5 Preparation for Future Learning

A second major problem with knowledge, in addition to its isolation, is that a knowledge assessment is a description of a purportedly stable mental state. Assessment designers try to ensure that they are detecting a stable state and not a temporary effect, for example, by doing test-retest reliability measures. Knowledge is taken as an end or start state; it is not about learning per se. This concern with capturing knowledge states rather than dynamic change has had major implications for assessment. The first has been an emphasis on problem solving. The second has been a focus on mastery. While both are good, they do not directly emphasize learning, and they miss experiences that prepare people to learn. We detail these points next, while describing an alternative type of assessment that looks at learning itself.

Problem Solving without Learning

The cognitive revolution has stressed problem solving and, unfortunately, largely left the learning emphasis of behaviorism behind. (Whatever one might think of the restricted explanations of behaviorism, its focus on the tight connection between learning and motivation was a great strength.) A perusal of the finest cognitive psychology textbooks (e.g., Anderson 2000) reveals scores of constructs that explain the knowledge organizations and processes that affect problem solving—schemata, priming, working memory, echoic buffers, attention shifting, and so on—but only a handful of constructs to explain learning, most of which emphasize memory encoding and retrieval (e.g., association and consolidation). Attempts to make knowledge more dynamic by using the active verb *knowing*, or *knowing in action*, suffer the same problems: they are about problem solving and not learning.

Users of knowledge assessments can *infer* learning by giving the same assessment as a pretest and posttest. This is quite rare in educational practice; few teachers give pretests to measure subsequent learning gains, in part because it detracts from instructional time. Even if they did, wouldn't it be more to the point to evaluate learning directly?

Bransford and Schwartz (1999) labeled most current assessments as *sequestered problem solving* (SPS). In the typical SPS assessment, students are sequestered (like a jury) from learning opportunities and outside resources that might contaminate the validity of the assessment. Learning during a test would be cheating. Consider the following informal example of the shortcomings of SPS assessment.

A late-night talk show host asks a group of students who have just graduated from Harvard a handful of tricky questions like, "Is the earth closer to the sun in June or December?" or, "What is the best way to reintroduce a baby eagle to the wild?"

To the delight of the audience, the Harvard grads fare no better on these questions than a group of high schoolers. This assessment purports to demonstrate that the experience of going to Harvard does not produce the learning gains one would expect. The television show *Are You Smarter Than a 5th Grader?* has capitalized on making otherwise-educated people look stupid. Adults often do not recall the facts that the show asks, and the fifth graders regularly do. Other examples include the frequent polls that indicate that Americans do not know basic facts that must have been covered in school or newspapers.

What these scenarios actually demonstrate is that SPS assessments do not tell the whole story, though they do capture the public mind for what it means to have learned. Imagine, instead, what would happen if both groups of students were given access to learning resources during their quizzes. The Harvard students would probably use the learning opportunity to produce responses that outshone those of the high schoolers. Similarly, the adults would probably be better at finding the answers to the questions than the fifth graders (although one never knows how they choose the adults to be on such a program).

Here is a second, simple thought experiment. Imagine that a firm wants to hire a financial analyst. Tom has just completed a two-week course in Excel—his first exposure to spreadsheet software. Sig has not learned Excel. Instead, using multiple spreadsheet packages over the past several years, he taught himself, achieving high levels of expertise. The company decides whom to hire by using a paper-and-pencil test of basic Excel operations that just happen to have been covered in Tom's course. Tom would probably do better on this SPS test. We suspect, however, that Sig would be more likely to serve the company well in the long run. His deeper understanding of spreadsheet structure and capacity to learn independently will enable him to learn and adapt on the job—for example, when the company switches to a new software package or when the employees are asked to learn advanced features of Excel on their own. The failure of SPS tests is one reason that all employers would prefer to hire people for a trial period to see if the employee adapts to and learns in their local context. Edwin Ghiselli (1966) reported that aptitude tests only predict 9 percent of job performance immediately after training, and subsequently drop to 5 percent after time on the job. This is because people learn on the job, and sequestered assessments like aptitude tests are not designed to predict people's future learning.

An alternative to a static SPS assessment is a *dynamic assessment*. Reuven Feuerstein (1979) introduced dynamic assessments as an alternative to the standard administration of IQ tests. During the test, he would assist children, and therefore would see if they could learn from his help to solve the IQ problems. By using a dynamic assessment, he was able to make more useful diagnoses about children's learning potential.

Bransford and Schwartz (1999), who were concerned that theories of transfer were only focusing on the application of knowledge rather than learning, proposed a dynamic assessment format they termed *preparation for future learning* (PFL). In a PFL assessment, there are resources for learning during the test, and the question is whether students learn from them. Bransford and Schwartz reasoned that PFL assessments would be sensitive to differences in instruction that would be missed by static SPS tests. As support, Schwartz and Taylor Martin (2004) contrasted SPS with PFL assessments in the context of teaching statistics to ninth-grade students. In this experiment, one factor was how students were taught. Half the students received direct instruction, and the other half completed a form of guided discovery called "inventing" (cf. Schwartz, Chase, Oppezzo, and Chin 2011). A week later the students completed a long test. At the end of the test, there was a problem beyond the edge of what they had been taught.

The experiment crossed the instructional factor with a second one that involved the test itself. For half the children in each instructional condition, the tests simply included the difficult target problem without any resources for learning during the test. Thus, the students who used this form of the test completed an SPS assessment. The other half the students in each instructional condition completed a PFL version of the test. In the middle of the test, there was a worked example that students had to follow to solve an accompanying problem. The worked example showed how to solve a new kind of statistics problem, and students had to follow the worked example to solve a similar problem on the exact same page. All the students did well at copying the worked example to solve the associated problem. The main question, though, was whether the students would learn from the worked example. Unbeknownst to the students, the information in the worked example part of the test held the key to solving the hard problem at the end of the test.

Figure 5.1 shows the main results for how well students performed on the difficult problem at the end of the test. First,

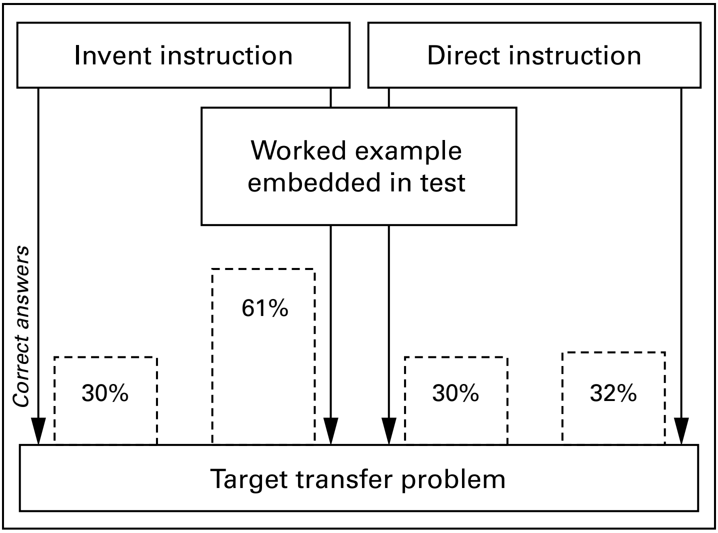

Figure 5.1
Improving the assessment of instruction by evaluating students' abilities to learn during a test
The students represented by the arrows in the middle of the graph received a learning resource as part of their test. This "learning during the test" assessment made it possible to differentiate the value of the two forms of instruction shown in the upper boxes. The students who did not receive the learning resource (the outer arrows) performed the same, which would have erroneously led to the conclusion that the two instructional treatments had the same value. (Adapted from Schwartz and Martin 2004.)

notice the levels of performance on the far right and far left of the graph. These are the levels of performance of the students who received the SPS version of the test, which did not include the worked example resource. By this SPS assessment, both forms of instruction led to the same outcomes, and one might naturally conclude that there is no advantage for one type of instruction over the other. Next, look at the two center bars indicating how well students did on the PFL version of the assessment. These are the students who received the worked example in the middle of their test. In this case, the invent instruction students doubled the performance of the direct instruction students. The invent instruction had better prepared students to learn during the assessment and then spontaneously use this learning. The simple point of this extended research example is that the use of dynamic assessments that include opportunities for learning can tell us a good deal more about the effects of instruction than can tests that simply measure static knowledge.

Mastery without Change

Knowledge provides a poor account of change. Either one has it or one does not, and knowledge-based accounts have trouble explaining how knowledge bootstraps itself from one state to another. This was highlighted in the learning paradox offered by Plato (*Meno*, 80.d):

> But how will you look for something when you don't in the least know what it is? How on earth are you going to set up something you don't know as the object of your search? To put it another way, even if you

come right up against it, how will you know that what you have found is the thing you didn't know?

Knowledge-based theories cannot easily give an account of where fundamentally new knowledge comes from. This is one reason that people argue that certain human abilities are innate; they cannot explain how the abilities could have been learned (Chomsky 1966; Pinker 1994). It is also one reason why it has been so difficult to produce a satisfactory knowledge-based account of conceptual change. Where can new concepts come from, if they do not come from ideas one already has? But if they come from prior concepts, then they must not be new concepts. Such is the knot of knowledge-based accounts.

In assessment, the description of knowledge as a state has led to an emphasis on mastery. Tests emphasize the mastery of skills and knowledge: Do students have the state of knowledge or not? If they exhibit mastery, they have knowledge. If they do not exhibit mastery, they do not have knowledge.

How does a mastery emphasis interact with the goal of seeing whether students are prepared for future learning? One assumption appears to be that if we want to assess someone's preparation for future learning, we should see if they have mastered the past. This seems like the rationale behind the Scholastic Assessment Test (SAT). The test tries to predict college success by seeing if students have mastered the mathematics, reading, and writing from earlier lessons.

The idea of looking at prior mastery to predict future learning is reasonable, but there is a catch. Tests of mastery presuppose knowledge in a mature form, implying that anything short of mastery does not count as knowledge and cannot be assessed.

Yet this is not true. First, people have earlier forms of understanding that do not comprise full-blown, declarative or procedural knowledge but that are nevertheless crucial for future learning. Michael Polyani (1966) referred to this as tacit knowledge. Harry Broudy (1977) described it as *knowing with*, as distinguished from *knowing that* and *knowing how*. Second, it is possible to assess these earlier forms of understanding, if we create assessments that allow learning during the test.

The late Russian psychologist Lev Vygotsky ([1934] 1987, 200) neatly captured the peril of mastery assessments:

> Like a gardener who in appraising species for yield would proceed incorrectly if he considered only the ripe fruit in the orchard and did not know how to evaluate the condition of the trees that had not yet produced mature fruit, the psychologist who is limited to ascertaining what has matured, leaving what is maturing aside, will never be able to obtain any kind of true and complete representation of the internal state of the whole development.

Knowledge-based assessments that can only detect mature forms of knowledge miss many of the important precursors of learning. One of PFL assessments' key benefits is that they can detect immature forms of understanding that SPS assessments miss. This makes them well suited to studying the types of informal learning that are so prevalent in today's information-rich ecosystem, because these types of learning often yield significant experiences that cannot be detected by mastery-focused tests.

For example, PFL assessments can be useful for evaluating digital game-based learning. James Paul Gee delivered a keynote to game researchers at the 2010 International Conference on the Foundations of Digital Games in which he mentioned that games could and should be used as preparation for future learning. It is easy to imagine how digital games might serve in this role. The rich experiences offered by gameplay are likely to produce tacit knowledge that is not easily measured by SPS tests of knowledge but that might be detected by PFL assessments. Players of the popular video game *Portal*, for instance, have ample opportunity to experiment with conservation of momentum, yet lacking some formal explanation of the phenomenon, the players would probably not do well on a physics test given in an SPS format. A PFL physics test, however, could provide a link between the players' experiences in the game and the formal physics concepts, thereby revealing the advantages for those who played *Portal* compared to a control group that did not play the game before taking the test.

To test the idea that the experience provided by digital games could serve as preparation for future learning, we built a game designed to help students learn the basic concepts of probability distributions (Arena and Schwartz 2010). Statistics is a notoriously difficult topic for people to learn and reason about (Nisbett, Krantz, Jepson, and Kunda 1983). One reason is that people naturally reason about single outcomes rather than distributions of outcomes. For instance, they often believe that the goal of statistics is to predict a single outcome versus a pattern of outcomes (Konold 1989). Making a "point prediction" reflects a causal form of reasoning as opposed to a statistical one. To the novice, random often connotes "without pattern," so the idea of statistics, where randomness supports inferences, can be difficult to grasp. We believed that one way to address this problem was to provide students with experiences that would give them intuitions about patterns of randomness (i.e., distributions) and probability before they received explicit instruction. Thus, our goal was not to produce a game that would teach students everything they needed to know about probability distributions (a bleak prospect for a game). We instead wanted to produce a game that would give students experience interacting with and thinking about probability distributions informally, so that when they received a subsequent exposition about probability, their gameplay experiences would help them understand the formal concepts. To that end, our game makes no explicit references to probability terms or concepts.

Figure 5.2 (plate 3) shows that in the game *Stats Invaders*, players simply shoot aliens dropping from the sky while trying to determine which of two displayed patterns (actually probability distribution curves) best describes the pattern in which the aliens are dropping. Once students have made their determination, they pick one of the two distributions on the right. This launches a bomb that is tuned to explode a mother ship in the same distribution (the game is a bit "male"). If the player picks the right bomb, the mother ship hidden above the descending aliens is destroyed. If not, the player loses a life—one of the standard tricks of video games.

In a study of the effectiveness of the game, we adapted the same research logic from before, where we compared SPS versus PFL measures of learning. As a pretest, a group of community college students completed a brief questionnaire about randomness and probability distributions with questions like, "What is the purpose of finding the pattern for a type of random event?" and, "What is the defining characteristic of a uniform distribu-

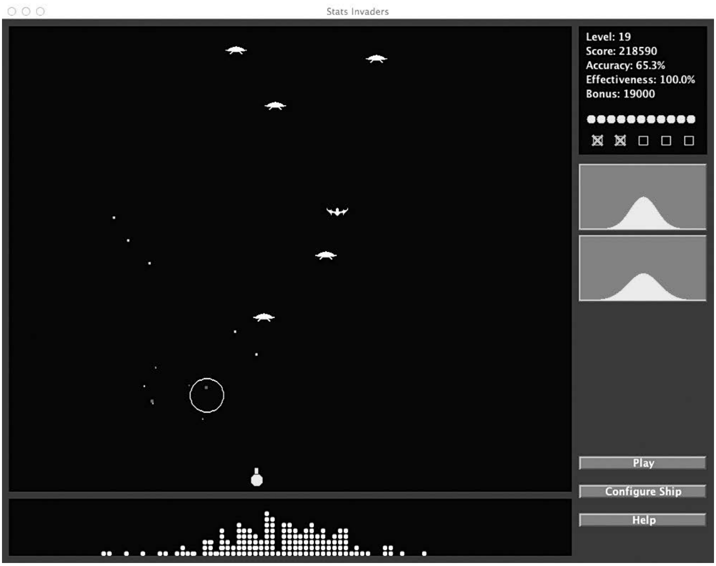

Figure 5.2 (Plate 3)

Stats Invaders: A video game to support the development of statistical intuition

Players shoot at descending aliens in the style of old arcade games. The aliens fall according to a horizontal statistical distribution (e.g., they are more likely to descend in some locations than in others). To end a round, players have to set a bomb tuned to the frequency of the "mother ship" by picking which of the shapes on the right side best describes the shape of the ship dropping the aliens. The goal is to help students develop intuitions about patterns within chance.

tion?" Then we had half the students play our game and half not. Next, half the students in each condition received a short reading passage on patterns in randomness. Finally, we gave all the students another test just like the first one. There were two key questions. Would playing the game improve performance on the test in its own right? And more important, would students who had played the game learn more from the passage than those who had not played the game?

Figure 5.3 shows gains from the first test to the second. Students from both groups learned from the passage, but students in the gameplay condition learned much more than students in the no-gameplay condition. The game players were able to relate what they were reading to their recent experiences in the game. Playing the game developed the earlier forms of knowledge that prepared the students to learn when given a chance. The game had created a time for telling (Schwartz and Bransford 1998). Notably, if we had not given the students a chance to learn, this benefit of the game would have gone undetected. Assessments that involve learning as part of the test can diagnose the benefits of intuitively compelling experiences that are missed by static knowledge tests. In chapter 7, we provide an example of a different type of PFL assessment that we used to evaluate whether popular commercial video games prepare students for future learning.

In the two cases of PFL assessments for statistics learning, the PFL assessment ultimately depended on a knowledge posttest. Students received posttests that examined whether they could answer questions based on the knowledge that had been told to them. This was valuable because it helped to validate the PFL

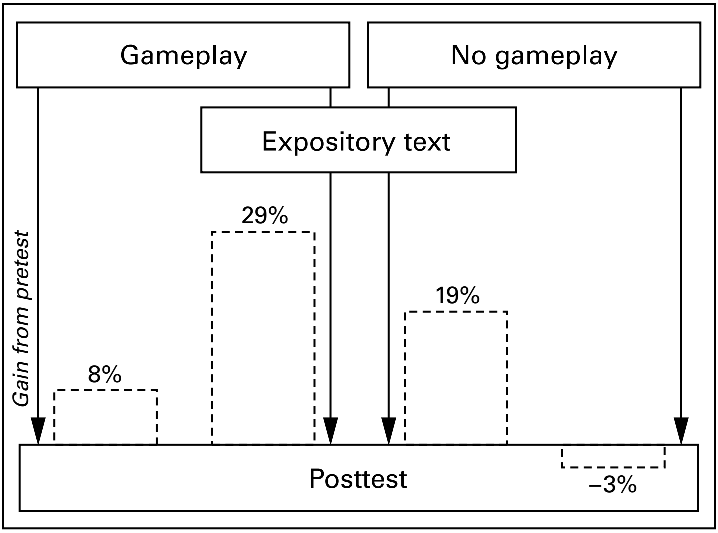

Figure 5.3
Benefits of gameplay in preparing students for future learning
Students who played a specially designed game to develop statistical intuitions learned more from an exposition on randomness than students who had not played the game.

approach to assessment using current standards and terms. One can only imagine trying to convince a policymaker about the value of new forms of assessments, if those assessments completely abandoned familiar terms. It would be like arguing that the gross domestic product (GDP) should be replaced by gross domestic happiness (GDH). Measuring happiness is not going to get very far in policy debates, despite the fact that it is closer to the goals of the citizens. To convince people that the GDH is a useful index, one would show that it supplies useful data that capture the value of the GDP measure, but that it also provides even more value. Therefore, our strategy is to first show that new forms of assessment offer added value, for example, by capturing what people think they care about in current terms (knowledge gains), while also capturing new information.

Our next task is to show that it is possible to use a choice-based approach to a PFL assessment. For example, Ryan Baker, Sujith Gowda, and Albert Corbett (2011) found that student choices of how to use help in an intelligent tutoring system predicted students' subsequent learning. The next chapter provides further relevant cases. After that, the final task is to show that choice-based assessments predict (and can change) future choices, thereby completely removing knowledge tests from the loop. We have not yet completed this final step (nor have we finished with the others). As neatly stated by José Ortega y Gasset (1960, 200), "Reason is not a train leaving at a fixed hour." We do not think people are ready to fully abandon knowledge until many people, including ourselves, help set forth strong evidence and reasons that we can do better.