

Running head: SOME IMPLICATIONS OF EXPERTISE RESEARCH

Some Implications of Expertise Research for Educational Assessment

Robert J. Mislevy

University of Maryland

Keynote address at the 34th International Association for Educational Assessment (IAEA) Conference, Cambridge University, September 8, 2008.

1. Introduction

These are exciting times in the world of educational assessment; days of urgent demands, unprecedented opportunities, and tantalizing challenges: Demands for consequential tests in schools and nations, at larger scales and with higher stakes than we have seen before. Opportunities to assess learning viewed from a growing understanding of the nature and acquisition of knowledge. Opportunities to draw upon ever-expanding technological capabilities to construct scenarios, interact with examinees, capture and evaluate their performances, and model the patterns they convey. And challenges abundant, encapsulated in a single question: How can we use these new capabilities to tackle assessment problems we face today?

This presentation focuses on insights we can leverage from the study of cognition, particularly focusing on the nature and acquisition of expertise. Cognitive science continues to reveal the nature of human knowledge, learning, and reasoning. We begin with a simple quiz that brings some profound principles to light. We take a closer look at these principles in sections that review limitations and capabilities of human cognitive processing and the emerging “sociocognitive” perspective on cognition, and relate them to some of the key findings from expertise research.

Our ultimate interest, of course, is what this tells us about assessment. As we proceed, we will point out implications for developing better assessments, and illustrate them with examples from research studies and applied projects. Assessment can be a very complex subject, but for the lessons we need to draw, a quotation from Sam Messick (1994) lays out the essential ideas quite neatly:

[We] would begin by asking what complex of knowledge, skills, or other attribute should be assessed, presumably because they are tied to explicit or implicit objectives of instruction or are otherwise valued by society.

Next, what behaviors or performances should reveal those constructs, and what tasks or situations should elicit those behaviors?

We will conclude by considering the prospects for bringing the insights of cognitive research into educational assessment more broadly.

2. A Quiz

Let's start with a little quiz based on some classical results from cognitive research that make us aware of how we think. The reader can take Task 1 and Task 2 himself, so I will present the tasks first and give the answers, and what they tell us about cognition, a couple pages later. Task 3 is hard to do by yourself, so the reader can try it on a friend.

2.1 *The Questions*

Task 1: What is this a picture of?



Figure 1. What is this a picture of? (Photographer: RC James)

Task 2: Which cards need to be turned over?

Each of the cards shown in Figure 2 has a number on one side and a letter on the other side. Consider the rule “If there is a vowel on one side, there is an even number on the other.” Which cards do you need to turn over to make sure the rule is not violated?

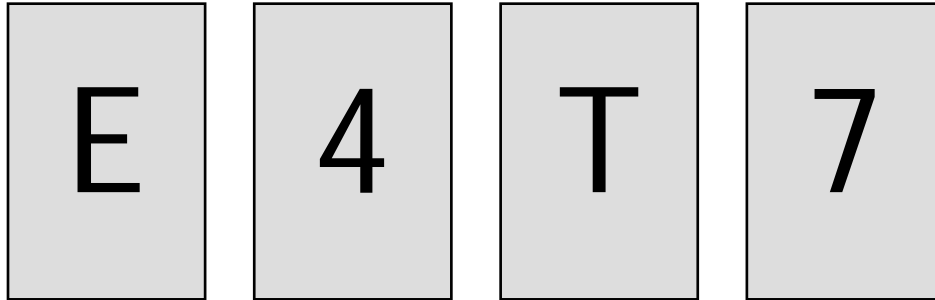


Figure 2. Which cards need to be turned over?

Task 3: A Little Story

Read the following story aloud to the subject:

The members of the Smith family exchanged gifts with one another. Mary gave Ed a stuffed hippo. Ralph gave a toy car to Tina. Harold received four gifts. It was the holiday season at last.

Chat with the subject a few minutes. Then ask her whether each of these three sentences, presented one at a time in writing, was in the story she heard:

1. Mary gave Ed a stuffed hippo.
2. Harold received ninety gifts.
3. Ralph gave Tina a toy car.

Finally, ask this question:

There were five sentences in the story. In order, what were the second-to-last words in each of the sentences?

2.2 The Answers

Task 1: What is this a picture of?

Figure 1 is a picture of a dog – specifically, a Dalmatian, sniffing the ground. Figure 2 points out his nose, his ear, and the curve of his back. If you didn't see the dog at first, you may have felt it “click” into place when you finally do. There are a number of interesting things about this.

First, once you see it, it is easy to see again. The picture hasn't changed at all, but you have. This phenomenon drives home how what we perceive depends partly on sensations the world presents us, but also on what we bring to the occasion. Second, if you had never seen a dog or similar animal, you would not see him here, perhaps even with lots of help. Third, once you know the dog is there, you can "see" where his left back leg is, even though there is hardly any information in the picture itself about where the lines would be. This perception comes almost entirely from what you already knew about dogs. It is a pattern you built up from many experiences, and you bring it to bear to make sense of this ambiguous and noisy picture.

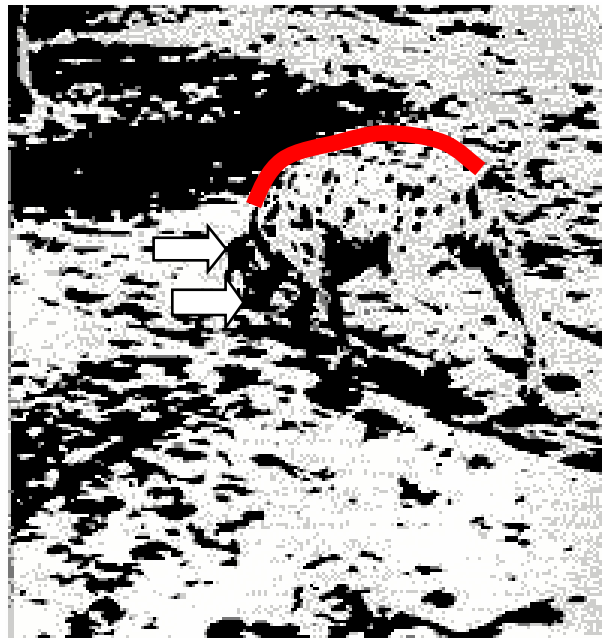


Figure 3. It's a Dalmation!

Task 2: Which cards need to be turned over?

Let's re-express the cards and the rule from Task 2. Suppose every card represents a customer in a restaurant, and every customer is characterized by their age and what they are drinking. The legal age for drinking alcoholic beverages is 21. If we re-label the cards from Task 2 so that "vowel" becomes "under 21," "consonant" becomes "21 or over," "even number" is "non-alcoholic beverage," and "odd number" is "alcoholic beverage," we obtain Figure 4. Whose identification do we need to check to make sure the law is not being violated?

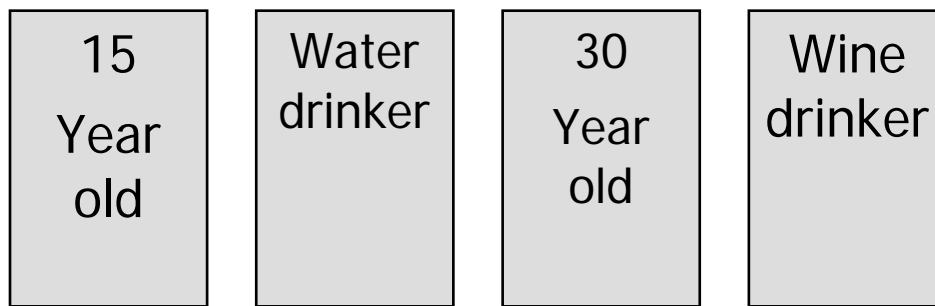


Figure 4. Whose identification do we need to check?

Even though the transformed problem is logically equivalent to the original problem (named the “Wason task,” after its originator, psychologist Peter Wason), most people have little difficulty determining that we need to check the beverage of the 15-year old and the age of the wine drinker. In its original form, subjects often forget to check the wine drinker or unnecessarily check the age of the water drinker. These errors are consistent with what is called a “confirmation bias.”

The transformed problem is formally identical problem, but it’s easier because it is no longer an exercise in abstract logic. It is reasoning in a familiar social situation, for which we have a ready schema. It’s called the permission schema, and we use it in legal systems, baking recipes, and hundreds of additional ways in everyday life. Even three-year olds understand it, in situations they operate in.

I call two points to your attention. First is the situated nature of cognition. Reasoning is not simply abstract and formal, although as we will see, abstract and formal thinking clearly do play an important role in expertise and in assessment. Rather it is a combination of reasoning through formal models as they are linked to understandable, human, story structures, which are linked in turn to the particular and unique situations we live in.

Second, notice that our understanding of the situation – the way that we encode the relationships in the situation – connects directly to strategies for action. Once we understand the problem as a familiar social situation that accords with the permission schema, we know whose IDs to check.

Task 3: A Little Story

Which sentences were present in the story?

1. Mary gave Ed a stuffed hippo. YES
2. Harold received ninety gifts. NO
3. Ralph gave Tina a toy car. NO

Why do many people miss the last one? For the same reason we usually find the first two easy. Our memory is not a tape recording in our head, but rather the trace of our processing, namely, a constructed understanding of a familiar kind of event.

The first sentence did happen to be in the story, and we say ‘yes’ because we remember not so much the words but the incident. Incidentally, most people can tell you what color they think the stuffed hippo was. Like the Dalmatian’s leg, this is information we bring to the party, supplied automatically by our accumulated personal experiences with actual events, things we’ve read or seen on television, and so on.

The second sentence wasn’t in the story, and even though it differs by just one word, that word – ninety rather than four—makes for a very different kind of event, something notably discrepant from the situation we were constructing in our minds to understand the story.

The third sentence was not in the story either, but this is a hard one to catch. A sentence that did appear in the story, “Ralph gave to Tina a toy car,” has a different surface structure but describes the same situation. What we remember is based on the understanding we constructed rather than a literal facsimile of the sensations; unless we are especially attuned to the details of surface structure, they usually fade quite quickly. This phenomenon accounts for discrepancies we find among witnesses to the same event, and is only now being addressed in the conduct of judiciary proceedings.

There were five sentences in the story. In order, what were the second-to-last words in each of the sentences?

The answer is: One, stuffed, to, four, holiday. Hardly anyone gets this right. The information is clearly present in the sensory input, but we don’t encode it for this particular (and unusual) feature, or a thousand other ways we might parse and encode what we hear. We understand it and remember it in terms of some human scenario. In

any situation we are bombarded with sensory input, only a fraction of which we can attend to. What we perceive is closely linked with what we understand, indeed what we can understand. This determination also depends in part on our goals, beliefs, and emotions at the time of experience. And what we understand is closely linked with what we do next.

3. A Closer Look at Cognition

Summary test scores, and factors based on them, have often been thought of as “signs” indicating the presence of underlying, latent traits. ... An alternative interpretation of test scores as samples of cognitive processes and contents, and of correlations as indicating the similarity or overlap of this sampling, is equally justifiable and could be theoretically more useful. The evidence from cognitive psychology suggests that test performances are comprised of complex assemblies of component information-processing actions that are adapted to task requirements during performance. (Snow & Lohman, 1989, p. 317)

3.1 Limitations and Capabilities of Human Cognitive Processing

Research areas within cognitive psychology today include perception and attention, language and communication, the development of expertise, situated and sociocultural psychology, and neurological bases of cognition. All share two complementary premises. The first is that people are essentially alike in terms of cognitive machinery and processes. The second is that within broad ranges set by our cognitive machinery and the nature of the world, the contents of learning and thinking are shaped by culture.

This human endowment entails both remarkable capabilities and surprising limitations. Examples of our capabilities include learning to use language as toddlers; apparently unlimited long-term memory capacity; and rapid and effortless recognition, judgment, language use, and coordinated bodily action. We are equally subject to sensory limitations, biases in reasoning, working memories of only about seven chunks of information, and unawareness of most of our own cognitive activity. We will call these *processing limitations*. They are wired in and universal.

We are also subject to what we will call *knowledge limitations*. Surveying the literature on expertise, psychologist Tim Salthouse listed a number of recurring difficulties that people new to a domain routinely encounter:

- Not knowing what information is relevant
- Don't know how to integrate information

- Not knowing what to expect
- Not knowing what to do and when to do it
- Lack of production proficiency

These pervasive bottlenecks are central to our concern. An expert in a domain is somebody who can deal with these problems. She became an expert despite her processing limitations by learning things. She takes advantage of the remarkable capabilities we do as humans possess by organizing traces of experiences that can be put to use in complex and fast moving situations despite our unavoidable processing limitations. Acquiring expertise is capitalizing on the kinds of learning we are good at, to get around what we are not good at. Patterns and associations are the key: building up effective ones, using some until they become automatic, being able to assemble them in flexible ways that suit an evolving situation, learning how and when to use tools that have been invented to support our work. Accordingly, it is the features of situations that stump novices but experts navigate that provide us with insights about how to construct assessment tasks.

3.2 *A Sociocognitive Perspective on Cognition*

In the introduction we noted exciting advances into the nature of knowledge, and how people acquire it and use it. A synthesis has begun to emerge across fields as varied as psychology, neuroscience, anthropology, linguistics, philosophy, and computer science, which in broad strokes integrates research on individual cognition with field studies of social interaction and laboratory investigations of cells, bodies, and brains. Linguist Dwight Atkinson (2002) calls it a “sociocognitive” approach, to emphasize the interplay between the patterns in the physical world and the social world to which we become attuned, and the patterns we develop and employ internally. Bristol University’s Guy Claxton (2001) calls our attention to “the intricate complexity of the unique moment in which a person interacts with an unprecedented material, social and cultural setting.”

The next section show how these ideas play out in one particular area in which much progress has been made recently, namely reading comprehension. In simplified form, Walter Kintsch’s CI model illustrates the key themes, and enables us to extend the paradigm more generally. This view of cognition will set the stage for a discussion of expertise.

3.3 Kintsch's Model for Reading Comprehension

It is an everyday occurrence for Riva to read a newspaper article to find out what happened in the football game last night, but what is happening inside her head as she works her way through the article? The following series of figures illustrates the process in terms of Walter Kintsch's Construction-Integration (CI) model for reading comprehensions.

Figure 5a shows the *surface text* on the left, the sensation the world provides to Riva. For it to have any chance of making sense, the writer depended on many patterns and conventions that had developed over hundreds of years through the interactions of billions of people, in the form of the letters, the syntax, and the words of the language itself. As Riva begins to work through the text (the rectangle represents her focus), she constructs from the surface text a representation of the propositions that the sentences on the page portray.¹ The propositions are not the same as the sentence; remember the difference between "Ralph gave Tina a toy car" and "Ralph gave a toy car to Tina." Kintsch calls the result the *text base*. His research, and that of many others, tells us that our working memory of the surface text fades within seconds after we construct the text base. Furthermore, skilled readers construct pretty much the same text base.

But constructing the text base is not comprehension. As we move through the text, our understanding will be constructed jointly from the information in the text base and the information we bring to the experience. Figure 5b shows that the elements in the text base activate elements, or patterns, from long term memory (LTM). This is the construction, or C, phase of Kintsch's CI model. These are not lists of propositions from a storage library, but patterns of concepts, impressions, feelings, traces of past experiences, and so on, all kinds of bits and pieces that resonate in some way with the input. These activations are automatic; they are based only on associations, not necessarily sensible, meaningful, or useful. If we see the word "bank" in a text, then associations with river banks, banks for financial transactions, and turning airplanes can all be activated. (Each of them is in turn associated with lots of other patterns in LTM,

¹ Kintsch's main interest begins after the processing of photons as patterns of light and dark, recognition of letters, patterns of letters forming words, recognizing the words, and identifying the words as phrases in a language in which the reader has at least a modicum of skill. There is a lot going on here, to be sure, but Kintsch focuses on the largely (but not entirely) subsequent processes of comprehending this text. Some of the kinds of processing Kintsch addresses in the CI model are taking place at these lower levels as well, but for skilled readers the processes at this earlier stage have become automated. We note in passing that learning to read in a second language brings the focus back to these processes.

and we will see that the nature of these associations is one of the primary ways experts differ from novices.)

Note that the Figure 5b shows the activated elements as being influenced by context as well as by the text base. The physical and social context, what we have been working on, our purpose for reading, even the time of day, can influence the likelihood of elements being activated. Subjects in a psychology experiment were shown a picture of a boy getting a box from table. When they were asked what might be in the box, subjects tested just before lunch were more likely to say there was food in the box than subjects tested after they had eaten.

The Integration, or I phase, of the CI model concerns the coincident patterns among the activated elements. Those that are compatible with others remain active. Those that don't will fade quickly, in less than half a second. The elements that do resonate with others will persist, and their integration constitute what Kintsch calls the *situation model*—effectively, Riva's understanding of the text at this point. It is comprised of both information from the text and information from Riva's past experiences. Different readers with different experiences can form substantially different situation models of the same text. For example, I understand perfectly well the propositions in this excerpt from Alan Pollack's analysis of the Beatles' "Strawberry Fields Forever":

On paper, it doesn't look so far out, but do you hear the opening chord as V, especially when the a-minor chord is implied in the second half of the measure? Similarly, toward the end of the phrase I hear the B-flat chord as IV of F and expect F to be the home key only to be fooled by that sort of forced 6/4 -> 5/3 plagal cadence at the end. Note, by the way, how the final measure of the intro contains an additional two beats!²

As a non-musician, however, I cannot connect the noun phrases connect with patterns of sounds, patterns of musical notation, or lived experiences of music I have written, heard, or played. My situation model is bare and impoverished in comparison with Mr. Pollack's. No one else's situation model will be exactly like his, but people who have studied music will have richer models that are similar to his in more ways to his than mine was. The phrases activate experiences from their past, using those terms and concepts to describe musical experiences that enable stronger similarities between their situation models and Pollack's—similar in the ways he intended as he wrote them.

² Downloaded from <http://www.icce.rug.nl/~soundscapes/DATABASES/AWP/sff.shtml>, August 10, 2009.

Riva continues to read, and the process continues as in Figure 5d (although more continuously and less discretely than the picture suggests). Note here that in addition to context and the text itself, the preceding state of the situation model also contributes to the construction of the next state.

Elements in LTM that have been activated and remain active in a situation model have their own associations with other elements of LTM, which can be pursued as well. Figure 5e shows Riva acting in a way that makes sense in light of her understanding of the text and the purpose of her reading. Figure 5f shows that this action changes the context of the situation, which in turn influences Riva's successive understanding and possibilities for further action. Think here of the interplay among the text, the person, and the context when we assemble a barbecue grill from the instruction manual.

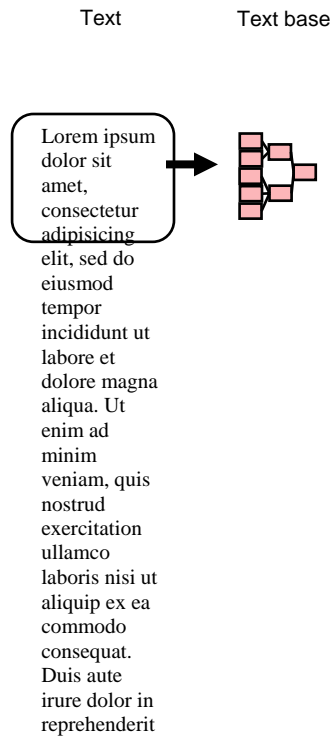


Figure 5a. Text to text base.

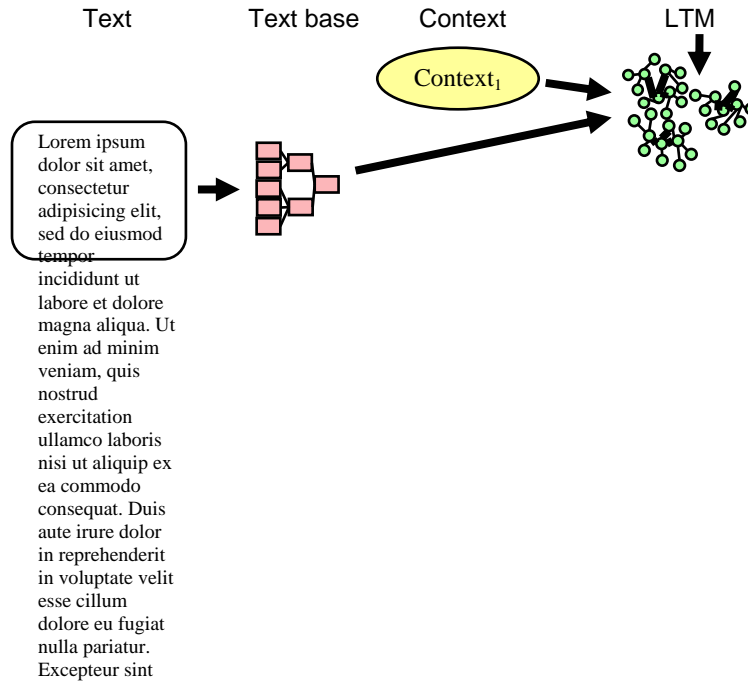


Figure 5b. The “Construction” phase.

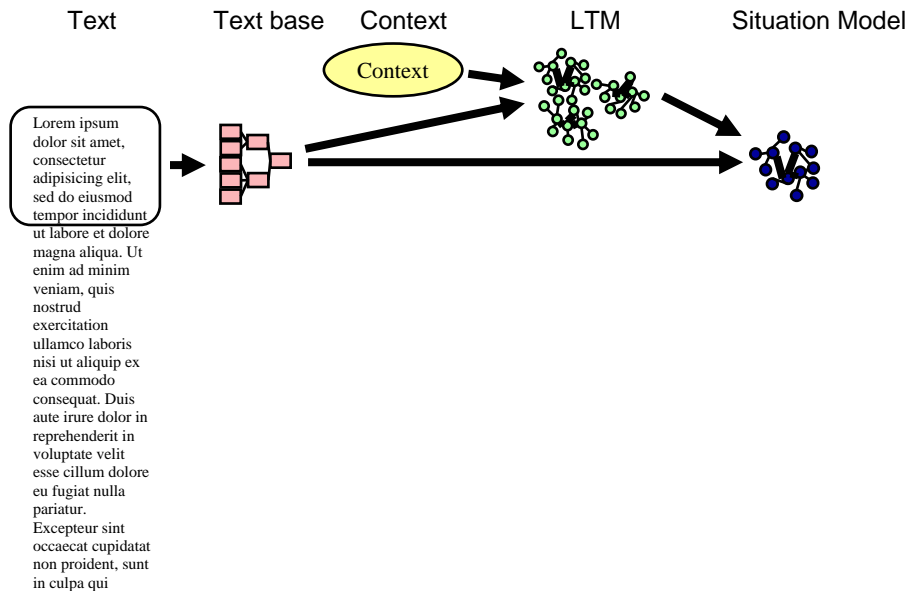


Figure 5c. The “Integration” phase.

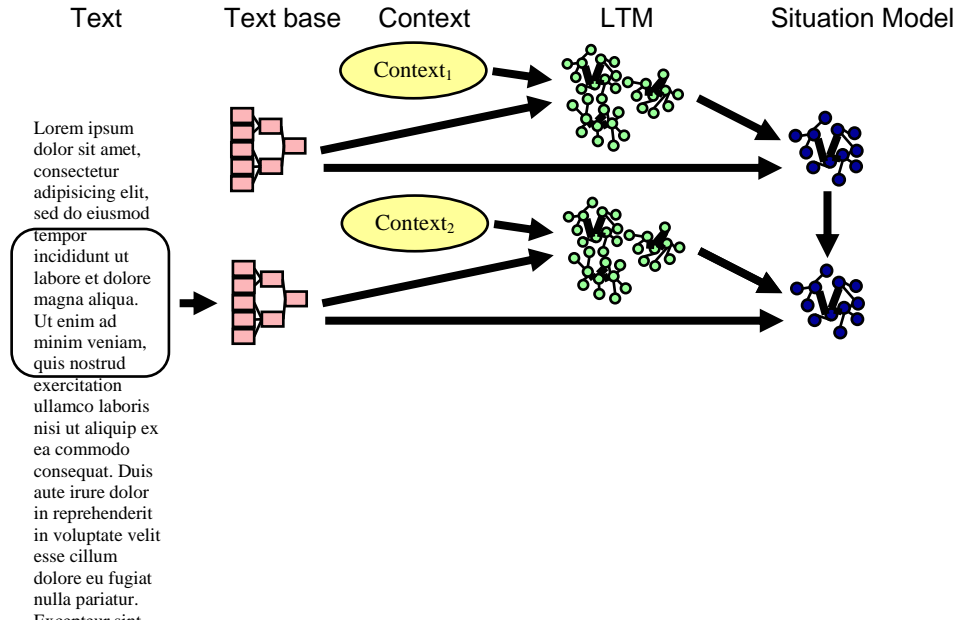


Figure 5d. The process continues.

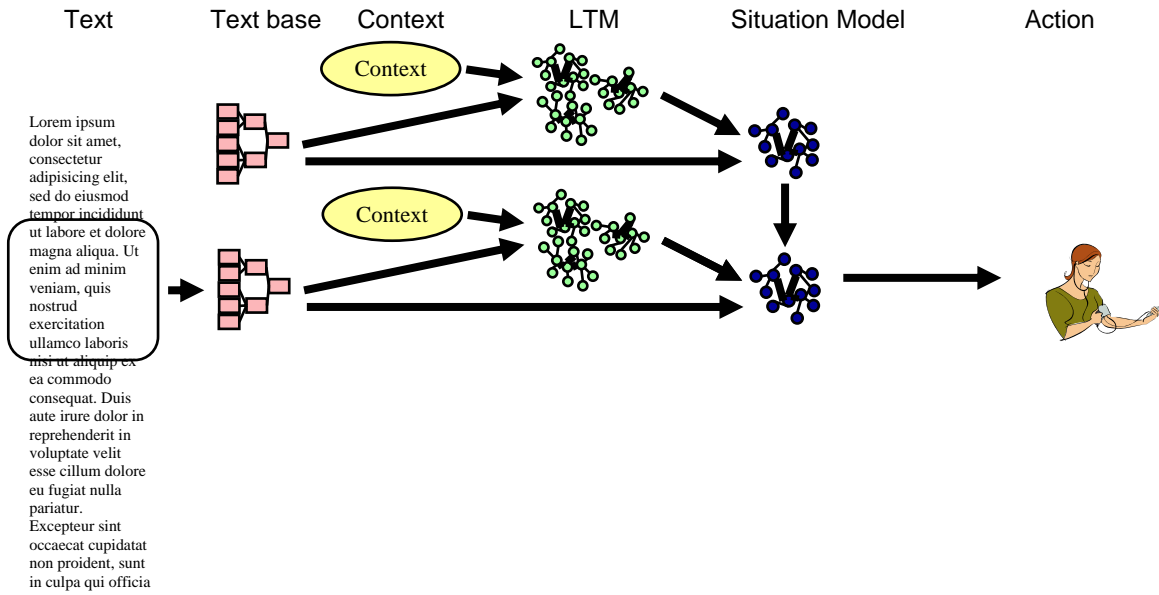


Figure 5e. An action based on the situation model.

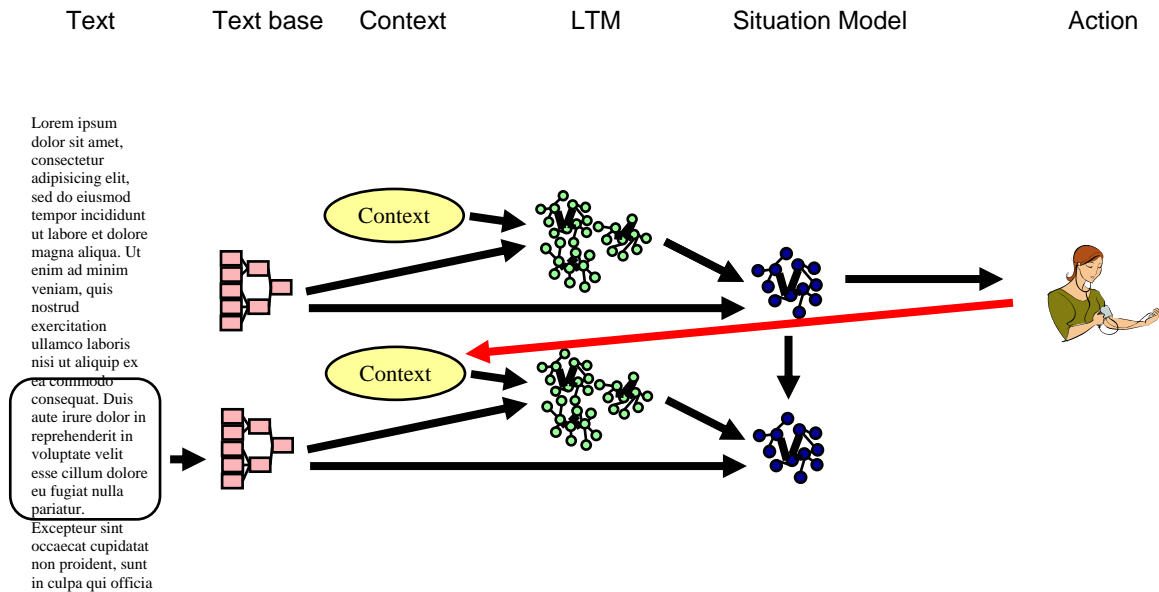


Figure 5f. The action changes the context in which the text is understood.

3.4 Extending the CI Paradigm

Although Kintsch's focus was reading comprehension, he argues that the essential nature and many of the processes in reading characterize comprehension more generally; for example, when we watch movies, prepare meals, have conversations, engage in work and play, and negotiate situations in day to day life. The elements from our long term memories that are activated are at many different levels of granularity, from the shapes of printed letters and the way vowels sound when they are preceded by different consonants, to patterns of physical movement and balance needed to ride a bicycle, to how grammatical constructions signal relationships in time, to how to order food in a restaurant, to what it means to be married or to be sick in our culture, to musical and mathematical notation, to our relationships with our family members.

In one sense every one of these elements is highly personal, arising as traces of the unique lived experiences of an individual human. On the other hand, many share key features across people for a number of reasons: We are similarly-constructed beings. We live in the same physical world that shapes our experiences and provides metaphors for more abstract relationships (e.g., Lakoff & Johnson, 1980). We exist in social groups, interacting in patterns with others that structure our actions and give meaning to them (Gee, 1992). Patterns of interaction structured in terms of, for example, language, symbol systems, customs, procedures, and institutions provide not only experiences with

shared elements but ways of perceiving them, thinking about them, acting in them, communicating with others about them, and contributing to them. It is in these terms we can think about the nature of expertise.

4. Key Findings from the Expertise Literature

In most domains of knowledge, we develop very powerful theories when we are very young. School and the disciplines are supposed to reformulate those theories and to make them more comprehensive and more accurate. As long as we stay in school, we can maintain the illusion that the effort has succeeded, but ... once we leave school, the illusion disappears and there is a 5-year-old mind dying to get out and express itself...

No one has to tell a kid that heavy objects fall more quickly than light objects. It's totally intuitive. It happens to be wrong. Galileo showed that it was wrong. Newton explained why it was wrong. But, like others with a robust 5-year-old mind, I still believe heavier objects fall more quickly than lighter objects....

The only people on whom these engravings change are experts. Experts are people who actually think about the world in more sophisticated and different kinds of ways. ... In your area of expertise, you don't think about what you do as you would when you were five years of age. But I venture to say that if I get to questioning you about something that you are not an expert in, the answers you give will be the answers you would have given before you had gone to school. (Gardner, 1993, P. 5)

A surprising finding from the early research on expertise, starting with Dutch psychologist (and master-level chess player) Adriaan de Groot's (1946) studies in the domain of chess, was that experts in domains are not very different from anyone else in terms of scores on familiar mental abilities tests. Rather, they differ in what they know, how it is organized, and how they are able to use it (Ericsson, et al., 2006). In the terminology of Section 3, they carry out the same kinds of processing, in the same ways reflected by the CI model, under the same limitations, as the rest of us. It is the elements that they can bring to bear from long term memory, which they have developed typically over years of reflective practice, that make them different.

Drawing on this sociocognitive perspective, this section looks at four important aspects of expertise:

- Organization of knowledge
- Knowledge Representations
- The Importance of Interaction

- Social Aspects of Expertise

We will see how the nature of the elements of knowledge and their associations enable experts to overcome the roadblocks Salthouse pointed out, and what these findings suggest about assessment design.

4.1 Organization of Knowledge

Chi, Feltovich, and Glaser (1981) asked novices and experts in physics to sort cards depicting mechanics problems like those shown in Figure 6 into stacks of similar tasks. They found that novices grouped problems in terms of surface features such as pulleys and springs. Experts organized their stacks in terms of more fundamental principles such as equilibrium and Newton's Third Law, each stack containing a mixture of spring, pulley, and inclined plane tasks.

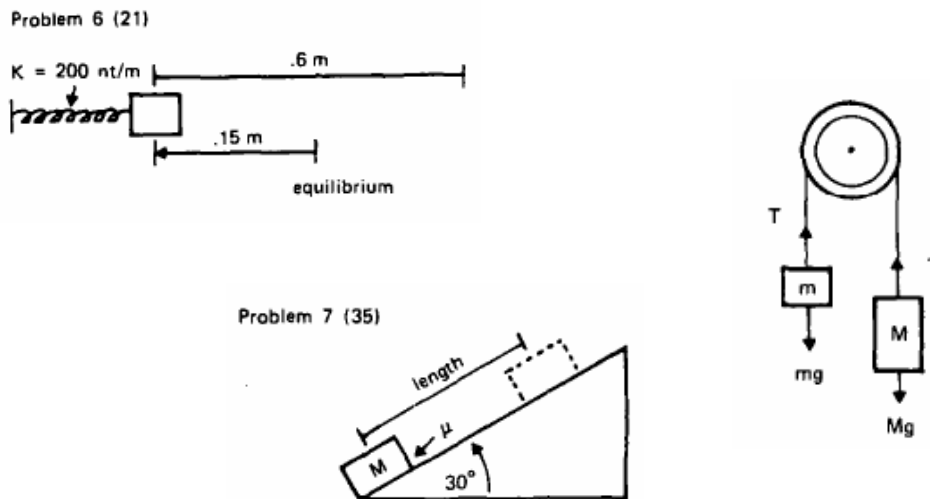


Figure 6. Some of the tasks from the Chi, Feltovich, and Glaser (1981) study.

The internal processes of both experts and novices were essentially the same in terms of constructing a situation model. The experts, however, activated elements the novices simply didn't have—schemas that are not at all obvious on the surface, revolving around hard-won insights by scientists over hundreds of years. These schemas, after years of study and experience, could be activated by features of seemingly unrelated physical situations that could be perceived only through their lenses. The experts saw physics Dalmations; the novices only saw spots.

Experts also differ from novices by being able to solve problems like these, by decomposing forces, chaining equations, solving for unknown variables, and so on. But in another study Larkin (1983) found that before they lay a problem out formally, the physicists first constructed a “story” for the real world situations in terms of physics principles. While they ultimately used symbol manipulation to work out the details of a solution, the narratives built around underlying principles that linked the symbolic representations to the particulars of real world situations, and guided the course of the procedures. Just as when our recognition of a permission schema fits a situation we encounter, *meaning* intertwines with *perception* on the one hand and effective *action* on the other. In this case, the associations are organized around physics principles, connected to strategies and procedures, available whenever patterns in unique real-world situations activate them.

Having developed elements of long-term memory around deep principles of physics, having built through experience effective actions to tackle problems, and having attuned their perceptions of real-world situations to patterns that activate physics schemas, these physics experts show us how acquired knowledge breaks through the bottlenecks. What information is relevant? That which accords with schemas as they are assembled to tackle the problem at hand. How should the information be integrated? In terms of the schemas, as they are assembled with one another and related to the real-world problem. What should be expected? What the models of processes and relationships would indicate, as it plays out in this unique situation. What should be done? Strategies are available as activated links from the models assembled to describe the situation. Will production be proficient? Yes, if there has been enough experience with problems in the domain for working through solutions, monitoring progress, reformulating when necessary, and so on, as actions successively move the real-world situation closer to solution.

These findings emphasize for us the importance of the interplay between the structured and abstracted knowledge that is (appropriately) addressed in schooling and the unique situations in which this knowledge is useful. To make the broad and generative power of formal structures available to students, learning experiences should illustrate them in multiple settings and make explicit the correspondence between formal structures and particular situations. One should not think of assessing, say, students’ “proportional reasoning proficiency” solely by means of facility with formal structures. A richer space of claims addresses proficiency with formal structures, proficiencies in activities that correspond to proportional reasoning in relevant contexts, and awareness of

correspondences among contexts and between contexts and formal representations. Note that the Chi et al. study used familiar problems—straight application of physics formulas to textbook exercises—but in an unfamiliar way: It used them to reveal whether subjects' were thinking about them in accordance with the underlying principles. We need ask what is important in nonstandard situations and why, what will happen next or what should be done, in the ways indicated by the models and principles that are the target of learning.

4.2 *Social Aspects of Expertise*

The first generation of expertise studies, including De Groot's studies of chess masters and the physics research discussed above, were carried out under the information-processing perspective. The focus was on the internal representations and problem solving processes of individuals. This perspective is complemented by an ethnographic approach that studies behavior in more naturally occurring settings, with particular attention to use of tools, knowledge representations, and socially-structured patterns of interaction. Examples include Scribner's (1984) study of dairy workers and Lave's (1988) analysis of how people use arithmetic reasoning not formally, but embedded in practically useful schemas, when they shop for groceries. Much of the information required by experts resides in the environment and situation as well as in their heads; cognition is both supported, in terms of tools and knowledge representations, and distributed, in terms of the interpersonal interactions required to accomplish a goal. Current work in expertise research integrates both perspectives, as both are indeed integral to expertise. Examples include studies of nurses in emergency rooms (Patel & Kaufman, 1996) and members of an airline flight crew (Hutchins & Klausen, 2000).

I was member of a team of researchers from the Dental Interactive Simulation Corporation (DISC), Chauncey Group International, and Educational testing Service that carried out a cognitive task analysis for to design task templates and scoring methods for a simulation-based performance test in dental hygiene. While existing multiple-choice content tests and hands-on tests of skills provided good evidence of these aspects of proficiency for licensing decisions, they provided less information about the knowledge hygienists use when they assess patients, plan treatments, and monitor progress. To investigate the nature of this knowledge, we compared groups of approximately ten hygienists each at the levels of expert, recently-licensed, and student novice, as they talked through eight hypothetical cases that represented important and recurring challenges.

The researchers and a team of experts recruited by DISC met to analyze the protocols. The mission was to abstract, from specific actions of individual subjects in particular cases, general characterizations of patterns of behavior—a language that could describe solutions across subjects and cases not only in the data at hand, but in the domain of dental hygiene decision-making problems more broadly. In line with the goal of assessment, the group sought patterns that would be useful in distinguishing hygienists at different levels of competence. We referred to the resulting characterizations as *performance features* (Cameron, et al., 1999; Mislevy, et al., 1999).

Not surprisingly, some of the performance features related to the organization and use of knowledge as described in the previous section. For example, one dealt with interpretation of multiple forms of information about a patient, such as radiographs, hard and soft tissue charts, and probing depths. The novices were usually able to identify important cues in these representations—they did see the Dalmatians—but the striking difference was how the experts summarized the information. They produced not simply lists of cues, but integrated representations of a patient’s underlying condition, which would plausibly project down to the views each representation showed. The experts could thus reason through the “stories” of various dental conditions. While the novices has reached a stage in which cues in the stimuli appropriately activated schemas for recognition and interpretation at the level of the stimulus forms, the experts were further able to associate patterns of multiple cues with etiologies.

Beyond the performance features associated with information-processing capabilities, we found several ways that experts differed consistently from novices that highlighted the social and contextual nature of the role of a dental hygiene professional. We note two of them, and another aspect of the social nature of practice that held implications for assessment design.

Tailoring language and vocabulary consistently distinguished experts and novices. Even when novices were correct about the substance of a point that needed to be communicated, they were less apt than experts to use terminology and conversational patterns that best suited to a particular interaction; overly technical explanations to patients, for example, or colloquial and incomplete communication of findings to other dental professionals. In contrast, the experts would note a patient’s reactions and questions to a comment, and adjust their tone, their terminology, and their depth of explanation in response. Figure 7 elaborates the CI sequence to suggest the iterative and joint model-building that occurs when two people converse. The experts’ situation models included a model of the patient that concerned the nature and level of explanation

they would understand and be comfortable with. This capacity allowed them to build their explanations in ways that cues suggested would help the patient build an appropriate model, which would be quite different than those they would use for a dentist, who would need to build a technical situation model to support her work.

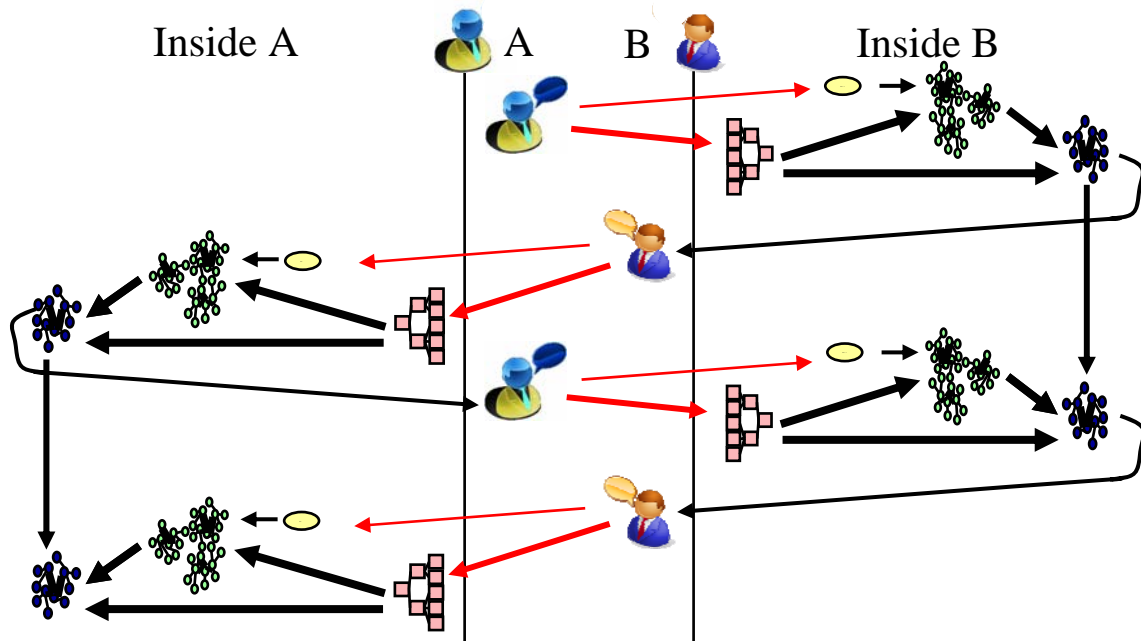


Figure 7. A look behind conversation from a CI perspective.

Scripting behavior describes a phenomenon the team saw at several phases of subjects' interactions with clients, including assessment, treatment planning, and patient education. A context, a situation, or a cue can activate a script. Novices learn to follow sequences of actions as "chunks", and carry them out with minimal tailoring to individual clients' needs. This is manifest when a novice produces a standard treatment plan or a patient education lecture with elements that are irrelevant or inappropriate for a particular client. For experts, knowledge is more finely variegated. Their densely interrelated web of smaller sub-themes and patterns can be flexibly assembled in response to individual clients and evolving interactions. The experts' resulting situation models better take into account contextual features and integrate them with substantive knowledge per se.

The social nature of what it means to be a dental hygienist also came into focus when we thought about how to capture in an assessment context the kinds of thinking that a candidate does. We saw that much of the difference between experts and novices was in the sense they made of cues, hypotheses they developed, treatment plans or additional

diagnostics they felt would be needed, and expectations of what to look for in a follow-up appointment. Each of these kinds of situations would be good bases for developing focused tasks, whether multiple-choice or open-ended. But how could this thinking be captured in the more continuous, interactive format of a simulated working through of a case?

Our recommendation provided at once good evidence from the perspective of assessment and high fidelity to actual practice: It was to have the candidate fill in a simplified standard dental insurance form. Not only do such forms capture key results of an examination or treatment plan and its rationale, they are integral parts of practice today. They are built around key features of interactions with patients and the underlying substance, and constitute a central way that actors in a widely distributed system communicate with one another. No matter how much you know about dental science and no matter how skilled you are with procedures, you cannot be a functioning hygienist without being able to acquire, use, and produce information in collaboration with others using these particular forms of knowledge representation.

4.3 *Knowledge Representations*

While many cognitive psychologists study the ways that knowledge is represented inside peoples' heads, researchers interested in expertise and in the social nature of knowledge are keenly interested in external forms for representing knowledge. Examples include maps, graphs, wiring diagrams, bus schedules, written language, mathematical formulas, object models for business systems, and insurance forms in dental offices. Good knowledge representations take information and relationships that are otherwise hard to think about (too much, too complicated, too fast, too small), and express it in forms that play to our strengths. They help us overcome human processing limitations in several ways: They supplement both limited working memory and faulty long-term memory over time and in volume. They coordinate action among people. They provide common ways of thinking about things of common interest—and they are especially powerful when those ways of thinking are not obvious and have been hard won through the work of many people over considerable time.

Knowledge representations thus connect the information-processing and sociocultural perspectives on expertise. Learning to work with knowledge representations of a domain—to perceive, express, manipulate, transform, explain, and share information—is essential to becoming an expert in that domain. The structures around which knowledge representations are organized come to be reflected in learners' internal representations.

They come to view real-world situations in these terms, much as when we recognize a situation as an instance of the permission schema. The tools, the procedures, and the strategies associated with knowledge representations become activated and available for action as these connections are developed. In all, this familiarity with the knowledge representations of a domain helps us overcome Salthouse's knowledge bottlenecks.

Well constructed representations build on fundamental principles in a domain, in an abstracted form. They gain power when we use them to understand real-world situations in their terms, as shown in Figure 8. The real-world situation in the lower left is viewed through the perspective of the knowledge representation. Part of becoming an expert is learning the range of situations in which this can be done and how to do it. Formal ways of thinking and acting within the representational space takes place in the middle plane, and again part of becoming an expert is developing this facility. At the top of the figure are other representational forms that are connected with the primary one, which to experts are associated with it and can be activated to re-represent a problem—moving back and forth between diagrams, equations, and graphs, for example. At the lower right is the reconceived real-world situation, seen through the eyes of the representation.

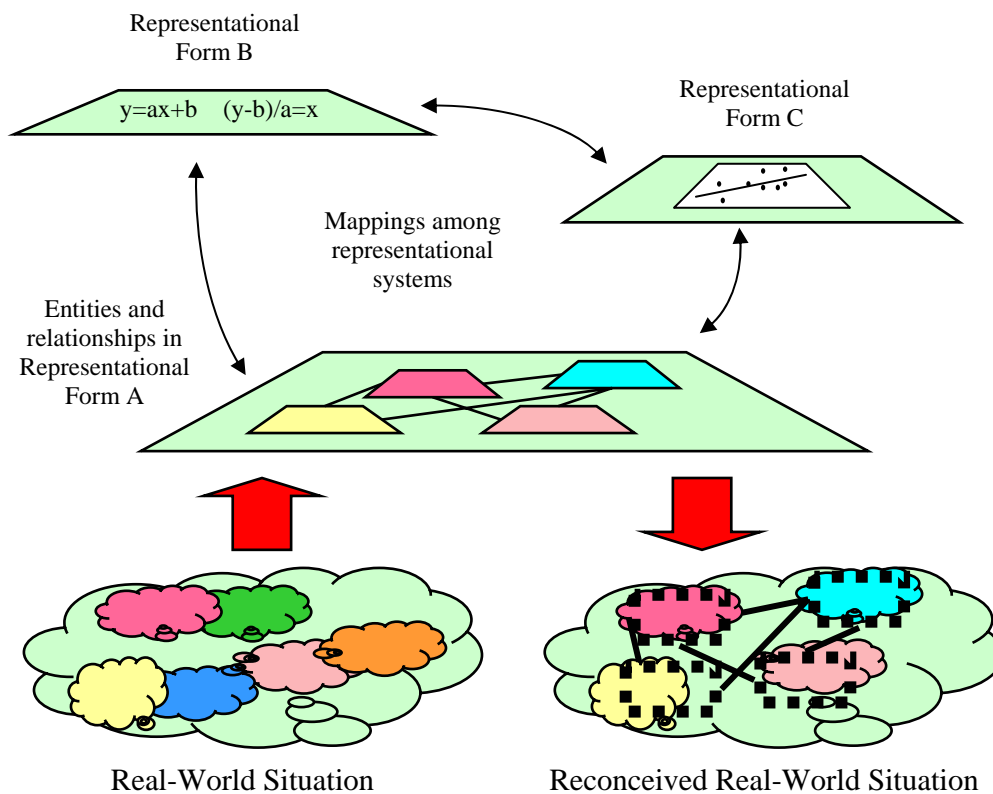


Figure 8. Knowledge representations.

Knowledge representations play a central role in learning in any domain of school or work, and correspondingly in assessment for those arenas (Gitomer & Steinberg, 1999). Recalling Messick's guiding questions for assessment design, when we ask what situations can provide us with evidence about a person's capabilities we know that information provided in one or more forms of knowledge representation will usually be an essential part of those situations. When we ask what we need to see people say, do, or make, it will often be producing, transforming, or building knowledge representations in those domains.

An example of an assessment project in which knowledge representations play a central role is the National Council of Architectural Registration Boards' Architectural Registration Examination (ARE). In 1997, the ARE became the first fully computerized licensure test to incorporate automated scoring of complex constructed responses in operational testing (Bejar & Braun, 1999). ARE tasks require a candidate to design a building according to a list of requirements and constraints, and execute architectural drawings to detail the solution using computer-aided design (CAD) technology. CAD tools and representations are today integral to being an architect, in much the same way that insurance forms are integral knowledge representations in medical professions. Figure 9 shows one of the problem statements with its list of constraints, and Figure 10 shows a solution in the candidate's work space.

Irv Katz's (1994) studies of expert and novice architects' design solutions helped ground a family of CAD design problems for the ARE. He found that the design process was invariably iterative: experts and novices alike started from an initial solution that met some constraints, and modified it repeatedly to accommodate more, always working from the representation generated thus far. But while both experts and novices continually revised aspects of provisional designs as they progressed, the novices' rework was more often substantial and discarded much previous work. The novices had encountered conflicting and hard-to-meet constraints when they were further along, whereas the experts identified and addressed these challenges early on. Varying the number of constraints, the challenge of meeting them, and the degree of conflict among them are systematic and cognitively relevant ways to vary task difficulty.

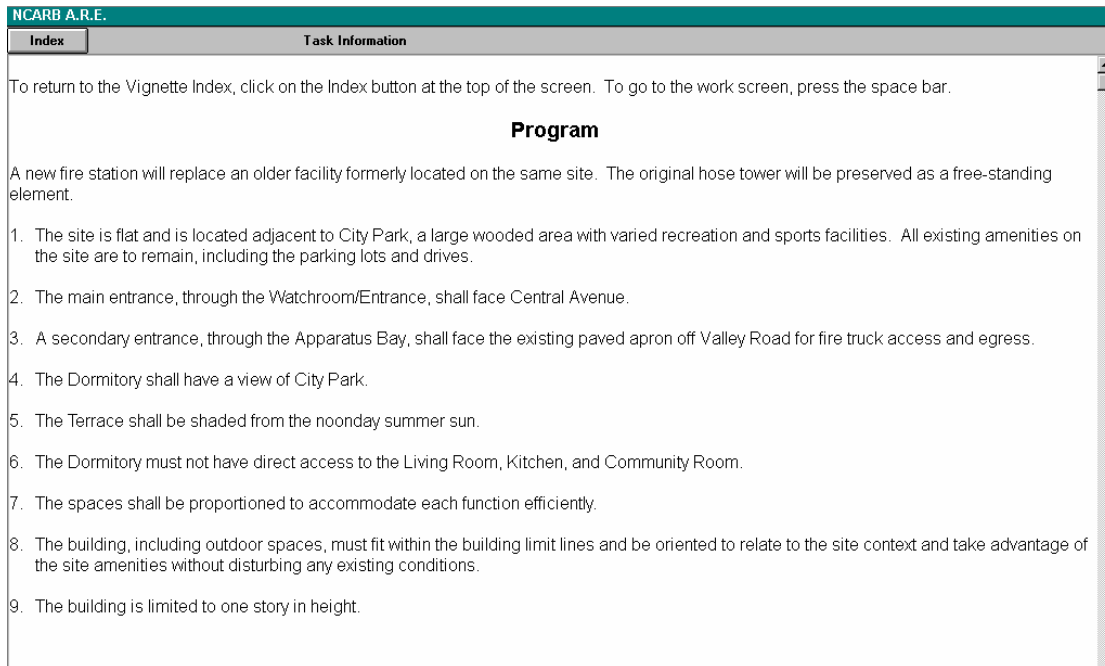


Figure 9. Goals and constraints of an ARE task (from Braun, Bejar, & Williamson, 2006)

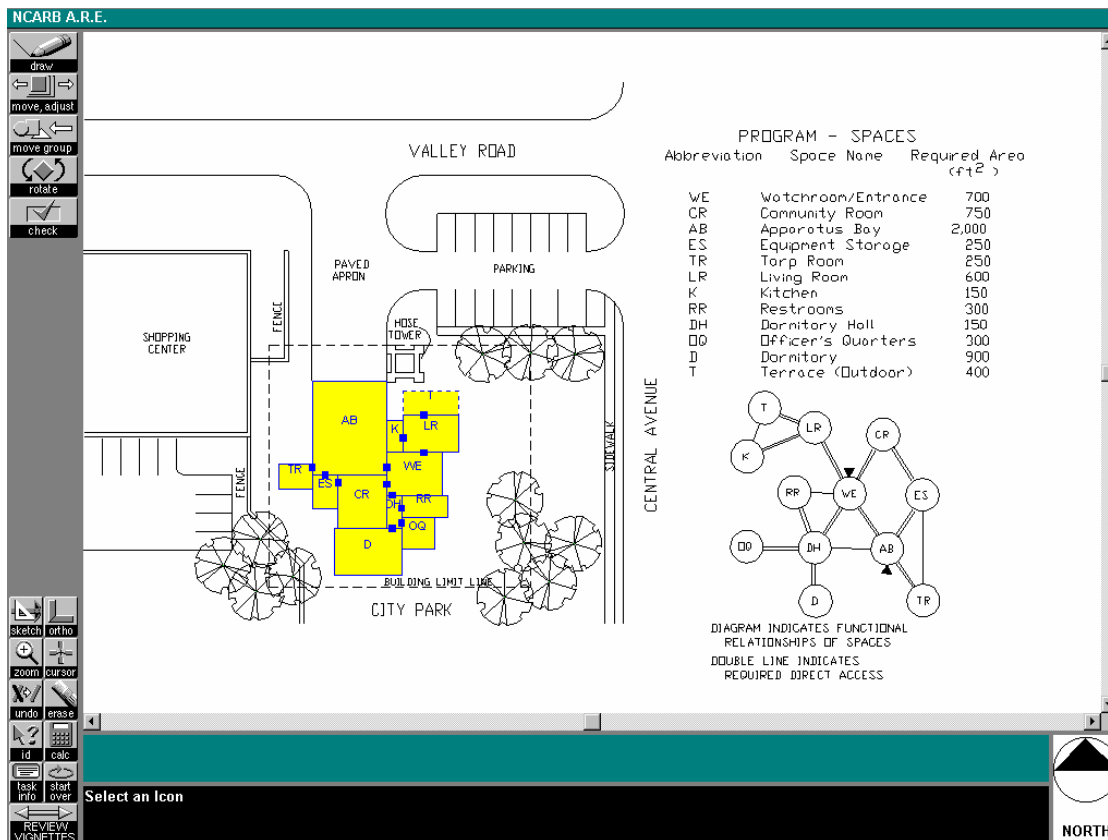


Figure 10. Solution for ARE task (from Braun, Bejar, & Williamson, 2006)

Beyond the CAD environment itself as a key form of knowledge representation, this kind of iterative solution directly illustrates one important advantage of knowledge representations and indirectly illustrates another. Note that design solutions are iterative for even expert architects, far too complex to be detailed without aids. We can think of the solution process in terms of the sequence of CI figures, with each cycle in solution starting from a provisional design *external* to the candidate, but capturing the outcomes of thinking thus far. The internal situation model the candidate constructs activate elements from LTM about constraints and ways of meeting them, while the details of the thinking that produced the provisional design need not remain active. The CAD representation can be thought of as an external component of working memory. The indirect illustration is that the final product can be shared with others. Any construction site is a whirl of activity by people with different areas of expertise and different responsibilities. The shared use of knowledge representations such as these site plans, blueprints, wiring diagrams, and project schedules support the complex of coordination their work. Each expert shares information as needed with other professionals by means of these knowledge representations they have jointly learned to use.

4.4 *The Importance of Interaction*

The Architectural Registration Examination shows that it is not necessary for all the details of the complex activities that people carry out to be somehow fully and explicitly modeled in our heads. Rather, the full complexity of activity is more usefully conceived as residing just as much in our environment. Simon (2001) tells of an ant making its way across a sandy beach. Its path is quite complex, as it tries one direction then another, backs up to avoid a chasm, crosses to different grain that is stable—continually winding, reversing, crossing over, moving forward. The ant did not plan this path. At each moment, the ant is interacting with features of its immediate surrounding, in accordance with a relatively small repertoire of behaviors tuned to ant local environments.

Human behavior is more complicated than ant behavior, of course, not to mention affect, reflection, and motivation. But it too is characterized by the interplay between a person and a situation, and mediated by the patterns through which the person interprets the situation. The capability to recognize in situations the patterns we have acquired, and to act through such interpretations to modify the situations or ourselves accordingly, that helps enable complex human behavior despite our limited attention and working memory. Static definitions, concepts, and relationships are necessary but not sufficient for

expertise; knowledge must be in a form that is primed for perception in and interaction with particular situations, as we saw in Larkin's study of physics experts.

Assessing students' capabilities to recognize patterns and react accordingly calls our attention to situations in which a student perceives and acts, the situation changes, and the student acts in response. The focus is on what happened before, what is missing, what to do next and why, all of which are addressed by reasoning through the patterns that are targets of learning. We need to be able to make inferences about students' capabilities to act in dynamic -- recognizing and understanding patterns, and knowing how to react to them or modify them.

What are some implications for other parts of the assessment argument? The environments must be able to provide feedback, offer affordances, and respond accordingly in ways both overt (a patient's results from a diagnostic test) and hidden (an allergenic response to a dental treatment only becomes apparent after a day of simulated time). The data concerning the situation in which an examinee acts is evolving, so at some grainsize, we need to track salient aspects of the state of the environment and perhaps its history in order to evaluate an examinee's actions.

Our last assessment example illustrates these ideas, and connects other aspects of expertise we have already discussed. It is the Hydrive intelligent tutoring system for learning to troubleshoot the hydraulic system in the F-15 aircraft. Like the ARE project, Hydrive was based in part on the results of a cognitive task analysis. Steinberg and Gitomer (1996) found that experts needed knowledge about the system itself, familiarity with troubleshooting tests and procedures, and strategies for when and how to apply the procedures. They coordinated these aspects of knowledge through systematic sequences of actions that are well described by the analyses in Alan Newell and Herb Simon's (1972) landmark volume *Human problem solving*, in terms of problem spaces, operators, procedures, and strategies. In particular, an active path in a hydraulics path is the chain of components from a signal, such as "pilot steps on rudder pedal", to an expected output, such as "rudder deflects" – and if the pilot steps on the pedal and the rudder doesn't move, then one or more of those components has a fault.

Hydrive contains a simulation of the hydraulics system, with whatever faults a task designer decides to enter (in operation, students work through a series of increasing complex problems). In a given problem, the student carries out troubleshooting actions in a very open-ended form, including replacing components as well as running tests. The simulator provides feedback to student actions, depending on the state of the system. Now evaluating a student's actions in the evolving solution is fundamentally different

from evaluating the correctness of responses in a static situation. Testing the output pressure of a pump half way along the active path in Hydrive suggests expert-level splitting of the problem space if it is the first action, but poor system knowledge if the test follows an earlier action that eliminated it from the problem space. The evaluation of an action is thus conditional on the current state of the system and the actions the student has already taken; that is, what the examinee *could* know from actions taken thus far.

A Hydrive solution is a dialogue much like the one shown in Figure 6, but with the simulator playing the role of a very limited conversationalist! The evolving situation model of a Hydrive problem-solver concerns possible states of the hydraulics system. He uses sensory input in the form of the system's outputs, the provisional model from previous actions, and, from LTM, knowledge of the system components and their usual interactions. An expert is able to construct a representation that has the right components in the right relationships, like the example in Figure 11. These components and their configuration, and their current behavior activate problem-solving strategies. They in turn activate procedures that are likely to provide useful new information about the system. A novice sees the same input, but lacks some of the expert-level knowledge connections. Figure 12, for example, is a diagram of a system in the same state as depicted by the expert who drew Figure 11. Not surprisingly, this situation model will not support effective troubleshooting actions.

Implications for assessment that fully exploring students' capabilities requires observing them in environments that act and react with the student. Ideally, as with ARE and Hydrive, and in the role-playing and conversation-based tasks that are used in language assessments, actual interactions take place. Technology helps us carry this kind of testing out in large scale systems, by providing suitable stimulus environments and allowing us to evaluate students' actions automatically. But even in static forms of testing such as multiple choice, tasks can be developed that at least evoke evidence of the kinds of knowledge, and the capabilities to build situation models, that are essential to interactivity. These would be questions in a real-world situation that strike at Salthouse's bottlenecks: What will happen next? What might have led to this situation? What is a good thing to do next? What would you expect to happen if you do it?

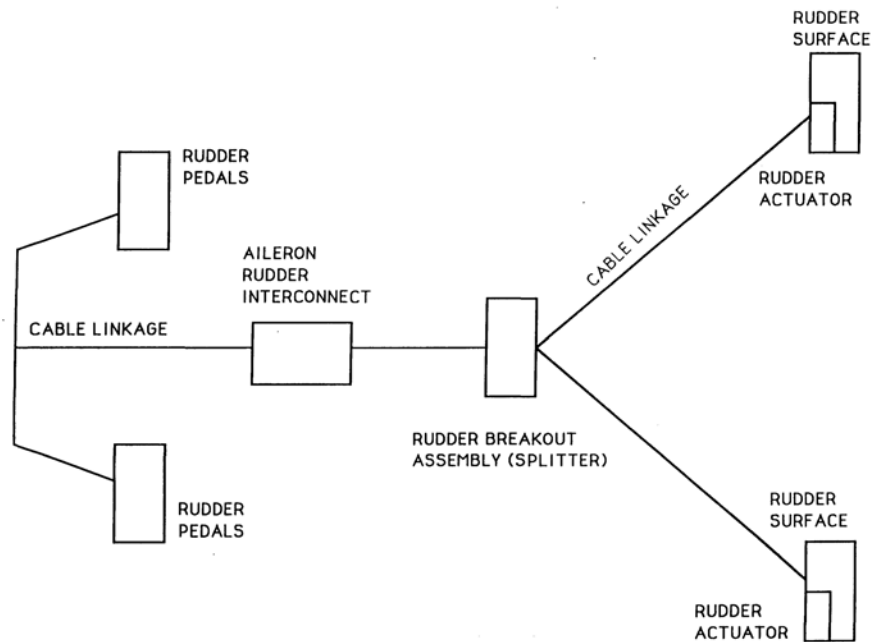


Figure 11. Expert’s drawing of the relevant parts of the system for a Hydrive problem (from Steinberg & Gitomer, 1996)

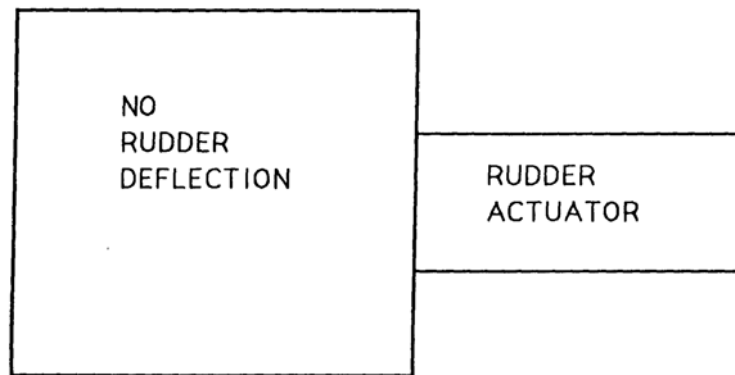


Figure 12. Novice’s drawing of the relevant parts of the system for a Hydrive problem (from Steinberg & Gitomer, 1996)

5. Conclusion

Cognition is always and inevitably interplay between the particulars of situations and the generalities embraced in patterns. The use of patterns “in our head” is embodied and

situated, constructed and always reconstructing. Many patterns are social, as to their construction, acquisition, and use. This include concepts, tools for thinking and acting as individuals and with others (in particular, symbol systems and knowledge representations), and ways of interacting with people and situations to accomplish human purposes. Some of these patterns, these ways of perceiving, interpreting, and acting, we express explicitly and want students to learn to use, as the students participate in activities in which they are useful.

Assessment is structuring situations that evoke evidence about students' thinking and acting in terms of these patterns. It is an exercise of meaning-making too: narratives about what students know and can do, in what kinds of situations; narratives cast in some conception of the nature of knowledge, its use, and its acquisition. The perspective suggests the kinds of stories that should be told—the kinds of claims one might make about students, the kinds of things one needs to see them say, do, or make to ground the claims, and the kinds of situations we might find evidence.

Educators are beginning to integrate insights from cognitive psychology and developments in technology into assessment practices. Some insights can be applied directly in familiar forms of testing. Others need to take advantage of technological environments and produce more complex performances, and lie outside traditional and familiar practices. In these cases, what is needed is good exemplars, extensions of psychometric methodologies, and principled frameworks for the design, implementation, and operation of assessments. These developments are beginning to take place. We, as the community with interests in learning and assessment, are moving to our next level of expertise.

References

- Atkinson, D. (2002). Toward a sociocognitive approach to second language acquisition. *The Modern Language Journal*, 86, 525-545.
- Bejar, I. I., & Braun, H. (1999). Architectural simulations: From research to implementation: Final report to the National Council of Architectural Registration Boards (ETS RM-99-2) Princeton, NJ: Educational Testing Service.
- Braun, H., Bejar, I.I., & D.M. Williamson (2006). Rule based methods for automated scoring: Application in a licensing context. In D. M. Williamson, R. J. Mislevy, & I. I. Bejar (Eds.), *Automated scoring of complex tasks in computer based testing* (pp. 83-122). Mahwah, NJ: Lawrence Erlbaum.
- Cameron, C.A., Beemsterboer, P.L., Johnson, L.A., Mislevy, R.J., Steinberg, L.S., & Breyer, F.J. (1999). A cognitive task analysis for dental hygiene. *Journal of Dental Education*, 64, 333-351.
- Chi, M.T.H., Feltovich, P., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science*, 5, 121-152.
- Claxton, G. (2001). Education for the learning age: A sociocultural approach to learning. In Wells & Claxton (Eds), *Learning for Life in the 21st Century*. Oxford, UK: Blackwell.
- de Groot, A.D. (1946). *Het denken van den schaker. Een experimenteel-psychologische studie*. Amsterdam: Noord-Hollandsche Uitgeversmaatschappij.
- Ericsson, K. A., Charness, N., Feltovich, P., & Hoffman, R.R. (Eds.) (2006) *Cambridge handbook on expertise and expert performance*. Cambridge, UK: Cambridge University Press.
- Gardner, H. (1993). *Educating the unschooled mind*. Washington, D.C.: Federation of Behavioral, Psychological, and Cognitive Sciences.
- Gee, J.P. (1992). *The social mind: Language, ideology, and social practice*. New York: Bergin.

- Gitomer, D.H., & Steinberg, L.S. (1999). Representational issues in assessment design. In I.E. Sigel (Ed.), *Development of mental representation* (351-370). Mahwah, NJ: Erlbaum.
- Hutchins, E., & Klausen, T. (2000). Distributed cognition in an airline cockpit. In Y. Engström & D. Middleton (Eds.), *Cognition and communication at work* (15-34). New York, NY: Cambridge University Press.
- Katz, I.R. (1994). Coping with the complexity of design: Avoiding conflicts and prioritizing constraints. In A. Ram, N. Nersessian, & M. Recker (Eds.), *Proceedings of the Sixteenth Annual Meeting of the Cognitive Science Society* (485-489). Mahwah, NJ: Erlbaum.
- Lakoff, G. & Johnson, M. (1980). *Metaphors we live by*. Chicago: University of Chicago Press.
- Larkin, J. (1983). The role of problem representation in physics. In D. Gentner & Stevens, A. L. (Ed.), *Mental Models* (pp. 75-98). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Lave, J. (1988). *Cognition in practice: mind, mathematics and culture in everyday life*. Cambridge University Press.
- Messick, S. (1994). The interplay of evidence and consequences in the validation of performance assessments. *Educational Researcher*, 23(2), 13-23.
- Mislevy, R.J., Steinberg, L.S., Breyer, F.J., Almond, R.G., & Johnson, L. (1999). A cognitive task analysis, with implications for designing a simulation-based assessment system. *Computers and Human Behavior*, 15, 335-374.
- Newell, A., & Simon, H.A. (1972). *Human problem solving*. Englewood Cliffs, NJ: Prentice-Hall.
- Patel, V. L., & Kaufman, D. R. (1996). The acquisition of medical expertise in complex dynamic environments. In A. Ericsson (Ed.), *The road to excellence: The acquisition of expert performance in the arts and sciences, sports and games* (127-165). Mahwah, NJ: Erlbaum.
- Scribner, S. (1984). Studying working intelligence. In B. Rogoff & J. Lave (Eds.), *Everyday cognition: Its development in social context* (9-40). Cambridge, Mass. Harvard University Press.

Simon, H.A. (2001). *The sciences of the artificial* (4th ed.). Cambridge, MA: MIT Press.

Snow, R.E., & Lohman, D.F. (1989). Implications of cognitive psychology for educational measurement. In R.L. Linn (Ed.), *Educational measurement (3rd Ed.)* (263-331). New York: American Council on Education/Macmillan.

Steinberg, L.S., & Gitomer, D.G. (1996). Intelligent tutoring and assessment built on an understanding of a technical problem-solving task. *Instructional Science*, 24, 223-258.