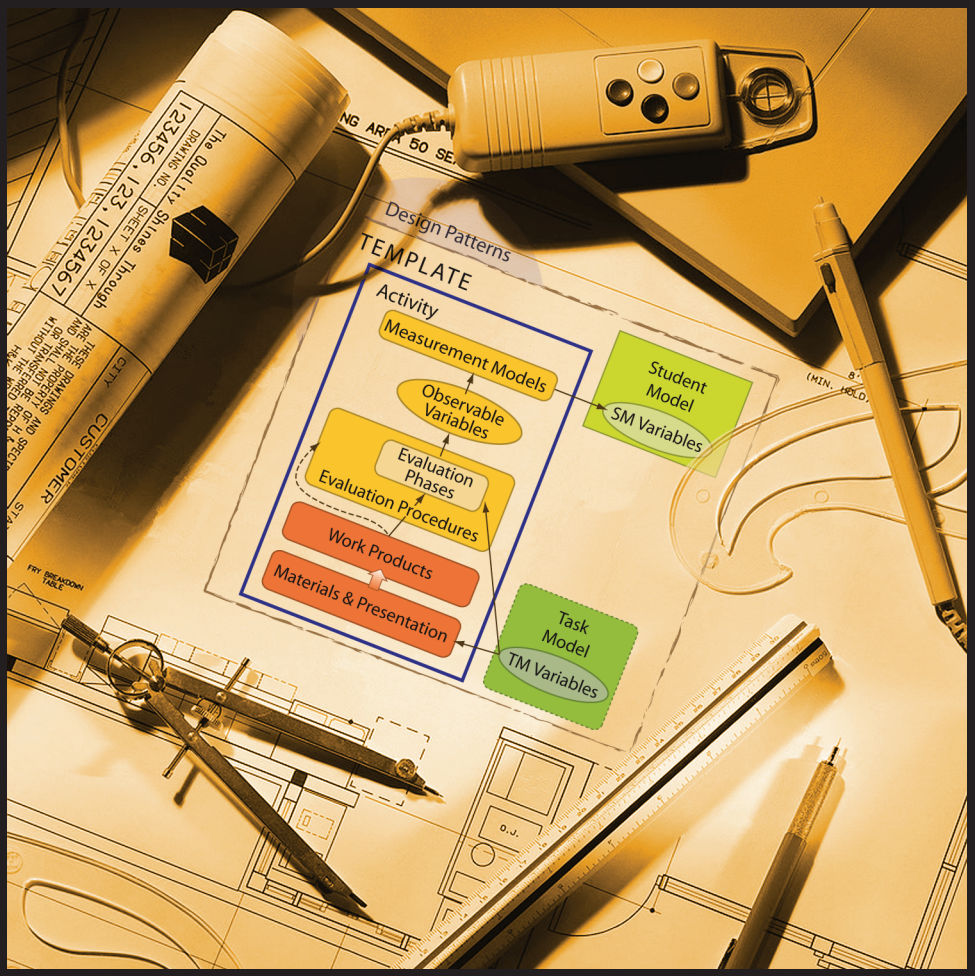


PADI Technical Report 20 | March 2009

Design Pattern for Assessing Cause and Effect Reasoning in Reading Comprehension

PADI | Principled Assessment Designs for Inquiry

Robert J. Mislevy, University of Maryland
Taslima Rahman, University of Maryland

Report Series Published by SRI International

SRI International
Center for Technology in Learning
333 Ravenswood Avenue
Menlo Park, CA 94025-3493
650.859.2000
http://padi.sri.com

PADI Technical Report Series Editors
Alexis Mitman Colker, Ph.D., *Project Consultant*
Geneva D. Haertel, Ph.D., *Co-Principal Investigator*
Robert Mislevy, Ph.D., *Co-Principal Investigator*
Lynne Peck Theis and Ron Fried, *Documentation Designers*

PRINCIPLED ASSESSMENT DESIGNS FOR INQUIRY
TECHNICAL REPORT 20

A Design Pattern for Assessing Cause and Effect Reasoning in Reading Comprehension

Prepared by:
Robert Mislevy, University of Maryland
Taslima Rahman, University of Maryland

Acknowledgments
This material is based on work supported by the National Science Foundation under grant REC-0129331 (PADI Implementation Grant).

CONTENTS

FIGURES

TABLES

ABSTRACT

This paper addresses reading assessment task design under a constructive, model-based, view of reading comprehension. We draw on Kintsch's (1998) Construction-Integration (CI) theory to ground the discussion psychologically, and design concepts from evidence-centered assessment design (ECD) to provide support for task authors. Specifically, the paper presents a *design pattern* to help test developers craft reading comprehension tasks that involve Cause and Effect (C&E) reasoning. The work is offered as a joint effort to (1) consider assessment implications of research on Cause and Effect reasoning in reading comprehension, and (2) provide a grounded example of the use of the ECD framework in an educationally important area of assessment. The resulting C&E *design pattern* addresses key considerations for stimulus materials, prompts, potential work products, and ways to evaluate them.

1.0 Introduction and Overview

"It is clearly false to assume that comprehension is an ability that can be measured once and for all, if only we had the right test. Instead, 'comprehension' is a commonsense term for a whole bundle of psychological processes…"

(Kintsch & Yarbrough, 1982, p. 834).

Despite research into the nature of reading that led to Kintsch and Yarbrough's statement decades ago, current assessment of K-12 students' reading proficiency in large-scale testing continues to develop under the assumption that comprehension is a unitary skill that comprises decoding, word recognition, and answering questions given a text passage to read. While few would deny the importance of these skills in constructing meaning from text/print materials, this skill-based approach does not, in itself, provide a principled basis for constructing tasks or evaluating students' proficiencies in reading. Reading assessment tasks typically have students read passages and answer questions about what was read. Achievement is determined based on the number of correct responses students make. Such assessments provide at best only global indications of proficiency, and offer no means to gather evidence on different aspects or balances of knowledge and skills, as might be desired for evaluating instructional methods or providing feedback for individual students.

Kintsch and Yarbrough (1982) recommended the development of "a collection of different tests, each tuned to some specific aspects of the total process," and noted that "to construct such a collection will require the guidance of a fairly sophisticated theory of prose comprehension" (p. 834). One line of research since that time has led to the development of specific tests to diagnose and remediate reading difficulties in the word recognition and decoding skills emphasized in primary-level reading instruction (Rayner, et al., 2001). However, decoding does not guarantee comprehension (Massey & Heafner, 2004), and the importance of word recognition abilities diminishes with fluent reading (Jackson & McClelland, 1979). Many middle and high school students read texts fluently but fail to understand what they read. Hannon and Daneman (2001) went further by developing tests that provided evidence about text memory, text inferencing, knowledge integration, and knowledge access.

Although separate tests that focus on single process hold value for diagnosing students' deficiencies and assessing specific learning outcomes, real-world tasks that involve reading generally require assemblies of processes jointly. Therefore, we also would want to be able to assess students' capabilities in performing complex tasks. We seek a conceptual framework for constructing tasks to assess aspects of the knowledge and processes involved in reading comprehension in purposeful combinations. Such a framework could support the development of highly focused tests, but it could also be used to design tasks that require targeted combinations of knowledge and processes. This broader goal requires guidance from a fairly sophisticated theory of prose comprehension.

Contemporary theories of reading comprehension are constructive and model-based (Graesser, Gernsbacher, & Goldman 2003). Comprehension of text passages is seen to involve the cognitive processes by which a reader integrates information in a text—including not only the content of its propositions but its use

of genre, cohesion devices, and rhetorical structures—with the knowledge of these patterns and the subject matter that the reader brings to the encounter, in order to construct an understanding. "Both the basic comprehension of literal text meanings and the use of knowledge necessary to go beyond the literal (propositional meaning) are accounted for. … In fact, text research has increasingly focused on the fact that a reader may understand several levels of text information, including information about text genre and communication contexts, as well as the text itself and the referential situation" (National Research Council, 1998, p.64).

This presentation addresses task design under this constructive, model-based view of comprehension, applying the ideas specifically to assessment of Cause and Effect (C&E) reasoning in reading comprehension. Throughout the later sections in this report, we examine different types of reading comprehension tasks to aid the discussion. Our emphasis is on ways to provide guidance for test developers. We draw on Kintsch's (1998) Construction-Integration (CI) theory to ground the discussion psychologically. To ground the discussion with respect to task design, we use concepts from evidence-centered assessment design (ECD) (Mislevy, Steinberg, & Almond, 2002, 2003, Mislevy & Haertel, 2006, Mislevy & Riconscente, 2006). ECD provides principles, patterns, and examples to guide the work of task developers from theoretical grounding to the practical elements of operational tasks.

The ECD design layer called *Domain Modeling* concerns how information about the domain can be organized in terms of the elements of an assessment argument, at a narrative rather than technical level. It provides conceptual guidance, but does not yet specify mechanical details such as psychometric models, particular stimulus materials, or final scoring methods. A representational form called an *assessment design pattern* (Mislevy et al., 2003) helps test developers sketch out design spaces for tasks that address targeted aspects of proficiency.

The present report, then, presents a *design pattern* to help test developers design reading comprehension tasks that involve Cause and Effect (C&E) reasoning in prose contexts. We take this pass at working through assessment implications of research on C&E reasoning in reading comprehension, and in doing so, provide a grounded example of the use of the ECD framework in an educationally important area of assessment. The resulting C&E *design pattern* addresses key considerations for stimulus materials, prompts, potential work products, and ways to evaluate them.

Section 2 of this report sketches the key findings from research on reading comprehension—and C&E reasoning in particular—that the *design pattern* draws on. It highlights Kintsch's (1988, 1998) Construction-Integration (CI) theory of comprehension, Lakoff's (1987) prototype structure for causation, and the cause-and-effect rhetorical frame. Section 3 reviews relevant aspects of the ECD framework, and introduces the forms of assessment arguments and *design patterns*. Section 4 presents the *design pattern* and illustrates attributes of the *design pattern* relevant to assess C&E reasoning with examples from different grade levels, topic areas, and types of assessments. Section 5 concludes by noting how *design patterns* for other aspects of reading comprehension could be developed and how they could be used in developing reading comprehension tasks.

2.0 Background

Research on the nature of reading comprehension traces back to Bartlett's (1932) early investigations of the role of schemas in understanding; that is, the role of a reader's prior knowledge about the elements, relationships, and processes that typify situations like the one being presented in a text. Such structures have been studied under the terminology of "frames" (Minsky, 1975; Fillmore, 1976), "schemas" (Rumelhart, 1975, 1980), "scripts" (Schank & Ableson, 1977), "schematic superstructures" (van Dijk & Kintsch, 1983), and "idealized cognitive models" (Lakoff, 1987). At many levels and in many aspects, information from such structures is activated from long-term memory and synthesized with information from the text in working memory. Comprehension can be thought of as the resulting mental model of the situation, from which further reasoning or action proceeds.

More recent work has shed light on how these processes occur. Kintsch's (1988, 1998) Construction-Integration (CI) theory, building on earlier work with van Dijk (Kintsch & van Dijk, 1978, van Dijk & Kintsch, 1983) is probably the best known and most influential contemporary framework for comprehension. Section 2.1 summarizes key points of CI theory as they relate to the purposes of this presentation. Section 2.2 reviews key elements of Cause & Effect (C&E) reasoning, and Section 2.3 discusses rhetorical structures that signal C&E relationships in textual material. Taken together, this brief survey serves as grounding for a *design pattern* for assessing C&E reasoning as a component of reading comprehension. Section 2.4 outlines some implications for assessment that will form the basis of the *design pattern*.

2.1 The Construction-Integration Theory of Comprehension

Kintsch's focus has been reading comprehension, although he argues that the essential nature and many of the processes involved in reading characterize comprehension more generally. Comprehension, according to Kintsch (1998), occurs when and if text elements that enter into a reader's mental process achieve a stable state in which the majority of the elements are meaningfully related to one another. The elements that enter into comprehension process include perceptions, concepts, ideas, images, and emotions. We form connections among things that were previously disparate, namely, the ideas expressed in the text and relevant prior knowledge. The CI theory of comprehension posits two phases (with repeated cycling as a reader continues through a text). The construction (C) phase is initiated by features of stimuli in the environment and activates associations from long-term memory—whether they are relevant to the current circumstances or not. The associations include the patterns that constitute schemas, with the probability of activation depending in part on the strength of similarity of stimulus features and aspects of the elements of the schema. In the integration (I) phase, only the aspects of activated knowledge—both from contextual input and long-term memory—that are mutually associated are carried forward. The result, the *situation model*, is the reader's understanding of the text. The notion of schemas is important in CI theory as a metaphor for patterns of interrelationships provoked and synthesized through associationist processes rather than more rigid structures and a top-down processing model that characterized work on schemas in the 1970's.

In this presentation we take CI theory as described in Kintsch (1998) as representative of a number of related theories of comprehension. Among those

noted in Grabe's (1999) review of this area for language testers are Rayner and Pollatsek (1989), Just and Carpenter (1987), and Bruer (1993). A number of features of these models are particularly relevant to our purposes. Van Dijk and Kintsch (1983) distinguish three levels involved in text comprehension, namely the *surface structure* of a text, the *text model*, and the *situation model*. The *surface structure* of a text concerns the specific words, sentences, paragraphs, and so on that constitute the text—that is, the particular features of the stimulus. The *text model* is the collection of interconnected propositions that the surface structures convey, and corresponds roughly to what might be called the literal meaning of a text. The text model would be the essentially the same for all readers with sufficient knowledge of the language. The *situation model* is a synthesized understanding that integrates the text model with the knowledge the reader brings to the encounter (also shaped by goals, affect, context, etc.), and constitutes that reader's comprehension of the text. Readers with different knowledge, affect, or purposes would produce situation models that differ to varying degrees.

Example Task 1 (see Figure 1) presents a reading comprehension task in the specialized area of medicine. A physician and a non-physician reading the paragraph in this task would produce rather different situation models. Strictly speaking, there is sufficient information in terms of formal content of the propositions to answer the Items 1 and 2 on a superficial level and Items 3 and 4 correctly. However, ideal answers to Items 1 and 2 are made deeper and richer by building around mechanisms and relationships not given in the paragraph content. Correct answers to Items 3 and 4 can give a misleading impression of the nature of the respondent's understanding.

Schemas of many kinds and at different levels are involved in the comprehension process. Some schemas involve the subject matter of the text, and, like the medical example in Example Task 1, are essential to what might ordinarily be called comprehension. Some concern cultural models (Strauss & Quinn, 1998), such as what it means to be married, how people are motivated by jealousy or ambition, or, of particular relevance to our purposes, prototypical situations involving causation. Other schemas concern the structure of the text. Writers use genres, grammatical forms, and marker words and phrases to convey relationships and intentions to the reader, and a reader who is familiar with these devices has more resources for building a situation model that resonates with the writer's intensions than a reader who is not. Still other schemas involve elements of the context of the interaction and the reader's purpose. Thus, we build a different situation model from a bus schedule if we actually need to catch a bus than if we just have to answer a test item about it.

From the perspective of CI theory, a person constructs situation models, or mental representations, of much the same kinds, in much the same ways, from different kinds of stimuli or from mixtures of stimuli. Different perceptual systems are employed which initiate the processes, and schemas associated with those systems are brought into play. When we watch a movie, the conventions and grammar of filmography help us establish chains of events, characters' inner states, and causal relationships. In reading comprehension, writers similarly use genres, rhetorical frames, and text markers that help a reader activate schemas that interact with content knowledge. Thus, reading comprehension tasks that address cause-and-effect structures involve both C&E relationships inherent to the subject matter being discussed and the surface structures of the text that are used to convey the information. The following sections address these topics in turn.

Figure 1. Example Task 1: Specialized Medical Knowledge (Brattström & Wilcken, 2000, p. 315)

Both markedly and mildly elevated circulating homocysteine concentrations are associated with increased risk of vascular occlusion. Here we review possible mechanisms that mediate these effects. Inborn errors of homocysteine metabolism result in markedly elevated plasma homocysteine (200–300 μmol/L) and thromboembolic (mainly venous) disease: treatment to lower but not to normalize these concentrations prevents vascular events. Mild homocysteine elevation (>15 μmol/L) occurs in ≈20–30% of patients with atherosclerotic disease. Usually, this is easily normalized with oral folate and ongoing trials are assessing the effect of folate treatment on outcomes. Although there is evidence of endothelial dysfunction with both markedly and mildly elevated homocysteine concentrations, the elevated homocysteine concentration in atherosclerotic patients is also associated with most standard vascular risk factors, and importantly, with early decline in renal function, which is common in atherosclerosis. Decline in renal function alone causes elevated plasma homocysteine (and cysteine). These observations suggest that mild hyperhomocysteinemia could often be an effect rather than a cause of atherosclerotic disease. Data on the common C677T methylenetetrahydrofolate reductase polymorphism supports this, in that, although homozygosity is a frequent cause of mild hyperhomocysteinemia when plasma folate is below median population concentrations, it appears not to increase cardiovascular risk. Indeed, there is recent evidence suggesting an acute antioxidant effect of folic acid independent of its effect on homocysteine concentrations. This antioxidant mechanism may oppose an oxidant effect of homocysteine and be relevant to treatment of patients with vascular disease, especially those with chronic renal insufficiency. Such patients have moderately elevated plasma homocysteine and greatly increased cardiovascular risk that is largely unexplained.

1) Explain the cause & effect relationship between increased homocysteine levels and atherosclerosis.

This is an open-ended item. The following answer would receive full credit on a scale of 0-4: Increased homocystein levels are the effect of atherosclerosis. This is evidenced by the following: Increased levels of homocystein levels are associated with most standard vascular risk factors including a decline in renal function. Early decline in renal function is common in atherosclerotic patients, and can also be directly attributed to elevated homocysteine levels. Hence, it can be concluded that atherosclerosis leads to increased homocysteine levels.

2) What conclusions can be drawn about the effects of oral folate treatment in patients with Hyper homocysteinemia?

This is an open-ended item. The following answer would receive full credit on a scale of 0-3: Oral folate not only normalizes the homocysteine levels and the renal functions, it also does not increase the cardiovascular risk. In addition, it has a beneficial antioxidant effect and actually has a positive effect in the treatement of patients with cardiovascular disease.

3) The effects of oral folate treatment include:

a. It helps elevate homocysteine levels and the renal function
b. It helps regulate the renal functions
c. It helps lessen the cardiovascular risk
d. Both b and c
e. Both a and c

4) Indicate whether each statement is True or False:

Antheroseclerosis is caused by Hyperhomocysteinemia _____
Antheroseclerosis is an effect of Homocysteine _____
Antheroseclerosis is the symptom of vascular and renal dysfunction _____

NOTE: The above text is an excerpt from the abstract of Brattström and Wilcken (2000).

2.2 Cause and Effect Reasoning

Cause-and-effect reasoning is central to human reasoning in everyday life as well as in the disciplines. A dictionary definition is straightforward: One event, the cause, brings about, through some mechanism, another event, the effect. This relationship is illustrated in the top frame of Figure 2: "Max hits the window" is one event, which causes a second event, the "Window breaks."

Writing from a psycholinguistic perspective, Lakoff (1987, pp. 54ff) proposes that reasoning about causation extends from a direct-manipulation prototype that is basic to human experience. He characterizes an idealized cognitive model for causation in terms of the following cluster of interactional properties:

1. *There is an agent that does something.*
2. *There is a patient that undergoes a change to a new state.*
3. *Properties 1 and 2 constitute a single event; they overlap in time and space; the agent comes in contact with the patient.*
4. *Part of what the agent does (either the motion or the exercise of will) precedes the change in the patient.*
5. *The agent is the energy source; the patient is the energy goal; there is a transfer of energy from the agent to patient.*
6. *There is a single definite agent and a single definite patient.*
7. *The agent is human.*
8. *a. The agent wills the action. b. The agent is in control of his action. c. The agent bears primary responsibility for both his actions and the change.*
9. *The agent uses his hands, body, or some instrument.*
10. *The agent is looking at the patient, the change in the patient is perceptible, and the agent perceives the change.'* (pp. 54-55).

Lakoff claims that the most representative examples of causation have all of these properties (e.g., Max broke the window). Less prototypical instances that we still consider causation lack some of the properties: indirect causation lacks Property 3, and billiard-ball interactions that characterize much reasoning in the physical sciences just have properties 1-6. In case "a)" in Example Task 2 (see Figure 3), there is one agent, one patient, and a single event. This is a prototypical event. Case "b)" mentions two events in which people build dams in the Everglades and the wildlife is devastated. Property 3 is lacking because the causal relationship is indirect; the intervening events are that damming reduces water flow, reduced water flow causes decline in sources and living areas for wildlife, and less food and living areas causes the wildlife to decline. All of the events after the first do not have a human agent, so Properties 8, 9, and 10 are lacking as well. In Example Task 3 (see Figure 4), the agent 'student' adds vinegar to the baking soda in the patient 'volcano,' and an eruption results. This example of causation maintains Lakoff's interactional properties at the surface level, although a more sophisticated reader's situation model would link to an explanation for the chemical reaction between baking soda and vinegar.

Figure 2. Types of Cause and Effect Relationships

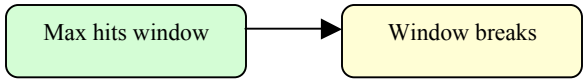
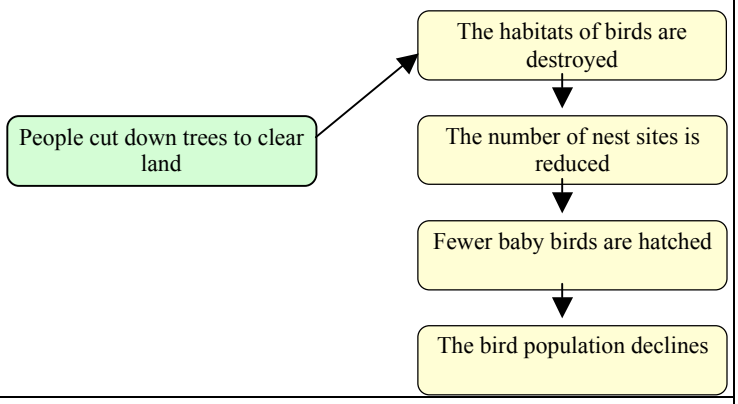
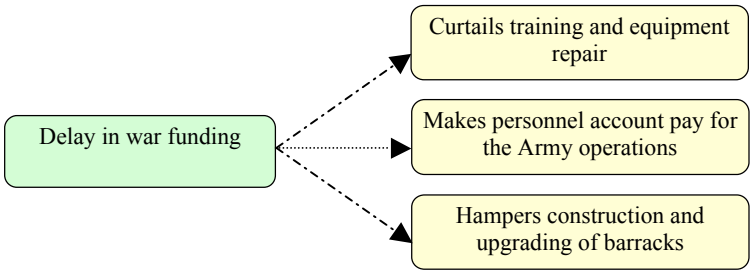
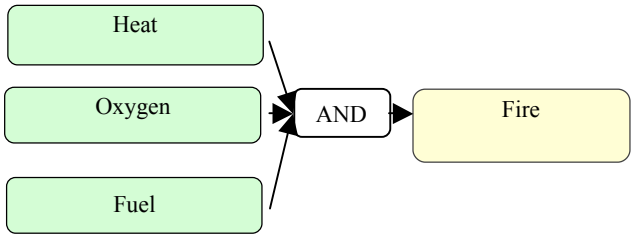
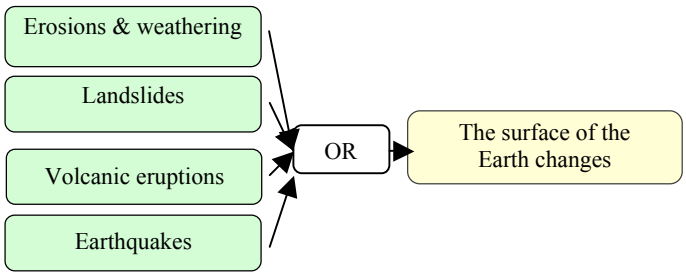
Type 1:One to one connection Max hit the window with the bat, and it broke.	
Type 2: Chained cause and effects '*When people cut down trees to clear land, they destroy the habitats of birds. This reduces the number of nest sites. As a result, fewer baby birds are hatched, and the bird population declines.*'	
Type 3: Multiple effects *A delay in billions of dollars of supplemental war funding for the Pentagon would cause the Army to curtail training and equipment repair necessary to prepare units in the United States for deployment. If the new funds did not arrive in time, the Army would prioritize spending in Iraq, Afghanistan and elsewhere and cut back in other areas, officials said. Next, the Army would request the authority to shift money from its personnel accounts to pay for operations. The construction and upgrading of barracks, fitness centers and other facilities would also be affected, Pentagon spokesman Bryan Whitman said, (Tyson, 2007).*	
Type 4: Multiple (catenated) causes *Combustion requires heat, oxygen, and a fuel source. To put out a fire ,it is sufficient to remove any one of the three.*	
Type 5: Multiple (alternative) causes *'The surface of the Earth changes. Some changes are due to slow destructive processes such as erosion and weathering. Some changes are due to rapid constructive and destructive processes such as landslides, volcanic eruptions, and earthquakes. These changes all affect topography of the Earth's surface.'*	

Figure 3: Example Task 2: Use of Cause & Effect Signal Words

Combine each of the following pairs of sentences into one using appropriate cause or effect transitions (in bold) from the list at the bottom. Write down your selection of the transition word next to the pair of sentences.

a) Max pushed the heavy box. The box slid along the floor. _____

b) The Everglades was dammed and drained. Wildlife was devastated. _____

c) The sun never drops below the horizon on first day of summer in North Pole. North Pole has 24 hours of daylight on the first day of the summer. _____

d) Wetland areas soak up rainwater like a sponge. They provide water storage and flood protection. _____

Due to Therefore So Consequently Because Because of Since

NOTE: Examples of C&E signal words use are adapted from the Pearson website *Super Read! Strategies for Effective Reading in Biology*. Downloaded January 19, 2008 from http://www.phschool.com/science/biosurf/superread/unit9/9strategy1.html

Figure 4: Example Task 3: Cause & Effect with a Hands-On Task

A model volcano

Make a model volcano by molding the sides from play dough. Put a small pile of baking soda in the middle, with a few drops of red food coloring. Then drip in some vinegar.

What happens? Why?

NOTE: This task is excerpted from http://cornerstonevalues.org/cause.html [Last accessed on January 19, 2008).

The basic C&E schema can be extended in various ways as depicted in Figure 2. These structures are less basic than Lakoff's (1987) cognitive prototype, but extend it in ways needed to support increasingly sophisticated C&E reasoning. In each instance, a diagram of the structure and an example of text representing that structure is given.

- An effect can also cause something else to take place; i.e., causal chains (Type 2 in Figure 2).
- A cause can result in multiple effects (Type 3 in Figure 2).
- An effect can require multiple causes, which must all be present for the effect to occur (Type 4 in Figure 2). This is an "AND" relationship between multiple causes and an effect.
- An effect can be entailed by different causes (Type 5 in Figure 2). This is an "OR" relationship between multiple causes and an effect.

Complex situations can require several of these relationships in combination. To anticipate the discussion of task design, Lakoff's research might lead us to expect that all other things being equal, C&E situations closer to the prototype might be easier to reason through than those that depart from it. Over-riding conditions would be familiarity with similar situations and the nature of the inference that is to be drawn.

For centuries, the fundamental nature of causation and the challenge of inferring cause from observations have occupied philosophers as eminent as David Hume and statisticians of the stature of R.A. Fisher. Cause and effect stands at the center of scientific reasoning and legal reasoning, troubleshooting, and human relations, each with its own explanatory mechanisms by which causes entail effects. To a large extent, formal education is learning about causal mechanisms in the various domains and using them to solve problems (reasoning from effects to causes) or produce desired outcomes (reasoning from causes to effects).

Causal relationships also play a prominent role in literature. Human actions narrated in stories, novels, and plays involve relations akin to physical causality. Causal models used in narratives can be "…different from the unambiguous, contradiction free system of science" (van Dijk & Kintsch, 1983, p. 47). When narrating, a text typically leaves some crucial causal relationship implicit, and readers have to supply the missing link from their own knowledge.

Knowledge about causal relations is often crucial for interpreting a text (Norman & Rumelhart, 1975; Schank & Abelson, 1977; Warren, Nicholas, & Trabasso, 1979). However, "as it turns out, people are often not good at this task, and arrive at misinterpretations that grossly distort the actual causal relations in the systems" (van Dijk & Kintsch, 1983, p.46; also see Graesser, 1981).

The basic cause-and-effect structure and its elaborations can be viewed as schemas, which, among others, can be activated in the CI process for building situation models. That is, among the elements of knowledge activated from long-term memory in the comprehension process, schemas developed from previous experience suggest connections among elements appearing in textual material, in addition to associations stimulated by specific elements of the text. General C&E schemas, if activated by a reader, influence the contents of the resulting situation model by adding propositions that relate elements from the text in ways that explain their connections and support further reasoning forwards or backwards in time. Specific knowledge of the real-world domain and the relationships underlying the particular context, again if activated by a reader, further enrich the situation model.

2.3 Cause and Effect Rhetorical Forms

Sentences are surface structures that convey predicate-argument schemas or propositions. Texts consist primarily of ordered collections of sentences in which propositions are related in various ways, including cause-and-effect relationships. Relationships within propositions or those connecting local propositions that are being processed together constitute microstructures in text, whereas larger organizing schemas are macrostructures. Van Dijk and Kintsch (1983) defined rhetorical forms in terms of strategies that experienced writers use to organize text when they write and experienced readers use when they read, in order to communicate certain kinds of information or intentions. Examples of rhetorical forms include argument, definition, compare and contrast, procedural description,

and—central to our purposes—cause and effect relationships. Rhetorical forms provide a means for forming both the microstructures and macrostructure of the text, such as its gist, overall organization, and main points.

For example, Kintsch and Yarbrough (1982) showed that students were better able to answer topic and main-idea questions for texts that were clearly organized according to a familiar rhetorical structure than for texts with identical propositional content but without such an organization. “Rhetorical cues and canonical ordering that distinguished the good forms of the text from bad forms facilitate the macroprocesses in comprehension, presumably because they permitted the successful use of rhetorical comprehension strategies” (Kintsch & Yarborough, 1982, p.833). Various cues in the “good” version of the texts triggered appropriate reading comprehension schemas, and the subjects used these schemas to organize their understanding of the texts. That is, the rhetorical frames may themselves be considered organizational schemas, which in conjunction with schemas associated with content, are integrally involved in the construction of an appropriate situation model.

With regard to the cause and effect rhetorical frame specifically, the relationship between causally connected propositions in a text may be explicit or implicit. In an explicit organization, adverbials such as *therefore*, *so*, and *as a result* are used to signal cause-effect relationships. In an implicit organization, propositions are presented in a linear order that can suggest but does not necessitate a causal relationship. In this case, knowledge about the content is required to build an appropriate situation model. Consider the following example:

> ***The snow was deep on the mountain. The skiers were lost, so they dug a snow cave, which provided them shelter.***

The prepositional representation would be as follows:

> ***Deep snow _ Lost skiers _ Skiers dig snow cave _ Cave provides shelter.***

The implication that the skiers were lost because of the snow is implicit. The implications that their being lost led them to build a cave and that the cave provided shelter are explicit.

An important goal of reading instruction is to attune students to the existence and use of cause-and-effect structures, the cause-and-effect rhetorical schema, and the techniques that writers use to signal its appropriateness. All contribute to reading comprehension. Becoming aware of cause-and-effect reasoning situations encountered in everyday life and their use across varied content domains makes students more likely to instantiate the cause and effect frame when they read. Knowing that writers use the cause-and-effect rhetorical frame in texts and studying varied examples of its use, both with and without markers, similarly increases the chances students will activate it when it is called for. Becoming familiar with text markers and the particular relationships they signal additionally increases the chances that students will activate cause and effect frames when they read and use it to construct appropriate meanings from texts.

2.4 Some Implications for Assessment

Formal education addresses cause and effect reasoning in two senses. The first, in subject domains, are the models and mechanisms of causal relationship in that domain, such as Newton’s laws in physics, principles of chemical reaction,

causes of revolutions, influences of geographical features on weather, and so on. Second, however, are the trans-domain capabilities of recognizing and reasoning through the cause-and-effect rhetorical frame and becoming attuned to the markers of the frame as they are used by writers—that is, the schemas related to cause-and-effect reasoning that can be considered an aspect of reading comprehension more generally. These latter capabilities are integral in "reading to learn" in content domains, that is, for coming to understand and reason through the specific C&E relationships in subject domains. Given that building schemas for cause–and–effect relationships in a domain is central to proficiency and that texts are correspondingly organized in cause–and–effect rhetorical frames, the failure to activate the appropriate schemas will seriously degrade comprehension and thus learning.

For this reason, cause-and-effect reasoning plays a prominent role in state and national statements of standards for education learning and assessment. The following examples illustrate C&E in K-12 assessments in both subject domains and reading itself.

The science assessments of Oregon (Oregon Department of Education, 2006a) and Vermont (Vermont Institute, 2001) require that students identify causes and effects. For example, the Vermont science assessment requires students to provide a prediction based on a testable hypothesis, where students are prompted regarding C&E relationships. The Massachusetts (MCAS) English Language Arts curriculum includes standards that require students beginning at 3rd grade to distinguish cause from effect (Massachusetts Department of Education, 2006). An Illinois reading assessment standard requires 5th grade students to represent the content of fiction passages using organizational patterns demonstrating C&E relationships, while students at 6th grade or above are required to identify such relationships from both fiction and non-fiction passages (Illinois State Board of Education, 2007).

The description of achievement levels for the National Assessment of Educational Progress (NAEP) indicates that a "Proficient" 4th grade reader should be able to recognize cause and effect relationships whether reading for literary experience or reading for information. The description for 8th grade indicates that students performing at the "Basic" level should be able to identify such relationship. In addition, 8th grade students should be able to make predictions when reading for 'Performing a task,' a more complex type of assessment task. Further description of the NAEP 12th grade science assessment addresses students' use of C&E reasoning with regard to solar system interactions. Specifically, "Proficient" students "recognize some inputs and outputs, cause and effects, and interactions of a system" (NAGB, p. 52), while "Advanced" students are able to "recognize cause-and-effect relationships within systems and can utilize this knowledge to make reasonable predictions of future events" (NAGB, p. 53).

Cause and effect reasoning is also used to distinguish levels of performance at signal achievement levels in many assessment systems. For example, Kansas designates five levels of performance: Unsatisfactory, Basic, Proficient, Advanced, and Exemplary. The performance level descriptors for 4th grade are marked by the following differences in expectations with regard to C&E relationships (Souther, n.d., slide 10):

- *Unsatisfactory* – The student performing at this level "is not likely to make connection or perceive relations in order to construct inferential meaning. This student struggles to determine cause and effect."

- *Basic* – The student performing at this level "makes minimal connection or perceives inaccurate relations in order to construct inferential meaning. This student inconsistently or inaccurately determines cause and effect."

- *Proficient* – The student performing at this level "makes obvious connections and perceives some relationship to construct inferential meaning. This student is likely to determine cause and effect."

- *Advanced* – The student performing at this level "makes connections and perceives complex relationships to construct inferential meaning. This student will link cause and effect."

- *Exemplary* – The student performing at this level "makes subtle or complex connections and perceives relationships to construct inferential meaning. This student will link cause and effect."

It is clear from these examples that creating tasks to assess aspects of C&E reasoning is important in assessing reading comprehension. Examples of such tasks abound. Each requires a student to interact with text in which a causal relationship of some kind is involved either explicitly or implicitly, create a situation model that captures the relationship in accordance with the principles of the content domain, and reason through the C&E structure in some way—for example, inferring potential causes for an effect, predicting effects from causes, organizing propositions in terms of causes and effects, or recognizing the possibility of alternate causes for an effect. Texts can vary in the degree to which text markers are used to signal that the cause-and-effect rhetorical frame is appropriate and how precisely the relationships among propositions are rendered in the surface structure. Tasks can vary as well in the degree to which content knowledge can be presumed or is itself a target of inference. The design pattern developed in Section 4.0 aims to make these design choices explicit, comment on the nature of the reasoning that is evoked, and illustrate their use.

3.0 Design Patterns and Assessment Arguments

The architect Christopher Alexander coined the term *design pattern* to characterize recurring problems and approaches for solving them in the domain of building projects and community planning. He lays out some 250 patterns, from "Front Porch" to "Mosaic of Subcultures" (Alexander, Ishikawa, & Silverstein, 1977). Once a designer recognizes she has encountered a situation to which a *design pattern* applies, the pattern provides support, rationale, and accumulated experience to help guide her design decisions. Extending the idea to software engineering, Gamma, et al. (1994) described twenty-three *design patterns* for object-oriented computer programming, with names such as "Builder," "Façade," and "Memento."

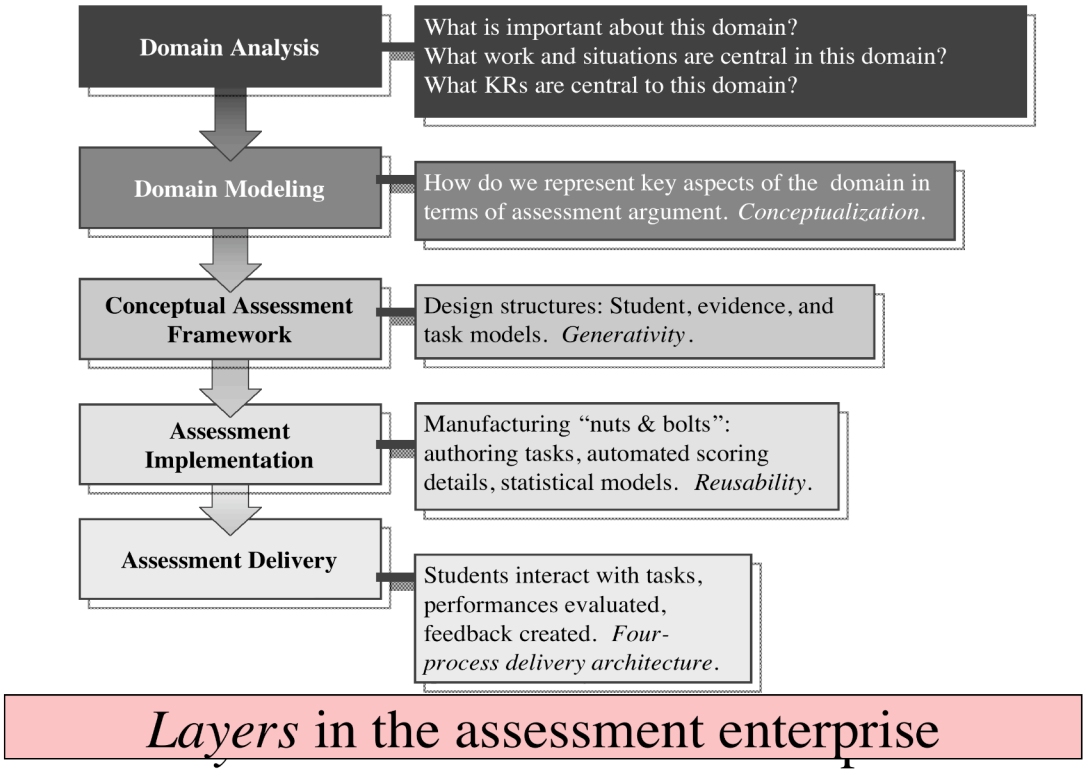
The Principled Assessment Designs for Inquiry (PADI; Mislevy & Haertel, 2006) project introduced *design patterns* for assessment tasks as a tool to support evidence–centered assessment design (ECD; Mislevy, et al., 2003). Assessment *design patterns* support communication among educators, domain experts, and assessment designers in a non-technical way about the meaningful aspects of inquiry around which assessment tasks can be built. They help designers "fill in the slots" of an assessment argument that builds around recurring themes in learning and approaches of obtaining evidence about them. This section provides a brief introduction to ECD and *design patterns* in order to set the stage for developing the *design pattern* for assessing C&E reasoning in reading comprehension.

3.1 Evidence-Centered Assessment Design

Evidence-centered design views an assessment as an evidentiary argument: An argument from what we observe students say, do, or make in a few particular circumstances, to inferences about what they know, can do, or have accomplished more generally (Mislevy, Steinberg, & Almond, 2003). The view of assessment as argument is a cornerstone of test validation (Kane, 1992, 2006; Messick, 1989). ECD applies this perspective proactively to test design. The ECD approach embodies design principles meant to instantiate an evidentiary argument in the objects and the processes of an operational assessment. The framework not only makes the underlying evidentiary structures more explicit, but also makes operational elements easier to reuse and to share.

In ECD, assessment is expressed in layers that provide structure for the different kinds of work and information at different stages of the process. Figure 5 shows the ECD layers. In the *Domain Analysis* layer, research and experience about the domains and skills of interest gathered—information about the knowledge, skills, and abilities of interest, ways people acquire them and use them, situations under which this knowledge is employed, indicators of successful application of the knowledge, and so on. The cognitive research on comprehension discussed in Section 2 is work in *Domain Analysis*. In the *Domain Modeling* layer, information from *Domain Analysis* is organized to form the assessment arguments. *Design patterns* are *Domain Modeling* tools that provide scaffolds for creating the substance of an assessment argument.

Figure 5: Evidenced-Centered Design Assessment Layers



While the other three remaining layers of the ECD framework are less directly related to the discussion of *design patterns*, they are noted for the sake of completeness. The *conceptual assessment framework* (CAF) concerns technical specifications for operational elements including measurement models, scoring methods, test assembly specifications, and requirements and protocols for delivery. An assessment argument laid out in narrative form at the *domain modeling* layer is expressed in terms of coordinated pieces of machinery, so to speak. The work in *assessment implementation* includes activities in preparation for testing examinees such as authoring tasks, calibrating items, finalizing rubrics, producing materials, producing presentation environments, and training interviewers and scorers, all in accordance with the assessment arguments and test specifications created in previous stages. The work in *assessment delivery* includes activities in presenting tasks to examinees, evaluating performances to assign scores, and reporting the results to provide feedbacks (see Almond, Steinberg, & Mislevy, 2003, and Mislevy & Riconscente, 2006, for further discussion on these layers).

3.2 Assessment Arguments

A closer look at assessment arguments shows how *design patterns* support assessment developments. Figure 6 extends Toulmin's (1958) general structure for arguments to the case of assessment arguments (Mislevy 2003, 2006). A series of logically connected claims are supported by data with warrants and open to alternative explanations. The *claim* indicates the proficiencies student could possess and the assessment seeks evidence of, in order to make valid and relevant inferences about the student. *Data* consist of students' behaviour observed in particular task situations, the salient features of those tasks, and other relevant information about the relationship between the student and the task situation (e.g., personal or instructional experience). Data provide support for the claim. *Warrants* indicate how responses in situations with the noted features depend on proficiency. A conception of knowledge and its acquisition—i.e., a psychological perspective—is the source of warrants that shapes the nature of claims an assessment aims to make and of the data needed to evidence them. *Alternative explanations* for poor performance are deficits in knowledge or skills required in carrying out a task. A claim may be qualified in light of alternative explanations, such as a lucky guess or unfamiliarity with assumed content.

Figure 6: An Extended Toulmin Argument Diagram for Assessment Arguments

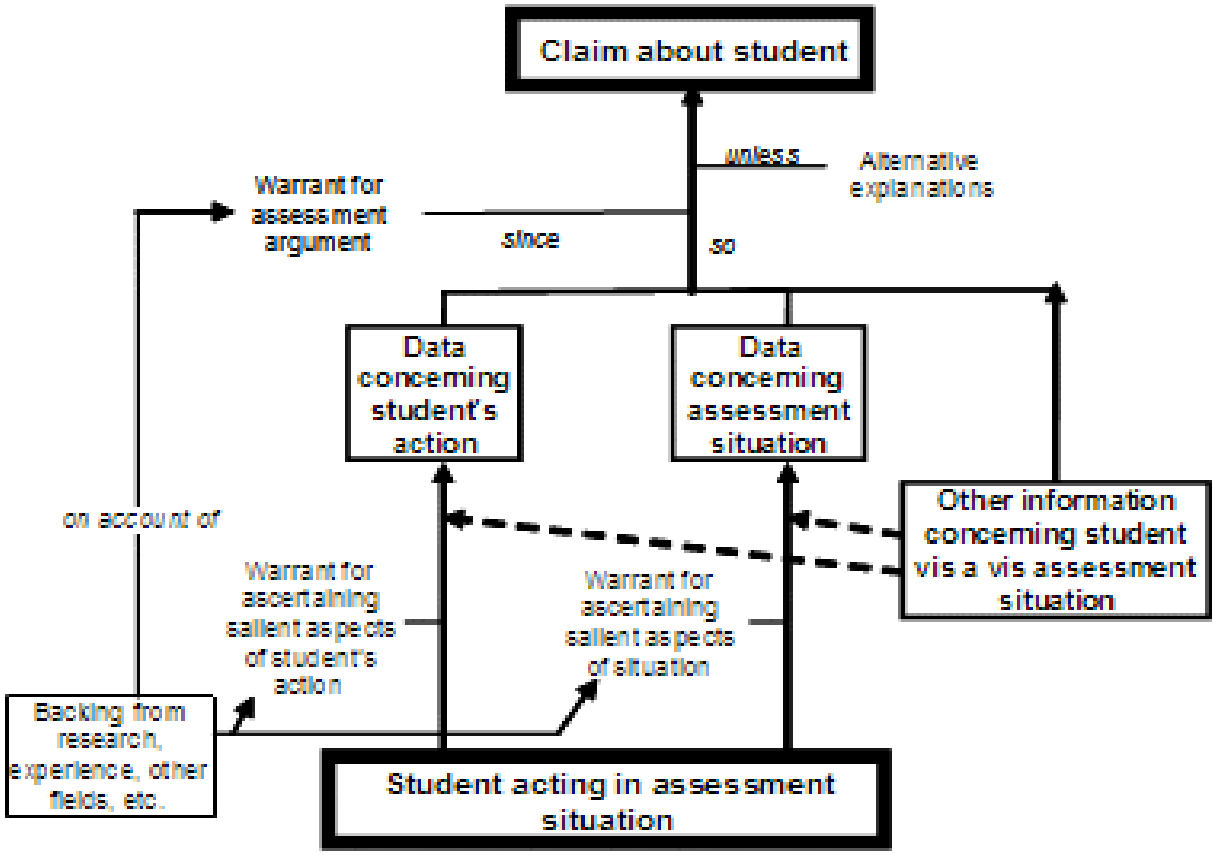


Figure 6 indicates the structure of an assessment argument but not its content. The assessment-design work in *domain modeling* essentially involves filling out this structure in terms of the claims, situations for observing students' actions, and ways of evaluating their performance as evidence about the targeted proficiencies. This reasoning is essential to good task design but often only implicit in test designers' work. *Design patterns* help a designer with this task by laying out important information about design choices as they arise for assessing the particular aspects of proficiency a *design pattern* has been built to address. The information in a *design pattern* is organized in categories ("attributes" is the technical term, from object modeling in software engineering) that focus attention on various elements in the argument structure and the relationships among them.

The major benefits of a *design pattern* in assessment development include reusability, transparency, and generativity (Mislevy & Riconscente, 2006). Reusability results because a *design pattern* encapsulates work completed at the *domain analysis* stage and organizes it in the form of an assessment argument. Transparency means that the rationale that underlies such task is explicit. A *design pattern* helps a prospective task designer get started readily by shaping the assessment argument in terms of aspects of knowledge and approaches that have been used in the past and supplying examples of tasks that illustrate the ideas. By articulating the critical components of the argument for tasks addressing the targeted capabilities, a *design pattern* thus sparks ideas for new tasks. Their particulars can be detailed with the content, purposes, constraints, and resources of the assessment at hand.

3.3 Attributes of a Design Pattern

A *design pattern* sketches the elements for a narrative structure concerning knowledge, skill, or abilities (KSAs) about which one wants to make a claim, the kinds of data that can provide evidence about acquisition of that knowledge or skill, and features of task situations that allow students to demonstrate them. Table 1 shows attributes of a *design pattern* in relation to the component of assessment argument. These attributes will be fleshed out next in the next section where we discuss the *design pattern* that was developed to assess reasoning in C&E rhetorical frame. (This table omits some less central attributes that appear in the extended version of the *design pattern* structure in Mislevy et al., 2003.)

Table 1: Attributes of a Design Pattern

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 KSA	The primary knowledge/skill/abilities targeted by this design pattern	Claim
Additional KSA	Other knowledge/skills/abilities that may be required by tasks motivated by this design pattern	Claim if relevant, alternative explanation if irrelevant
Potential Work Products	Things students say, do, or make that can provide evidence about the focal knowledge/skills/abilities	First stage in acquiring data about student's actions
Potential Observations	Features of work products that encapsulate evidence about focal knowledge/skills/abilities	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 knowledge/skills/abilities	Data concerning situation
Examples	Samples of tasks that instantiate this design pattern	
References	Research, applications, or experience relevant to task design under this design pattern	Backing

4.0 A Design Pattern for Cause and Effect Reasoning in Reading Comprehension

This section walks through the attributes of an assessment *design pattern* entitled "Assessing cause and effect reasoning in reading comprehension." The summary form of this *design pattern* appears in Table 2. The attributes discussed include Focal and Additional KSAs (related to claims in the assessment argument); Characteristic and Variable Task Features (related to data about the situation in which the examinee acts); Potential Work Products and Observations (related to data about the examinee's performance); and other pertinent but less central attributes. Points will be illustrated with the tasks introduced earlier and others drawn from existing assessments or constructed by the authors.

4.1 Summary and Rationale

The Summary simply states that this *design pattern* concerns the design of tasks that present students with situations to interact with written texts (and, as we will see, possibly additional forms of representation) that concern C&E situations and ask them to reason through the C&E schema to make inferences, predictions, or explanations. The Rationale, which corresponds to the warrant in Toulmin's (1958) argument structure, is that a student who has access to suitable information from long-term memory, activates it, and integrates it in a situation model with the appropriate C&E relationships, will be able to carry out the directed inferential task. Such warrants are backed by the research on C&E reasoning and rhetorical structures summarized in Section 2.

4.2 Focal Knowledge, Skills, and Abilities (KSA)

The primary organizing feature of a *design pattern* is the Focal KSA. Focal KSAs are the target of inference in an assessment, and they concern aspects of students' proficiencies in some domain at some grainsize. A *design pattern* can designate a cluster of related KSAs. In these cases, one design choice for a task author is whether to assess them all as a composite or to emphasize various aspects of them in ways that choices among Variable Features of the task will highlight. In measurement terms, Focal KSAs are the construct of interest, or the knowledge and skills that are meant to be assessed.

The focal KSAs that are called into play in assessing C&E reasoning include those listed below. This categorization into knowledge, skills, and abilities is somewhat arbitrary; it is clear that these are not lists of discrete, independently existing capabilities. The list merely groups aspects of the broader capability to instantiate a relevant C&E schema in the context of a text, integrate that schema with other information from the text and background knowledge, and reason through it.

- Knowledge of …
    - how phenomena —real or imaginary—can be connected in terms of causation.
    - how causal/logical connections between or among events are expressed in propositions.
    - what makes an event a cause and what makes an event an effect or outcome.

- Skill to …
    - distinguish between antecedent and precedent events described or portrayed.
    - distinguish causal connections from temporal relations or compare/contrast relations.
    - distinguish independent variables from dependent variables.
    - distinguish various forms of causal connections that may exists between and among events.
    - distinguish signal words for cause-effect relation from that of the other relations such as compare-contrast (e.g., ‘either …or’, ‘instead of’, ‘but’), temporal (e.g., ‘next’, ‘then’, ‘preceding’).
    - organize propositions as implied by C&E relationships, as indicated by rhetorical structure.
    - organize propositions as implied by C&E relationships, as indicated by marker words.
- Ability to …
    - reason in deductive or inductive mode; that is, from causes to effects (predictions), and from effects to causes (explanations).
    - trace relationships among causally connected events, for connections stated explicitly.
    - trace relationships among causally connected events, for connections stated implicitly.
    - hypothesize C&E structures (Broek & Kremer, 2000).
    - recognize signal words for cause-and-effect such as ‘accordingly’ , ‘as a result of’, ‘because’, ‘but’, ‘consequently’, ‘due to’, ‘for this reason’, ‘if … then’, ‘in order to’, ‘nevertheless’, ‘not only … but’, ‘on account of’, ‘since’, ‘so’, ‘that’, ‘thereby’, ‘therefore’, ‘this leads to’ (Irwin, 2002).
    - organize propositions in C&E structures in accordance with the relationships indicated by signal words.

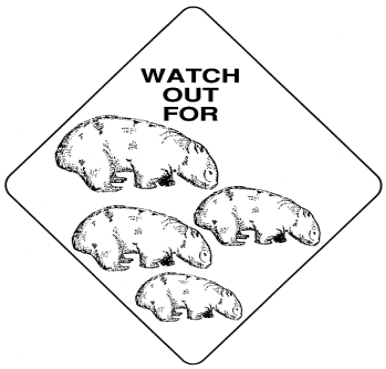
Table 2. A Design Pattern for Assessing Cause and Effect Reasoning Reading Comprehension

	Cause and Effect Reasoning in Reading Comprehension
Summary	This design pattern motivates tasks that require students to reason through cause-and-effect schema to make inferences, predictions, or provide explanations.
Rationale	Cause-and-effect relationship is central to human reasoning, in everyday life as well as in the disciplines. This reasoning stands at the center of scientific reasoning, legal reasoning, troubleshooting, and also in human relations. Knowledge about causal relationship is critical in interpreting text – expository or narrative. Failure to activate cause-effect schema can degrade comprehension and thus, learning.
Focal KSAs	Knowledge of how real or imaginary phenomena are connected in terms of causation, what makes an event a cause and what makes an event an effect, and how causal connections are expressed in propositions; Skill to distinguish various forms of causal connections, distinguish causal connections from logical or temporal connections, distinguish signal words for cause-effect relation from that of the other relations, organize propositions as indicated by rhetorical structure, marker words; Ability to reason in inductive or deductive mode for predictions or explanations, trace relationship among causally connected events whether stated explicitly or implicitly, hypothesize in cause & effect structures, and recognize signal words for cause-effect relationship.
Characteristic features	Prose presentation of a situation involving one or more cause & effect relationships; directive requiring examinee to reason about or through those relationships.
Add'l KSAs	Familiarity with the substantive relationship that is the basis of the cause & effect structure Ability to read, listen, observe, and respond orally or in writing, follow information portrayed/presented in various representational forms, identify elements of propositions. Knowledge of various types of text passages, and notations, signs, and symbols used in various disciplines, and requirements involved in various task situations. Skill to extract information presented in charts/graphs, pictures, flow-charts, etc., and to use them when making a response.
Variable features	Nature of C&E relationship(s) with regard to Lakoff's canonical C&E schema Structure of C&E relationship(s): simple, chained, multiple effects, multiple possible causes, multiple conjunctive causes C&E relationship(s) implicit, explicit, or mixed? Direction of reasoning required – cause to effect or effect to cause Representational forms – whether some information is presented in visual, oral, or hybrid form Length and complexity of prose passage. Degree of substantive knowledge necessary to perform the task.
Potential work products	Selection of adverbials or answer choices, completion of table/graphic organizer, sequencing cards or pictures to structure a narrative, oral or written explanations or predictions, identifying missing elements, re-enact a story or event
Potential observations	Is the C&E relationship appropriately explained or distinguished from other forms of relationship or appropriately depicted through a diagram or correctly constructed when the events are not presented in canonical order? Were C&E signal words used appropriately? How accurately a diagnosis is made given symptoms, or causes hypothesized given effects? Are there conceptual errors or misunderstandings that indicate the C&E relationship has not been properly constructed? Are rhetorical strategies being used appropriately in explanation?
Selected References	Irwin, J. (2002) Kintsch, W. (1998) Kintsch & Yarbrough (1982) Lakoff, G. (1987) National Assessment Governing Board (2004a,b) Rayner & Pollatsek (1989)

The focal KSAs called in Example Task 2 presented in Figure 3 includes ability to recognize signal words for cause-and-effect relationship and skill to organize propositions using adverbials to indicate C&E relationship. The task also demands knowledge of how causal connection between and among events is expressed. A test-taker without such knowledge can find it difficult to select the appropriate signal words in organizing the propositions to explicitly state the causal connections among events. Substantive knowledge is also involved in this task. In item a) of this task, the test-taker is required to know that an object is static. A static object needs some force to slide on the floor; in this case, Max's pushing provided the force necessary for the heavy object to slide. The task authors expected that this relationship would be familiar to all test-takers from everyday experience. Example Task 4 (shown in Figure 7), like Example Task 2, involves use of signal words for C&E relationships. However, the focal KSA primarily entailed in Example Task 4 is tracing relationship among causally connected events, for connections stated explicitly by use of the signal words.

Figure 7: Example Task 4: Cause & Effect with Various Question Types for a Narrative Passage

In this section, you will be given …minutes to read the following passage and answer questions



Wombats

As we rode along the highway sixty miles northeast of Adelaide, Australia, a diamond-shaped sign suddenly loomed ahead. Watch out for Wombats, it warned. We peered into the sparse scrub along the roadside and searched for the brown furry animals. In the distance we spotted a mob of red kangaroos bouncing out of sight, and near the road a crow like bird called a currawong was perched, but nowhere did we see any wombats. However, we later found out that this was not surprising because we were traveling during midday, and wombats are active mostly at night. It wasn't until we visited the animal reserve that we finally saw our first wombat and learned more about this funny-looking creature.

A wombat is a shy and gentle animal. But even if you lived in Australia and were willing to keep watch during the nighttime hours, it would be difficult to get to know one. As more and more people move into territories in which wombats live, they destroy the wombat's burrows and food supplies. In some areas where the wombat was once plentiful, it is now almost extinct. Animal reserves have been set up recently to protect the wombat. Perhaps with a little help these friendly creatures will gain prosper and multiply.

Why aren't wombats seen by people often?

a. Wombats look too much like koalas so people do not know which one is wombats
b. Wombats usually are active at night when very few people travel
c. There are not enough wombat-crossing signs to tell where wombats can be seen
d. Wombats are difficult to see because they live in trees

What would wombats do upon seeing people? Use the space below to write and explain your answer.

_____

_____

_____

NOTE: The text for Example Task 4 is excerpted from a narrative in public domain of NAEP Questions Tool: http://nces.ed.gov/nationsreportcard/itmrls/. Authors prepared the questions.

The Focal KSAs primarily involved in Example Task 5 (Figure 8) are ability to trace relationships among events described in a text passage and knowledge of what makes an event a cause and what makes an event an effect. The task calls for skills to distinguish various forms of C&E relationship that may exist among events, because the example depicts a situation where a single cause leads to multiple effects, and each effect is a cause for the event in next tier. The test-taker is required to organize the events to make the relationship more explicit.

Example Task 3 (see Figure 4) is a hands-on task that asks the student to portray cause and effect relationship in real phenomena. It requires both reasoning deductively to anticipate what will happen when the vinegar is added, and inductively to explain the outcome. The test taker needs the skill to distinguish various forms of causal connections. The causal relation involves an "AND" relation between multiple causes and an effect. The effect, namely the volcano eruption, requires multiple causes to occur, namely the baking soda, the vinegar, and bringing them together. The chemical reaction provides the substantive context of the task and would not be expected to be familiar from everyday experience. This C&E task, therefore, would be appropriate for assessing two rather different constellations of capabilities. Among students who are studying the chemical reaction that is the basis of the task, the target is the explanation in terms of the C&E relationships of that particular substantive model. Among students who are not known to be familiar with the model, the focus would be on structuring the events in terms of C&E relationships without having to provide the "hidden" steps in terms of the chemistry model.

4.3 Additional Knowledge, Skills, and Abilities

Cause-and-effect reasoning is always about some particular content, in some context, for some purpose, and evidenced in some manner. Any assessment task is necessarily contextualized, and building a suitable situation model depends on Additional KSAs as may be required, as well as the Focal KSAs that concern the instantiation and use of C&E schemas. Additional KSAs are unavoidable because it is always necessary that a task provides information in some manner, using some language and representational forms; that it provides a context drawing upon some particular situations; and that students must respond using some format, medium, and representational form. The Additional KSAs attribute of an assessment *design pattern* alerts the task developer to recurring considerations of knowledge that can be involved in tasks assessing the Focal KSA—additional knowledge or skills that can either enrich the assessment argument or subvert its intent. The task developer will make design choices by means of the Variable Features to circumvent some Additional KSAs, include others because they are relevant to the intended claim, or include others by necessity even though they introduce alternative explanations for poor performance but are unavoidable under the constraints of the project.

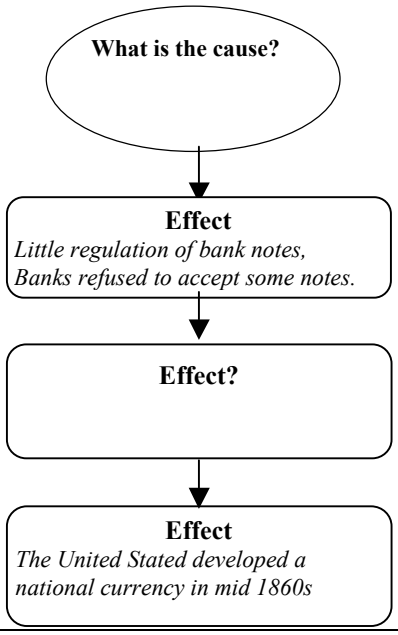
Figure 8: Example Task 5: Graphic Representation of Events Portraying Cause and Effect

Read the passage and complete the graphic representation with information appropriate for the empty boxes.

"The U.S. constitution states, "The Congress shall have Power … to coin Money," This passage gives the federal government alone the right to issue legal currency, or money, for the nation. However, until the mid-1800s, many state banks could print bank notes. These notes could be exchanged at the bank that issued them for their value in gold or silver coins.

As a result, there were many types of bank notes but little regulation of them. Banks often refused to accept or honor the value of other banks' notes. The issuing of so many types of bank notes disrupted economic transaction. The federal government, therefore, stopped the state's issuing of bank notes and developed a national currency in the mid-1860s.

In the past 100 or so years, the government has made some changes to the money issues. The U.S. Treasury discontinued printing the $500, $1,000, and $10,000 bills and introduced the $2 bill and Susan B. Anthony dollar coin. The treasury also has worked on redesigning paper money to make it harder to counterfeit. These new bills feature larger portraits than the older bills did. As Time magazine writes of the new $100 bill, Benjamin Franklin "now dominates the bill like a movie star in a newspaper advertisement" (Irwin, 2002).



NOTE: Adapted from examples presented in Irwin, J. (2002). *Reading Strategies for the Social Studies Classroom.* New York: Holt, Rinehart and Winston.

Additional KSAs*,* generally speaking, are other knowledge, skills, and abilities that may be required in a task that addresses the focal KSAs. They affect the validity of assessment arguments by weakening, strengthening, or conditioning the value of evidence from task performance. The concepts required to control Additional KSAs can prove challenging and subtle and are addressed explicitly only in advanced discussions of assessment (such as Messick, 1989, 1994). Indeed, one of the hallmarks of an expert test developer is handling Additional KSAs in a manner that optimizes the evidentiary value of tasks in light of a test's purpose, the target examinee population, and the constraints of administration conditions. A *design pattern* is particularly useful to the nonprofessional designing tasks by providing contextualized help for thinking about Additional KSAs.

Additional KSAs can degrade the value of information about an examinee's response when knowledge or skill is required or is beneficial to response in addition to the Focal KSA and lacking it can lead to poor performance. The medical content of Example Task 1 (Figure 1) provides an extreme example. A skilled reader can draw some C&E inferences by analyzing the structure of the propositions and capitalizing on signal structures in the text. If the reader happens to understand the underlying physical mechanisms, however, it is possible to fill in links that are not stated in the text and build a richer situation model that supports a wider range of inferences. Whenever the goal of a task is to assess an examinee's ability to assemble and reason through C&E structures as signaled in text, and *not* to assess familiarity with particular substantive causal relationships, then the content of the task should be familiar to all examinees. This can be accomplished by using content that is familiar to everyday life or more specialized content that is known to be familiar to all the members of the examinee population, as in the first item in Example Task 2 (see Figure 3).

To highlight the point that what is familiar depends on the targeted examinee population, we may consider medical situations that may be unfamiliar to a lay audience but are commonplace to health professionals. They require Additional KSAs that are necessary for success but do not raise alternative explanations for poor performance for this specialized population. This is the design choice made in the Occupational English Test (OET; McNamara, 1996), which is used to test the English language proficiency of medical professionals immigrating to Australia. The use of health related contexts and content in the OET reading comprehension test introduces a demand for Additional KSAs *presumed to be familiar in the examinee population* and thus enhances validity by mirroring the targeted language use situations.

As noted previously, C&E schemas are central to reasoning in substantive domains, through processes and relationships that characterize the phenomena that domain addresses. Comprehension of such relationships from prose, perhaps as supplemented by information in additional forms such as graphs and diagrams, can thus be an important target of learning and thus of assessment in those domains. C&E reasoning tasks addressed in this *design pattern* can be useful for assessment in subject matter domains. The same considerations concerning the psychological backing apply, although now the substantive schemas that must be synthesized with the text base are integral to the claim in the assessment argument. Such a task provides evidence about whether a student can build and reason through a situation model that uses a substantive causal relationship that is also at issue. In these cases, the test designer may deliberately chose values of Variable Features that lower the difficulty caused by rhetorical structures and signal words. Example Task 1 would illustrate this

reasoning when the goal is assessing a medical student's understanding of the physiology of the situation.

In sum, Additional KSAs may be avoided in order not to interfere with the inferences about Focal KSAs. Alternatively, if it is known that the intended examinees have sufficient levels of a given Additional KSA, it may be included in a task to add variety or realism without introducing construct-irrelevant variance. A intent to focus on C&E reasoning *per se* requires examinees be familiar with the context of a task, that is knowledge of the particular scientific model (e.g., biology, physics, medicine, history) and the language used as a medium to present the task.

Tasks assessing C&E reasoning in reading comprehension tests often require examinees be able to do the following:

- Read (decode, read with fluency), listen, or observe. That is, by whatever modality the surface presentation of information is conveyed, the knowledge, capabilities, and proficiencies necessary to acquire information in this mode are required.

- Identify elements of a proposition. Examinees need to be able to identify the predicate-argument scheme of the idea units expressed in sentences, a basic unit of a language.

- Follow information portrayed/presented in various representational forms—e.g., paper and pencil, on-line, and performance. That is, the conventions of a genre (e.g., novel, movie, theater, drama, and play) may be employed to convey information to the examinee. Lack of familiarity with the genre can present an alternative explanation of poor performance.

- Extract information presented in charts/graphs, pictures, or flow chart. The conventions of representational forms may be necessary to acquire propositions for the text base that are expressed in forms in addition to prose, as are often found in instructional and reference materials.

- Respond orally or in writing (paper and pencil or electronically). The capabilities and familiarity with conventions and expectation of response formats are ancillary to C&E reasoning, but they are required to perform well.

- Present information using graphic organizers such as path analysis. Similar to the preceding point about representational forms besides prose, except now with regard to responding.

- Distinguish various types of text passages such as narrative, expository, and persuasive. Familiarity with standard text forms and purposes provides information to organize propositions in text and to fill in gaps.

- Recognize notations, signs, symbols used in various disciplines.

- Recognize the requirement of various task types—e.g., multiple-choice, open-ended, fill-in-blank, and hands-on. Task types are themselves rhetorical forms used in the domain of assessment, and knowledge of how they are used to present information, expectations they entail about the kinds of reasoning that is required, and prototypical ways of

responding are often presumed, but are extraneous sources of difficulty for the student who is not familiar with them. Note, for example, that the different introductory examples use different response modes, each of which has its own knowledge requirements for being able to produce a response. Lack of familiarity with an interface or with the standards for evaluating open-ended responses are alternative, construct-irrelevant, explanations for poor performance raised by KSAs associated with response modalities.

Additional KSAs involved in Example Task 2 (Figure 3) are the ability to read, ability to identify predicate-argument scheme of the idea units expressed in sentences, and recognize that it requires a response in writing; i.e., the task requires selection of a signal word in writing. Additional KSAs in Example Task 5 (see Figure 8) are the ability to read and familiarity with expository passages. Although the task asks for the ability to respond in writing similar to Example Task 2, the response option additionally requires use of a graphic organizer, so familiarity with this representational form is an Additional KSA. The graphical form can be a useful aid to students for analyzing C&E relationships if they are familiar with it but a source of construct irrelevant variance in assessment if they are not. The Additional KSA of ability to read involved in Example Task 3 (see Figure 4) is joined by the capability to carry out an instruction to manipulate physical objects. Further, this task requires a constructed response to describe the outcome of that manipulation. The writing capabilities can be sources of construct irrelevant variance as well, as would be seen in a student who could explain the reaction orally but not in written form. As mentioned previously, medical knowledge is an Additional KSA in Example Task 1.

There are many ways students can be placed in situations that involve these constellations of knowledge and skill, all of which emanate from the use of schemas involving C&E—some generic, some content specific, some focused on prose conventions. This leads to the questions of whether they are available, whether they can be accessed to construct situation models, and whether they support appropriate reasoning in terms of C&E relationships. Features of task situations that are required to provoke C&E reasoning in text comprehension will be discussed next in Section 4.4 as Characteristic Features of tasks. Features that can be varied to adjust task difficulty, to focus on different aspects of C&E reasoning, or to include or exclude various additional elements of knowledge will be discussed in Section 4.5 as Variable Features of tasks.

4.4 Characteristic Task Features

Characteristic Features of tasks are central to evoking evidence about the Focal KSAs. All assessment tasks motivated by the *design pattern* need to possess them in some form. All tasks inspired by a "Formulating Scientific Explanations" *design pattern*, for example, involve a real world situation and a requirement for the student to identify, create, or support a claim about the underlying process or pattern that explains them. For C&E reasoning in reading comprehension, the prose presentation of propositions that correspond to a situation with causal relationships is required. The student must provide a response that in some manner indicates that he or she has constructed a situation model that appropriately incorporates the relationship. A great deal of latitude remains for task construction, but if these requirements are not met, it will be difficult to justify a task as providing evidence of C&E reasoning in text comprehension. We see

that these requirements are met in all example tasks in various ways. For instance:

In Example Task 2 (Figure 3), the student must make a selection from the given adverbials to connect the events described by two sentences in each case. The student must take the information presented in the propositions, use background knowledge to relate them in a C&E structure in a situation model, then use a rhetorical device—a compound sentence with a marker word—to explicitly signal the relationship.

In Example Task 5 (Figure 8) the student is asked to complete a graphic organizer. Compared to Example Task 2, the prose passage is more complex. The focus here is on identifying the C&E relationships the author has indicated in the passage, which form the macrostructure of the text and are signaled by markers at the sentence level. The same organizer could be used with a passage in which the C&E structure was less central, and in which C&E relationships were implicit rather than explicit.

Example Task 3 (Figure 4) involves a real world situation where test-takers are asked to make a prediction (identify effect) and explain their reasoning (identify cause or causes). The relevance of C&E reasoning is signaled by the form of the questions: 'What will happen? Why?' etc. A good response for this example will indicate the underlying rationale for the prediction or explanation—that is, why the effect event happened and how it happened.

Example Task 1 (Figure 1) is a typical task type used in domain areas: A passage describing a substantive situation, in which C&E relationships that are characteristic of the domain play a role. In such cases, the content of the propositions and the structure of the text are usually not sufficient for an examinee to perform well if he or she is not familiar with the domain. The substantive principles and relationships in the domain provide the schemas that are necessary to build a situation model upon which an adequate response can be built.

4.5 Variable Task Features

Variable Features of tasks are a primary tool of task developers. They can take different values in order to adjust the difficulty of tasks, to focus their evidentiary value on different aspects of the Focal KSA, or to incorporate or circumvent particular additional KSAs. In regard to reading comprehension, "Whether a coherent mental representation is constructed during reading also depends on the properties of the text. The same information can be conveyed in different ways, with certain forms being more user-friendly. A user-friendly text passage, for example a passage that does not include tangential details or connections that a reader must make are clearly delineated, or relevant background knowledge is explicitly presented, can reduce the demands on readers' cognitive capacities and inferential skills" (Broek & Kremer, 2000, p. 14). Properties such as these can be manipulated to vary the difficulty of tasks when test-takers are at certain ability levels.

More importantly, the features can be manipulated to increase, decrease, or avoid certain aspects of knowledge. For example, younger students can be given passages concerning everyday situations so that the focus is on awareness of the C&E rhetorical structure in the text and instantiating a situation model with the C&E schema for a familiar relationship. Older students can be

given passages with less familiar content (though probably not quite as unfamiliar as Example Task 1!) for building a situation model leaning more heavily on text cues and structure than contextual knowledge. Similar considerations arise with respect to the implicitness or explicitness of signaling. Excluding textual C&E signals places more of a burden on a reader to recognize and instantiate C&E relationships. Including them provides more scaffolding, as in Example Task 4 (Figure 7), and may be more appropriate for younger students when the focus is learning to use those markers, or for older students when the inferential focus includes the substantive relationships as in Example Task 2.

Categories of Variable Features that are germane to assessment of C&E reasoning in reading comprehension include the following:

- Structure of the C&E situation
  - Distance from the prototypical C&E structure.
  - Pattern of relationships—e.g., single cause to single effect, multiple. causes to a single effect or single cause to multiple effects, or an effect that becomes a cause for another effect.
  - Implicitness or explicitness of the C&E situation.
  - Type of text passages: narrative, expository, argumentative, conversational.
  - Purpose of reading—e.g., reading for literary experience, to perform a task, for information.
  - Degree of use of signal words.
  - Proximity of events that are causes and effects.
  - Length and readability of passages.
  - Whether passage includes irrelevant details.
- Degree of substantive knowledge required. This feature interacts with the number of intermediate propositions that may be missing in an explanation or narrative. It also interacts with whether substantive knowledge is also a targeted claim. As Example Task 1 (Figure 1) illustrates, when substantive principles or models are a target of assessment, these are exactly the ones the task designer can omit from a passage because the examinee's success in building a situation model depends on being able to employ that knowledge. On the other hand, when it is reasoning through C&E per se that is the target, then omitting C&E steps introduces irrelevant variance unless the situation can be presumed familiar to all the test takers.

- Direction of reasoning required. "Cause to effect" (deductive) reasoning through a given C&E structure is typically easier than "effect to cause" (inductive) reasoning. No less an expert than Sherlock Holmes tells us, "Most people, if you describe a train of events to them, will tell you what the result would be. They can put those events together in their minds, and argue from them that something will come to pass. There are few people, however, who, if you told them a result, would be able to evolve from their own inner consciousness what the steps were which led up to

that result. This power is what I mean when I am talking of reasoning backward, or analytically" (Doyle, 1986, p. 100).

- Representational forms. Prose text is usually present in reading comprehension, but other forms of presentation may be additionally included:
  - Visual (e.g., charts/tables, diagrams, pictures).
  - Oral (e.g., speech, narrations, sound bits).
  - Hybrid – a combination of visual and oral (e.g., movies, television).

4.6 Potential Work Products and Potential Observations

Central to an assessment argument are data in the form of what the examinee says, does, or makes. The attributes of a design pattern called Potential Work Products, Potential Observations, and Potential Rubrics all concern how to capture information from examinees' performances concerning C&E reasoning. There are many possible ways of getting evidence about the target KSAs from student performances and extracting the information about the Focal KSAs. Potential Work Products are examinees' responses or performances that hold clues about the Focal KSA. They are what students say, do, or make in task situations that provide evidence of understanding of what makes an event cause or effect and their relationship. Various features of the work products either by themselves or in combination provide examiners evidence that examinees know and be able to do C&E reasoning. The Potential Work Products for assessing C&E reasoning include the following:

- Verbal (oral or written) responses such as
  - written constructed responses, from short word or phrase to extended essays.
  - selection of an answer from given response choices.
  - completion of table(s) or diagram where elements are missing or identification of the correct elements in various diagrams to indicate C&E relationships, as in Example Task 5 (Figure 8).
  - indication (such as highlighting) of causes that contribute to an effect.

- Hands-on responses such as
  - sequencing cards or pictures to structure the narration of a series of events, as in Example Task 6 (Figure 9).
  - developing a scenario to illustrate or demonstrate the events (as a trial lawyer might in re-enacting a crime, to justify a particular C&E explanation for evidence).

- Oral responses such as explanations and summaries of C&E links.

Figure 9: Example Task 6: Sequencing Events to Portray Cause and Effect

The Lakota Sioux

The Black Hills of South Dakota were sacred to the Lakota. They called these mountains 'the heart of everything that is." When gold was discovered there in 1874, thousands of white people poured into the area. The Lakota fought back again. In 1876, Lieutenant Custer and more than two hundred soldiers attacked thousands of Lakota and Cheyenne warriors. The Indian warriors killed Custer and his entire force in what became known as the Battle of the Little Bighorn.

But the Lakota could not resist U.S. Army for long. One by one, Lakota leaders like Crazy Horse and Sitting Bull surrendered. In 1890, Chief Sitting Bull was murdered on Standing Rock Reservation, where the U.S. Army was holding him. His people fled to the Pine Ridge Reservation in South Dakota for protection. However, U.S. soldiers killed more than 250 Lakota. After this tragedy, the Lakota stopped fighting.

During this time, white hunters had almost wiped out buffalo herds. Without the buffalo, the Lakota way of life ended. Most Lakota then lived on reservations where life was very hard. There were few ways for them to make a decent living (Santella, 2001).

The following text boxes can be arranged in order of time to reflect cause and effect. Drag and drop them onto the scale in order to explains that *Lakota did not have any choice but living in the reservation.*

Buffalo herds were wiped out by white hunters.

Lakota way of life ended when buffalo herds were wiped out.

Most Lakota then lived on reservation where there were few ways for them to make a decent living.

All Lakota leaders were killed.

The Lakota could not resist U.S. Army anymore.

1	2	3	4	5

NOTE: The text for this example task excerpted from 'Santella, A. (2001). *The Lakota Sioux*. New York: Children Press. Authors created the C&E question for an online task.

In conjunction with suggesting Potential Work Products, a *design pattern* also prompts assessment designer to ponder about how to evaluate the work products to identify and characterize evidence about Focal KSAs that the Work Products convey. Potential Observations are salient characteristics of what students say, do, or make that constitute evidences of the Focal KSAs. They describe qualities, strengths, or extent of work that tends to distinguish those with more rather than less capability, on the whole or in selected aspects. Formalized and tailored to the specifics of a task, they are the basis of what are commonly called "item scores." The idea here is to determine, from an observable response, whether in interacting with the text the student has formed a situation model with appropriate C&E relationships and can reason through it as required. In the context of C&E reasoning, the following qualities can be examined. Note

that many of these qualities require particular kinds of Work Products, but others can be evoked from a variety of Work Product forms; the focus in the assessment argument is not so much on the form of the Work Product but the nature of the thinking it reveals. Also, most of the observations can be made with respect to familiar, everyday C&E situations or relationships in substantive areas such as science and history when C&E reasoning in the domain is at issue.

- Are ideas/events/propositions organized to indicate a C&E relationship? This characteristic can be identified from Work Products in the form of answers to multiple choice tasks when the choices include incorrect or mis-ordered alternatives, as in items 3 and 4 of Example Task 4. It can be identified from open-ended responses as to the appropriateness of the explained relationships, as in items 1 and 2 of Example Task 1. It can be identified from the construction of a form such as the diagram in Example Tasks 5 and 6 as to whether propositions have been appropriately mapped into a C&E schema.

- Is the C&E relationship being properly distinguished from among other possible connections, such as logical, correlational, and chronological?

- Are events that hold C&E relationships but are not presented in canonical order being correctly reconstructed? Note that this kind of observation can be implemented in Work Products that take the form of multiple choice questions, time lines, filling in or constructing diagrams such as those in Task Examples 5 and 6, or open-ended essays or oral explanations.

- Are events in movies, plays, or stories that hold C&E relationships but are not presented in canonical order being correctly reconstructed? As with Example Task 6, a text passage can be rearranged that does not follow a canonical order, and students can be asked to reconstruct the order.

- How accurately are causes hypothesized, given effects—e.g., diagnoses made from a given set of symptoms? That is, is the student instantiating an appropriate model and carrying out the backward or inductive inferences? This kind of observation (and the next one) can be obtained from multiple choice tasks where the alternatives include prior states that would generally not lead to the given effects. Open-ended responses are even more informative.

- How accurately are effects predicted, given causes, such as prognoses given symptoms? That is, is the student instantiating an appropriate model and carrying out the correct forward or deductive inferences?

- What is the quality of the rationale given for predictions, retrodictions, or explanations? Is an appropriate C&E structure proposed? Is the reasoning through the proposed model coherent? Are there conceptual errors or misunderstandings that indicate the C&E relationships have not been properly constructed (Graesser, 1981; p.21)?

- Are rhetorical strategies being used appropriately? Example Task 2 is a straightforward example that focuses on the use of rhetorical signaling devices, in a easy-to evaluate form. In open-ended tasks like the second item in Example Task 4, one evaluates the degree to which a

constructed response such as an essay uses these devices to communicate the targeted C&E relationships to the reader. This includes both signaling words and ordering of C&E relationships.

For any given task, the specific definition of an observable variable and the way its value is determined from a student's Work Product are given by an evaluation procedure or rubric. For an opened-ended task, the author has the design choice of how many and how detailed observable variables to implement. For example, to get detailed diagnostic information, observable variables can be defined at a fine grain-size, say at the level of detail of individual elements of KSAs discussed in Section 4.1. For overall proficiency, a global rating combining the use of many aspects of C&E reasoning may be combined holistically.

Figure 10, for example, presents a rubric for the open-ended response to Example Task 4. Note that this rubric encompasses both C&E reasoning and outside knowledge about the behavior of shy animals. This knowledge is expected to be available to students, but if it is not it can be an alternative explanation for a score that is less than it ought to be, based just on capabilities of reasoning through C&E relationships. Formulation of appropriate rubrics depends on the kinds and qualities of inferences one wants to make about students. Different observations can be derived from the same Work Product depending on which KSAs are of interest and the assessment's purpose.

Figure 10: Sample Rubric for the Open-Ended Task in Example Task 4

Extensive

This response demonstrates an in depth understanding of what wombats‘ are like (shy and gentle) and making a prediction appropriate to the nature. The prediction appropriate to the nature can be of wombats. For example:

> "Wombats will probably hide or run-away from people because wombats are shy animal. Also, wombats will not cause any harm to people as they are gentle animal.”

Essential

This response demonstrates an understanding of what wombats’ are like (shy and gentle). The response will suggest what wombats can not do suitable for ‘shy and gentle’. However, the answer is not as complete to receive ‘Extensive’. For example:

- "Wombats will stay away from people because wombats are shy and will not growl because wombats are gentle.
- "Wombats are shy, so wombats will run away and will not attack or growl as wombats are gentle.

Partial

This response demonstrates some understanding of the characteristics of wombats but could not make a prediction. Or, make a prediction without explaining why. For example:

- "It would be difficult for wombats to see people. Even when wombats see, they do not harm people. They are gentle.
- Wombats will stay away from people because wombats are shy.

Unsatisfactory

This response demonstrates little or no understanding of wombats’ nature (shy and gentle). Say something not relevant or appropriate. For example:

- "Wombats are so heavy, people can be smashed"
- "Wombats need protection from people"
- “Wombats will not see people. They are fuzzy animal”
- "Wombats sleep during the day. People travel during day"

Note: Authors prepared this rubric following what is available in ‘NAEP Question Tool’: http://nces.ed.gov/nationsreportcard/itmrls/

5.0 Conclusion

These are exciting times for educational assessment. On the one hand, there are increasing demands for consequential tests in schools and states, at larger scales and with higher stakes than we have seen before. On the other hand, advances in fields that bear on assessment arrive almost daily, from cognitive research, educational technology, and measurement modeling. Both of these developments outstrip traditional methods for designing, implementing, and carrying out assessment. Traditional methods rely on familiar item forms, item types, rules-of-thumb, and insights of test developers—mostly a matter of custom and procedure rather than underlying principles of assessment and the domain at hand.

In response, assessment researchers have been working to make explicit the underlying principles of assessment and develop methodologies that support the design of new forms of assessment that build on new understandings of cognition (e.g., the model-based research noted in Section 4.7) and new forms of assessment (e.g., simulation-based assessment). A view of assessment as evidentiary argument grounds this active line of work (National Research Council, 2001). All phases of assessment design are viewed in terms of their contribution to explicating an argument and implementing machinery to embody it. Theory-grounded backing for task design can draw variously upon cognitive studies from the information-processing, expertise research, situative psychology, and sociocultural literatures. Embretson (1998) illustrates the approach with psychological ability tests, for example. Bachman and Palmer (1996) provide practical guidance for task-based language tests that is grounded in psycholinguistic and sociocultural research. Baker (1997) and her colleagues structure tasks around "big ideas" in learning domains including social studies and middle school mathematics. We note that "big ideas" themselves often revolve around cause-and-effect relationships in the domain and that assessment itself is reasoning from students' responses as effects caused by the nature and organization of their knowledge and skills.

As one effort in this area, the Principled Assessment Design for Inquiry project (Baxter & Mislevy, 2004, Mislevy & Haertel, 2006, Mislevy & Riconscente, 2006), developed *design patterns* for building tasks around key aspects of reasoning in science (Mislevy et al., 2003). This presentation has applied the same structure to authoring tasks to assess students' capabilities in cause-and-effect reasoning as a component of reading comprehension. It builds on psychological research on reading comprehension, especially Kintsch's (1998) Construction-Integration (CI) theory of comprehension. The goal is to provide a motivating framework to guide the work of teachers, researchers, and test developers who wish to create reading comprehension tasks. It is neither a procedural guide nor a recipe book for this purpose; creative thinking and the art of assessment design are still required and, in large part, will always be required. What this application seeks to do, however, is to relate task authoring in this particular area of reading comprehension to contemporary research and to provide language and concepts for explicating the arguments that underlie cause-and-effect assessment tasks. If this initial effort proves useful, we hope it can lead to broader and more detailed development of a broad range of design patterns in the area of reading comprehension, drawing variously on research in cognitive, social, developmental, rhetorical, and substantive aspects of reading.. Used singly or in combinations, such a collection would help test developers create reading tasks that target the knowledge, skills, and abilities that match their assessment purposes.

References

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

Bachman, L.F., & Palmer, A.S. (1996). *Language testing in practice*. Oxford: Oxford University Press.

Baker, E. L. (1997). Model-based performance assessment. *Theory into Practice, 36*, 247-254.

Bartlett, F.C. (1932). *Remembering*. Cambridge, UK: Cambridge University Press.

Baxter, G. & Mislevy, R. J. (2004). The case for an integrated design framework for assessing science inquiry (*CSE Technical Report 638*). Los Angeles: The National Center for Research on Evaluation, Standards, Student Testing (CRESST), Center for Studies in Education, UCLA.

Brattström, L. and Wilcken, D. E. (2000). Homocysteine and cardiovascular disease: cause or effect? *American Journal of Clinical Nutrition, Vol. 72*, 315-323, (Retrieved July 28, 2007, from http://www.ajcn.org/cgi/reprint/72/2/315?ck=nck

Broek, P.V., & Kremer, K.E. (2000). The Mind in Action: What it means to comprehend during reading. In B. M. Taylor, M. F. Graves, & P. V. Broek (Eds.), *Reading for Meaning* (pp.1-31). New York: Teachers College Press.

Bruer, J. (1993). *Schools for Thought.* Cambridge, MA: MIT Press.

Doyle, A.C. (1887/1986). *A study in scarlet*. New York: Bantam Books.

Embretson, S.E. (1998). A cognitive design system approach to generating valid tests: Application to abstract reasoning. *Psychological Methods, 3,* 380-396.

Fillmore, C.J. (1976). Frame semantics and the nature of language. *Annals of the New York Academy of Sciences: Conference on the Origin and Development of Language and Speech, 280,* 20-32.

Gamma, E., Helm, R., Johnson, R., & Vlissides, J. (1994). *Design patterns: Elements of Reusable Object-Oriented Software.* Reading, MA: Addison-Wesley.

Grabe, W. (1999). Developments in reading research and their implications for computer-adaptive reading assessment. In M. Chalhoub-Deville (Ed), *Studies In Language Testing* (pp.11-46). Cambridge: Cambridge University Press.

Graesser, A. C. (1981). *Prose comprehension beyond word.* New York: Springer-Verlag.

Graesser, A.C., Gernsbacher, M.A., & Goldman, S. (2003) (Eds.). *Handbook of discourse processes.* Mahwah, NJ: Erlbaum.

Graesser, A.C., Olde, B., & Klettke, B. (2002) How does the mind construct and represent stories? In M. Green, J. Strange, & T. Brock (Eds.), *Narrative impact: Social and cognitive foundations* (pp. 229-262). Mahwah, NJ: Erlbaum.

Hannon, B. & Daneman, M. (2001). A new tool for measuring and understanding individual differences in the component processes of reading comprehension. *Journal of Educational Psychology, 93,* 103-128.

Illinois State Board of Education (2007). *Illinois Reading Assessment Framework Grades 3-8, State Assessments Beginning Spring 2006.* Springfield, IL: Illinois State Board of Education. Electronic version retrieved July 28, 2007, from http://www.isbe.net/assessment/pdfs/iaf_reading.pdf

Irwin, J. (2002). *Reading Strategies for the Social Studies Classroom.* New York: Holt, Rinehart and Winston.

Jackson, M. D., & McClelland, J. L. (1979). Processing determinants of reading speed. *Journal of Experimental Psychology: General, 108,* 151-181.

Just, M., & Carpenter, P. (1987) *The Psychology of Reading and Language Comprehension.* Boston, MA: Allyn and Bacon.

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

Kane, M. (2006). Validation. In R.L. Brennan (Ed.), *Educational Measurement* (4th ed., 17-64), Westport, CT.: American Council for Education/Praeger.

Kintsch, W. & van Dijk, T. A. (1978). Towards a model of text comprehension and production, *Psychological Review, 85,* 363-394.

Kintsch, W. & Yarbrough, J.C. (1982). Role of Rhetorical Structure in Text Comprehension. *Journal of Educational Psychology, 74*, 828-834.

Kintsch, W. (1988). The role of knowledge in discourse comprehension construction-integration model. *Psychological Review, 95,* 163-182.

Kintsch, W. (1998). *Comprehension: A paradigm for cognition.* New York: Cambridge University Press.

Lakoff, G. (1987). *Women, fire, and dangerous things: What categories reveal about the mind.* Chicago: University of Chicago Press.

Massachusetts Department of Education (2006). *Massachusetts Comprehensive Assessment System - Working with Massachusetts Curriculum Frameworks, Winter/Spring 2003.* Malden, MA: Massachusetts Department of Education. Electronic version retrieved July 28, 2007, from www.doe.mass.edu/frameworks/news/2003/wkshp.pps

Massey, D.D., & Heafner, T.L. (2004). Promoting reading comprehension in Social Studies. *Journal of Adolescent & Adult Literacy, 48*, 26-40.

McNamara, T. (1996). *Measuring second language performance: A new era in language testing.* New York: Longman.

Messick, S. (1989). Validity. In R.L. Linn (Ed.), *Educational Measurement* (3rd Ed.), (pp.13-103). New York: American Council on Education/MacMillan.

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

Minsky, M. (1975). A framework for representing knowledge. In P.H. Winston (Ed). *The psychology of computer vision* (pp. 211-277). New York: McGraw-Hill.

Mislevy, R.J., Chudowsky, N., Draney, K., Fried, R., Gaffney, T., Haertel, G., Hafter, A., Hamel, L., Kennedy, C., Long, K., Morrison, A.L., Murphy, R., Pena, P., Quellmalz, E., Rosenquist, A., Songer, N.B., Schank, P., Wenk, A., and Wilson, M.R. (2003). *Design Patterns for Assessing Science Inquiry. PADI Technical Report 1.* SRI International, Menlo Park, CA.

Mislevy, R.J., & Haertel, G.D. (2006). Implications for Evidenced-Centered Design for Educational Testing. *Educational Measurement: Issues and Practices, 25,* 6-20.

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. (2002). On the roles of task model variables in assessment design. In S. Irvine & P. Kyllonen (Eds.), *Generating items for cognitive tests: Theory and Practice* (pp.97-128). Hillsdale, NJ: Erlbaum.

Mislevy, R.J., Steinberg, L.S., & Almond, R. G. (2003). On the structure of educational assessment. *Measurement: Interdisciplinary Research and Perspectives, 1,* 3-62.

National Assessment Governing Board (2004a). NAEP: *Reading Framework for 2005 National Assessment of Educational Progress.* Washington, DC: U.S. Department of Education.

National Assessment Governing Board (2004b). *NAEP: Science Framework for the 2005 National Assessment of Educational Progress.* Washington, DC: U.S. Department of Education.

National Research Council (1998). *Preventing reading difficulties in young children.* C.E. Snow, M. S. Burns, & P. Griffin (Eds.), Committee on the Prevention of Reading Difficulties in Young Children. Washington, DC: National Academies Press.

National Research Council (2001). *Knowing what students know: The science and design of educational assessment.* Committee on the Foundations of Assessment, J. Pellegrino, R. Glaser, & N. Chudowsky (Eds.). Washington DC: National Academy Press.

New Zealand Foundation for Character Education. (2006). *Cause and Effect - The Law of Consequences* (Poem and Volcano Tasks). Electronic version retrieved May 16, 2006. http://cornerstonevalues.org/cause.html

Oregon Department of Education (2006a). *Reading/Literature: Test Specifications and Blue Print - 2005-2008, Grade 4.* Portland, OR: Office of Assessment and Information, Oregon Department of Education. Electronic version retrieved July 28, 2007 from http://www.ode.state.or.us/teachlearn/testing/dev/testspecs/asmtrdtestspecsg4.pdf

Oregon Department of Education (2006b). *Science Test Specifications and Blue Prints, Benchmark 3/grade 8, 2006-2007*. Portland, OR: Office of Assessment and Information, Oregon State Department of Education. Electronic version retrieved July 28, 2007 from http://www.ode.state.or.us/teachlearn/testing/dev/testspecs/asmtscitestspecsb3.pdf

Rayner, K., Foorman, B.R., Perfetti, C.A., Pesetsky, D., Seidenberg, M.S. (2001): How Psychology of Science Informs the Teaching of Reading. *Psychological Science in the Public Interest, 2*, 31-74.

Rayner, K. & Pollatsek, A. (1989) *The psychology of reading.* Englewood Cliffs, NJ: Prentice Hall.

Rumelhart, D. E. (1980). Schemata: Building Blocks of Cognition. In R.J. Spiro, B.C. Bruce, & W.F. Brewer (Eds), *Theoretical Issues in Reading Comprehension* (pp.33-58), Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers.

Rumelhart, D.E. (1975). Notes on a schema for stories. In D. G. Bobrow & A.M. Collins (Eds.), Representation and understanding: Studies in cognitive science. New York: Academic Press.

Santella, A. (2001). *The Lakota Sioux.* New York: Children Press.

Schank, R.C., & Abelson, R.P. (1977). *Scripts, plans, goals, and understanding.* Hillsdale, N.J.: Lawrence Erlbaum Associates.

Strauss, C., & Quinn, N. (1998). *A Cognitive Theory of Cultural Meaning*. Cambridge, UK: Cambridge University Press.

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

Tyson, A.S. (April 1, 2007) 'Pentagon Says Funding Delay Would Affect Rotations, Training, Repairs' in Washington Post, page-A4.

van Dijk, T.A., & Kintsch, W. (1983). Observation on the Status of Experimental Research on Discourse Comprehension in *Strategies of Discourse Comprehension* (pp.21-60). New York: Academic Press.

Vermont Institutes (2001). *Linking VT-PASS to Local Science Teaching and Assessment Through the VT-Pass Performance Task Template*. Electronic version retrieved July 28, 2007 from http://www.vermontinstitutes.org/assessment/linking.pdf.

Warren, W.H., Nicholas, D.W., & Trabasso, T. (1979). Event Chain and inferences in understanding narratives. In R.O. Freedle (Ed.), *New directions in discourse processing* (pp. 23-52). Hillsdale, NJ: Erlbaum.

Sponsor
The National Science Foundation, Grant REC-0129331

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

Subgrantees
University of Maryland
University of California, Berkeley. *Berkeley Evaluation & Assessment Research (BEAR) Center* and *The Full Option Science System (FOSS)*
University of Michigan. *BioKIDs*

SRI International