature to test different aspects of scientific skills or incorporate targeted Additional KSAs. For example, students may be required to read a measurement instrument and know how to record raw data in a table with multiple columns and rows. If students are working on the conclusion formulation phase of an investigation, task writers may want to consider what form of data summary is displayed to show the pattern of experimental results more clearly and thus support generation of the following explanation. When students make an inference, they might need to consider the evidence across representational forms, as illustrated in Scene 6 of the "photosynthesis investigation." Data transformation can also be used in the experimental study if that is related to the intent of a task.

4.4 Narrative Structures

Narrative Structures for the MCA-II storyboards are actually a Variable Task Feature, but one of sufficient importance to merit their own attribute in the *design pattern*. As noted earlier, Narrative Structures are recurring structures for organizing the flow of information and items in the contextualized sets of items that constitute an MCA-II science task. The selected Narrative Structure for a storyboard can be used to help build and describe the narrative “arc” of the entire storyboard (i.e. what is presented and demanded from the first scene to the last). For tasks that address experimental investigation, three Narrative Structures were identified that lend themselves particularly well (see Fulkerson, Nichols, Haynie, & Mislevy, 2009):

- Investigation. A student or scientist completes an investigation in which one or more variables may be manipulated and data collected.
- Change Over Time. A sequence of events is presented to highlight sequential or cyclical change in a system.
- Cause and Effect. An event, phenomenon, or system is altered by internal or external factors.

The Narrative Structure of Investigation is a natural structure for storyboards and tasks required to assess students’ knowledge of and capabilities with experimental investigation. Change Over Time also is a plausible structure if the storyboard topic presents a context with a system undergoing change. Cause and Effect is appropriate as well because experimental investigation is a problem-solving approach for studying a causal relationship. In short, all of these Narrative Structures might help guide the construction of a storyboard that would include the presentation of at least some phase of experimental investigation.

In the “photosynthesis investigation” example, while all three Narrative Structures could be argued for, Investigation probably is the best fit. That is because the storyboard begins and ends with phases of an experimental investigation and the “arc” of the story is presented as the study of a hypothesis. Change Over Time is less compelling because the change involved (rate of photosynthesis depending on light source distance) is fairly immediate and not the accumulation of a sequence or cycle. Cause and Effect is an acceptable fit, but one can imagine that a storyboard writer following this Narrative Structure would have gone more in the direction of looking at alternative causes for an effect over the course of several scenes.

In Minnesota, looking at the standards and benchmarks is a good way to get some direction for selecting the Narrative Structure. Another finding from the alignment study is that the Narrative

Structure attribute of *design patterns* noted earlier is reflected in the test specifications document[5] derived from the Minnesota Standards, in the form of "content limits." The test specifications go beyond the statements of standards themselves, by further suggesting the kinds of reasoning and some of the features of tasks that are appropriate to include on the MCA-II to assess students at the given grade levels. Some Narrative Structures are implied in the content limits, and others are explicitly articulated.

4.5 Potential Work Products and Potential Observations

The attributes of Potential Work Products and Potential Observations in a *design pattern* concern how to capture students' thinking in terms of something they do or make and identify the information that constitutes evidence about targeted aspects of Focal KSAs. In the online version of the *design pattern*, there are links between particular Work Products and Observations because certain Work Products support identifying certain Observations. Their association can be displayed more clearly in the horizontal view of the *design pattern* (see Section 5.0 below).

In the MCA-II, computer-based tasks are generated in three formats: multiple choice, open-ended response, and figural response. This is the structure of the Work Product. What is listed below and in the *design pattern* are ideas for the semantic content of the Work Product that can be considered as evidence of students' reasoning skills in experimentation. The design pattern (see Table 3) lists eight entries under Potential Work Products, but this list could easily be expanded. In any case, most of the Work Products given may be implemented in more than one of the structural forms available to the MCA-II.

The first Work Product option given involves having the student select, identify, or evaluate a measurable research question. The next two Work Products have to do with the different kinds of variables in an experiment. The second entry asks students to Identify or differentiate independent and dependent variables, and the third asks students to identify or differentiate variables that do and do not need to be controlled. The item in Scene 2 from the "photosynthesis investigation" is an example of the third type of Work Product. The fourth Potential Work Product entry is to have a student complete some phases of experimentation with given information. For example, a student could be given an experimental situation and asked to identify the first steps necessary in the conduct of the experiment. The fifth and sixth entries presume that a student has been presented with a set of experimental results. For the fifth entry, the student is asked to

---

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

generate or identify a data pattern from the results, while the sixth entry involves generating an interpretation, explanation, or conclusion from the results (for an example, see Scene 7 from the "photosynthesis investigation"). The seventh entry suggests a work product where a student critiques his or her peers on their choice of an experimental procedure or explanation of experimental results. While this activity is a better fit for classroom instruction, an assessment could present a hypothetical situation to evoke a similar exercise. Last, the eighth Potential Work Product involves giving a student unexpected or confusing experimental results, and having them identify possible reasons for these results.

.

Potential Observations are what assessment designers hope to generate from students' Work Products that would constitute evidence of the Focal KSAs. They describe qualities, strengths, or the extent of work that tends to distinguish more or less capability on the whole or in selected aspects. The design pattern lists eleven Potential Observations for tasks addressing experimental investigations:

- Accuracy in identifying a situation suitable for experimental investigation
- Plausibility of a measurable research question being raised
- Plausibility of hypothesis as being testable by a simple experiment
- Plausibility/correctness of design for a simple experiment
- Correct identification of independent and dependent variables
- Accuracy in identifying variables that should be controlled or made equivalent
- Plausibility/correctness of steps to take in the conduct of an experiment
- Plausibility of plan for repeating an experiment
- Correctness of recognized data patterns from experimental data
- Plausibility/correctness of interpretation/explanation of experimental results
- Accuracy in critiquing the experimental design, methods, results, and conclusions of others

While the above list is by no means comprehensive, it covers considerable grounds in terms of the types of student behavior that — if found to be accurate, correct, or plausible — would indicate one or more of the abilities listed under the Focal KSAs. Note that these qualities can be sought in from a variety of item types and work products. For correctness of recognized patterns in experimental data, for example, a designer could write a multiple choice item and ask students to identify the correct interpretation, pose an open-ended question about the interpretation of a given data set, or look for evidence of students' understanding of patterns in data they collect themselves. This attribute focuses attention on student thinking reflected in performance rather than the surface form of the item.

5.0 Working with Design Patterns

As explained at the beginning of Section 4.0, moving from a hard copy version of a design pattern to one online makes possible additional features for the user. In addition to the hyperlinks for details and examples already described, the system for online design patterns has two notable features aimed at emphasizing the relationship among attributes and particular entries within these attributes. The first of these is an association feature, and the second is a horizontal view feature.

The association feature is illustrated in Figure 4. Working within the regular vertical view of the design pattern, this feature involves highlighting an icon next to one of the entries (in this case, the icon next to FK9) and then having other associated entries become highlighted (i.e. bolded). In the screenshot example, we see that highlighting the Focal KSA entry about recognizing that experimental steps must be repeatable results in a linkage to a Potential Observation about being able to plan for the repeating of an experiment. This emphasizes to the assessment writer that what an assessment targets can be linked directly to the type of observation from student work that can be considered as evidence about the particular Focal KSA. Many types of associations are possible across design pattern attributes, including links among Focal KSAs, Characteristic Features, Potential Observations, and Potential Work Products and links among Additional KSAs and Variable Features.

Figure 4: Example of Vertical View Associations Among Attributes of the Design Pattern for Experimental Investigation

Fk8. Ability to recognize variables that are inconsequential in the design of an experiment details

Fk9. **Ability to recognize that steps in an experiment must be repeatable to dependably predict future results**

Fk10. Ability to recognize that random assignment to treatment conditions (i.e. levels of the independent variable) is an important way to rule out alternative explanations for a causal relationship details

Fk11. Ability to interpret or appropriately generalize the results of a simple experiment or to formulate conclusions or create models from the results

Ak1. Content knowledge (may be construct relevant) details

Ak2. Prerequisite knowledge from earlier grades details

Ak3. Prerequisite experience assessing or conducting component steps of an investigation details

Ak4. Ability to collect, organize, analyze, and present data details

Ak5. Familiarity with representational forms (e.g., graphs, maps) details

Po1. Accuracy in identifying situation suitable for experimental investigation

Po2. Plausibility of a measurable research question being raised

Po3. Plausability of hypothesis as being testable by a simple experiment

Po4. Plausability/correctness of design for a simple experiment

Po5. Correct identification of independent and dependent variables

Po6. Accuracy in identifying variables (other than the treatment variables of interest) that should be controlled (held constant) or made equivalent (e.g., through random assignment).

Po7. Plausibility/correctness of steps to take in the conduct of an experiment

Po8. **Plausability of plan for repeating an experiment**

Po9. Correctness of recognized data patterns from experimental data

The nature of many of the attribute associations can be displayed even more clearly with the alternative horizontal view option. Available in the online version, this feature can be accessed by clicking “View Horiz” in the upper-right hand corner of the standard *design pattern* page. By doing so, one can select which sets of attribute entries to view and have them displayed together side-by-side in of columns where linkages across particular instances of the attributes are made clear. Figure 5 shows an example of such a horizontal view, in this case arraying the side-by-side views of Focal KSAs, Potential Observations, and Potential Work Products.

Figure 5: Horizontal View Segment from Associations among Three Attributes of the Design Pattern for Experimental Investigation

Anchor: Focal knowledge, skills, and abilities	Associated: Potential observations	Associated: Potential work products
Ability to distinguish between experimental and observational methodology		
Ability to recognize that when a situation of scientific interest includes aspects that can be altered or manipulated practically, it is suitable for experimental investigation details	Accuracy in identifying situation suitable for experimental investigation	
Ability to recognize that the purpose of an experiment is to test a prediction/hypothesis about a causal relationship details	Accuracy in identifying situation suitable for experimental investigation Plausability of hypothesis as being testable by a simple experiment	Select, identify, or evaluate an investigable question details
Ability to identify, generate, or evaluate a prediction/hypothesis that is testable with a simple experiment	Plausability of hypothesis as being testable by a simple experiment	Select, identify, or evaluate an investigable question details
Ability to plan and conduct a simple experiment step-by-step given a prediction or hypothesis	Plausability/correctness of steps to take in the conduct of an experiment Plausability of plan for repeating an experiment Accuracy in critiquing the experimental design, methods, results, and conclusions of others	Complete some phases of experimentation with given information, such as selection levels or determining steps Critiques of peers on their choice of experimental procedures or explanations of experimental results details
Ability to recognize that at a basic level, an experiment involves manipulating one variable and measuring the effect on (or value [illegible]	Plausability/correctness of design for a simple experiment [illegible]	Identify or differentiate independent and dependent variables in a given scientific situation [illegible]

6.0 *Working with Design Patterns*

For task writers, the most challenging part of test design is working from a statement of standard or benchmarks to the production of a specific task that evokes a student's demonstration of his or her proficiencies with regard to those standards or benchmarks. It is at just this step that the talents of gifted and experienced test developers come into play while novice test developers have the hardest time. This is particularly true when the intended tasks are innovative as to format or use of technology, or address traditionally hard-to-assess proficiencies such as inquiry and model-based reasoning in science.

The Experimental Investigation Design Pattern can help fill this gap and make a seemingly mysterious process explicit. It can ground an evidentiary assessment argument to clarify why or why not task situations should evoke the proficiencies of interest, options for fine-tuning these task situations, and what to look for in student performances and how to evaluate them. Due to its versatility, this *design pattern* can greatly expand the design space that task writers can work from. By using an exemplar task, the "photosynthesis investigation," we found that a pre-existing storyboard can be analyzed in terms of already reflecting a number of the attribute entries from the experimental investigation design pattern. While it is clear that no one storyboard could involve all of the content within a design pattern as developed as the one for experimental investigation, what the design pattern does make clear is the design choices that a task writer has going forward. For example, this design pattern could prompt a writer to design other storyboards that emphasize different phases or aspects of the experimental investigation (e.g., ones revolving around the Focal KSAs of being able to generate or evaluate a hypothesis or plan for the replication of an experiment). In this sense, this *design pattern* can prompt task writers to consider test design from the wider perspective of a more complete and unified inquiry-based assessment argument.

This Experimental Investigation Design Pattern is a parallel piece to another *design pattern* for observational investigation. Because experimental and observational investigation are indispensable and complementary scientific methods, we suggest our users also refer to the Observational Investigation Design Pattern (Mislevy et al., 2009) to better understand the difference and similarities between these two types of inquiry skills and then use them appropriately.

References

Almond, R. G., Steinberg, L. S., & Mislevy, R.J. (2002) Enhancing the design and delivery of assessment systems: A four-process architecture. *Journal of Technology, Learning, and Assessment*, *1:*(online).

American Association for the Advancement of Science (1993). Project 2061: Benchmarks for science literacy. New York: Oxford University Press.

Bloom, B. S., Hastings, J. T., & Madaus, G. (Eds.). (1971) *Taxonomy of educational objectives, handbook I: Cognitive domain.* New York: McKay

Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design and analysis issues for field settings.* Chicago: Rand McNally.

Cox, D. R. (1958). Planning of experiments. London: Wiley.

Fleck, L. (1935) *The genesis and development of a scientific fact* (trans. Thaddeus Trenn, 1979). Chicago: University of Chicago Press.

Fulkerson, D., Nichols, P., Haynie, K., Mislevy, R. (2009). *Narrative Structures in the Development of Scenario-Based Science Assessments (Large-Scale Assessment Technical Report 3).* Menlo Park, CA: SRI International.

Hammrich, P.L. (1997, May/June). Teaching for excellence in K-8 science education: Using Project 2061 benchmarks for more effective science instruction. *Journal of Teacher Education, 48*, 222-232.

Mayo, D.G. (1996). Error and the growth of experimental knowledge. Chicago: University of Chicago Press.

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

Mislevy, R. J., Almond, R. G., & Lukas, J. F. (2004). A brief introduction to Evidence-Centered Design (*CSE Technical Report 722)*. Los Angeles: The National Center for Research on Evaluation, Standards, Student Testing (CRESST), Center for Studies in Education, UCLA.

Mislevy, R. J., & Haertel, G. D. (2006). Implications of evidence-centered design for educational testing. *Educational Measurement: Issues and Practice*, *25*(4), 6–20.

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.

Mislevy, R. J., & Riconscente, M. M. (2006). Evidence-centered assessment design: Layers, concepts, and terminology. In S. Downing & T. Haladyna (Eds.), *Handbook of Test Development* (pp. 61-90). Mahwah, NJ: Erlbaum.

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

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

National Research Council (2000). Inquiry and the National Science Education Standards. National Committee on Science Education Standards and Assessment, Washington, DC: National Academy Press.

Toulmin, S. E. (1958). *The uses of argument.* Cambridge: Cambridge University Press.

Wagner, P. (1983). The nature of paradigmatic shifts and the goals of science education. *Science Education, 67* (5), 910-924.

!

Sponsor
The National Science Foundation, Grant DRL - 0733172

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

Subgrantees
University of Maryland
Pearson