Choice-Based Assessments for the Digital Age

Daniel L. Schwartz & Dylan Arena

Stanford University, School of Education

August, 2009

Educational assessment is a normative endeavor: The ideal assessment both reflects and reinforces educational goals that society deems valuable. A fundamental goal of education is to prepare students to act independently in the world—which is to say, to make good choices. It follows that an ideal assessment would measure how well we are preparing students to do so. The argument of this paper is that current assessments, which primarily focus on how much knowledge students have accrued, are inadequate. Choice, rather than knowledge, should be the construct around which assessments are organized. Digital technologies make this possible, because interactive assessments can evaluate students in a context of choosing whether, what, how, and when to learn.

Our plan will be as follows. We will start by providing a few reasons that choice should be the central construct of assessment. Then we will discuss how technologies can drive changes in assessment practices and make choice-based assessment and instruction possible. To make headway, though, the third section argues that it will be important to dethrone knowledge as the primary construct of assessment, lest new technologies only make us more efficient at measuring the wrong thing. Next, as a feasibility demonstration, we will describe several studies that have successfully measured learning by assessing student choices. Building on this preliminary work, we envision a digital “choice-adaptive learning environment” that provides a model for making headway in the development of integrated instruction and assessment environments. Finally, we will briefly consider the place of choice in discussions of 21st century competencies and skills.

Before we go any further, however, we must clarify what we mean when we use the term “choice.” As suggested above, we take it as foundational that a primary goal of education is to help students develop abilities and understandings so they can make choices that maximize their chances of succeeding within and beyond school; and we believe, therefore, that choice should be the central construct of assessment. However, we recognize that not all choices are in the purview of education. Choice assessments should not be a back door for enforcing beliefs that fall outside the domain of public education (such as whether students make the “correct” choice about a political matter). Instead, they should indicate whether students can learn and adapt in productive ways. Thus, our discussion of choice-based assessments refers to learning-relevant choices such as how and what to learn, not all choices.

Choice is the Central Concern

Among the many stakeholders in education, choice is the central concern. Parents care about their children’s choices, for example, how they spend their free time, whether they try hard at school, and even whether they develop a sense of the possibilities from which they might choose. Parents hope for “good choices,” and arguably, many parents care about “good knowledge” (or grades) to the extent that it creates opportunities for choices later on. Yet despite the manifest importance of choice, current assessments only evaluate a degraded sense of choice, as in choosing an answer on a multiple-choice test.

Education professionals also value choice. They embrace schools for their abilities to help students choose and learn once they leave school. Schwartz, Bransford, and Sears (2005) interviewed superintendents to see what help learning scientists might provide to their endeavor. One possibility was that they would ask for help in achieving high test scores to increase their districts' standings. They did not. Instead, "[t]he surprisingly unanimous answer (they were surprised as well) was that they wanted us to help students make their own choices in the future. They wanted the students to be able to 'learn for themselves' and make informed decisions. They believed that well-designed school experiences could transfer to help children continue to learn once they left school" (p. 2). Ideally, assessments would provide information to these superintendents about how their schools are doing in this respect, but measurement of such outcomes is beyond the reach of current assessment practices.

Figure 1. Text materials of the BAR/BRI course that prepares students for the California Bar exam.

Outside of classrooms that exert strict curricular control, the need to make learning choices is the norm. These choices are where the rubber of school meets the road of life. One particularly vivid example involves preparation for the bar exam. Law school graduates need to pass a bar exam to become practicing lawyers, but law schools typically do not teach the specific knowledge needed for the exam. Law schools focus on broad issues and ways of thinking rather than the specifics of particular state codes. To prepare for the exam, students routinely take special courses independent of law school. These preparation courses provide an overabundance of learning resources such as readings, reviews, outlines, practice tests, case synopses, videos, live lectures, workshops, and on-line tutorials. Figure 1 shows the thousands of pages of textual materials for one such course, the California BAR/BRI (www.barbri.com). Across materials, the content is highly redundant, so rather than plowing through everything, well-educated law students choose the presentation format, activities, and timing of their study, as well as the social arrangements that they feel suit their learning needs for different topics within the curriculum. Their learning is driven by their choices of what, when, how, and with whom to learn.

In this example, experiences within law school help students make sense of the content of the materials, so they can make choices about how to navigate the mountain of resources to optimize their progress towards the exam. Of course, not everyone has to prepare for a bar exam, but we all face situations outside of school for which our school knowledge is not enough, such as comparing cars or camcorders. In these situations, the choices we make about how to learn what we need to find out will determine our success.

The modern workplace also puts choice at a premium. In prior generations, people could better anticipate a clear career trajectory and stable lifetime employment with one firm. In today's economy, jobs are no longer stable, and skills need to be updated frequently. New information and affiliated technologies appear daily, and workers must constantly adapt to new contexts, colleagues, jobs, and even careers, to keep up with changes in competition and industry structure (Benner, 2002). In addition, workers are expected to participate in their own growth, as demonstrated by the widespread use of continuous improvement programs across industries where the retooling of skills is considered part of the job itself (Appelbaum & Batt, 1994). In today's economy, at every level, the ability to accomplish difficult tasks is more likely to depend on one's ability to navigate the vast array of informational resources than it is to be based on static knowledge that could be measured by today's assessments. The question is no longer, "What do you know?" The question is now, "What can you successfully choose to learn?"

Finally, choice is at the center of a free society that emphasizes democracy and opportunity. Democracy depends on people's abilities to recognize and execute choices within the constraints that make society possible. Agency and participation are operationalized in choice. Questions of identity and inclusion matter because they contribute to the choices that people make.

Societies achieve their ideals of choice to varying degrees because of pre-existing conditions, biases, and ill-formed political structures. Schools should not further contribute to a loss of choice; instead, they should directly address issues of choice in developmentally appropriate ways. For obvious reasons, children should not have the freedom of choice with the responsibility it entails to the same extent as adults. At the same time, choice is at the center of our social philosophy, and therefore, it should be at the center of assessments that are increasingly the beacon of what schools should accomplish.

Why Technology Matters

Our proposal is that new digital technologies make it possible to use choice-based assessments that accord with people's intuitive sense of what education should be about. Choice-based assessments can better reveal how well students are learning specific topics and help change instruction to emphasize people's abilities to adapt and learn. Historically, technology has always had a powerful influence on instruction and assessment. To take a remote example, the Dark Ages had limited technologies for information storage. Therefore, information was carefully transmitted from teacher to pupil, like a flame from one candle to another. Monks painstakingly transcribed manuscripts letter for letter. One can imagine that assessments of monks largely involved their abilities to reproduce what came before without error, and a good deal of instruction also focused on errorless knowledge transmission.

Advances in technologies create fresh opportunities for changing assessment and instruction. In general, the assessment enterprise has been quick to adopt the efficiencies of new technologies, perhaps quicker than the instructional enterprise. Most people born before the 21st century are familiar with the #2 pencil and bubble forms. The bubble form was a technology that rapidly suffused the assessment enterprise. Professor Ben D. Wood, who helped design the IBM 805 that scored bubble forms, used the resulting income to endow graduate student fellowships at Teachers College, Columbia University. More recently, the bubble form is being replaced by fully automated computerized testing, which can collect, analyze, and transport data the moment an assessment is over.

The quick uptake of new assessment technologies creates the possibility of a benevolent Trojan horse. In the rush to embrace more efficient assessment technologies, new forms of assessments can be introduced that influence education in productive ways. One productive change would replace the common practice of using assessments at the expense of instruction. Testing requires time, and students are strictly forbidden from learning during the test except in incidental ways (e.g., Roediger & Karpicke, 2006).

Bransford and Schwartz (1999) labeled the typical assessment format "Sequestered Problem Solving" (SPS). Like a jury, students are sequestered from learning opportunities and outside resources that might contaminate the validity of the assessment. Learning during a test would be cheating. A major limitation of SPS assessments is that they are a retrospective assessment of what students were taught, when what we really care about is whether students are in a position to continue learning and growing. As we describe below, assessments that are designed to focus on learning during the test can be more revealing than SPS measures. The late Russian psychologist Lev Vygotsky neatly captured the peril of retrospective 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…" (1934/1987, p. 200)

An alternative to current assessments, which often detract from instruction, is to integrate assessment and instruction. There are several advanced technologies that already do this well. For example, Cognitive Tutors (Koedinger & Anderson, 1997) monitor how students are solving problems on the computer. These assessments are embedded within the learning tasks themselves. Much like a tutor observing how a pupil is solving a problem, the computer system adjusts instruction based on a model of what the student can do so far. Thus, one key improvement fashioned by technology is the ability to integrate assessment seamlessly into the process of learning itself. In this way, assessments can provide useful feedback to inform, rather than compete with, learning opportunities.

A second promise of technology is that it enables an alternative to the linear format of instruction. Most instruction currently takes the form of a pre-determined linear sequence. One historical reason for a linear curriculum that precludes student choice involves assumptions about efficiency in learning. School textbooks rarely ask students to choose how or what to learn, even when students are taking an elective course. The assumption is that these decisions should be left up to experts, who can make more efficient learning choices than a novice could. There is merit to this assumption, but it comes at too high a price when it removes all choice. If students never have a chance to experience feedback about good and bad learning choices in the relatively protected atmosphere of school, they will be under-equipped to learn how to make learning choices when they are on their own and the stakes are higher. A reliance on the efficient acquisition of knowledge at the expense of choice is a recipe for making routine experts rather than adaptive ones (Hatano & Inagaki, 1986).

Consider the case of teaching metacognition. Metacognition refers to ways that people can monitor and regulate their own problem-solving and learning. Most curricula that incorporate lessons on metacognition do not give students choices about learning. Therefore, the lessons on metacognition are inert: Without opportunities to experience the value of choosing to use metacognition or not, the instruction simply becomes more lessons without an authentic context of application. For instance, the metacognitive skill of time management cannot be taught effectively in schools where activities are regimented to the minute and enforced by bells and speakers, so parents often use homework to teach their children the skill of time management in a context of learning.

A second, more practical reason for a pre-determined linear curriculum involves the sheer challenge of monitoring student learning if students are allowed to follow different trajectories. Teachers of public high school in the United States can have roughly 150 students per day, which means teachers might face the prospect of tracking 150 different learning trajectories. It is much easier to track the position of each student against a single trajectory as specified by the curriculum. Even in the context of individualized computer instruction, it has been necessary to assume a single, idealized trajectory. The previously mentioned Cognitive Tutors, like all forms of programmed instruction, presuppose a sequential curriculum and an ordered sequence of learning, so it is possible to keep track of students and move them forward or backward in the sequence depending on their performance.

However, there are newer genres of technologies that do not presuppose a linear progression but still provide ample opportunities for assessment and learning. For instance, multi-player videogames like the World of Warcraft (www.worldofwarcraft.com) include a slate of choices about what to do, whom to be, what to learn, and how to learn. Here is a sample of the choices available to new players of the World of Warcraft at the time of this writing (more options are added frequently). Players choose one of over two hundred Realms in which to create their characters (there are four different types of Realm, determined by whether or not players are expected to act "in-character" and whether or not players are allowed to attack each other). They then choose one of two factions (Alliance or Horde); one of ten species (Blood Elves, Draenei, Humans, etc.); one of two genders (Male or Female); and one of ten careers (Druid, Hunter, Shaman, etc.). The players must then choose their appearances from among thousands of possible configurations—and all this is before the game has even begun. Once the players' characters enter the world, they choose their characters' professions, talents, friends, enemies, quests, worldview, and goals. They then embark on a completely open-ended adventuring career. Everything players of World of Warcraft learn about how to succeed in the game (which is a lot—it is a very complex game) is defined entirely by their choices, from the setup of the game to whom they interact with online. This profusion of choices might seem daunting, but it encourages players to feel a real ownership for and become strongly invested in their characters' success.

These types of choice-rich environments also include ample opportunities for assessment. With more choices and interactivity comes more information about the learner. Portfolio and project assessments have tried to capitalize on the increased information found in choice-filled environments, but they have proven relatively intractable for wide-scale assessments (Resnick & Resnick, 1994). Technology can help overcome the difficulties associated with increased information. Computers can automatically log all user behaviors in the system that might be of interest to a teacher, assessor, or researcher, ranging from chat logs to interpersonal distance to direction of gaze to activities engaged in—it is an ethnographer's thick description for free. Systems like World of Warcraft take advantage of this rich information to provide assessments to the players and to further regulate the available choices. They are instrumented with extremely refined metrics of player preferences, progress, and strengths. For example, there are metrics that indicate various powers and levels of accomplishment and access. These metrics are clearly available to players through gauges, points, game levels, ratings by peers, and so forth. They serve as powerful motivators to do even better (Reeves & Read, in press). In this case, assessment is built into an environment of high choice. Gee (2003) has made the compelling case that in this respect video games are a great model for effective instruction and assessment: In good video games, assessment is an integral part of the design of instruction and is built into the core mechanics of the learning interactions.

Ideally, advances in new technologies will make it possible for students to experience a broader range of learning choices and their effects. These technologies will avert problems of students perseverating in unproductive directions, because automated assessments of students' choices will be integral to the learning environments. The possibilities for guidance will be much greater than the current "Correct → Go-Forward / Incorrect → Go-Backward" methodologies. These assessments will not depend on assumptions of a strict linear curriculum, because the technology can accommodate much more complex data and analyses than were possible just a few years go. They will make it possible and efficient to monitor student learning choices and their consequences. They will be useful for everyday instruction and high-stakes assessments, because both will involve learning.

The Isolation of Knowledge-Based Assessments

Before continuing our positive case for choice-based assessments, we will now consider the negative case against the current state of affairs, namely, knowledge-based assessments. As a topic of inquiry and debate, the construct of knowledge has fueled great advances in scholarship, but it is not ideal for achieving the practical and normative aims of education. To mention just one shortcoming, knowledge assessments are inherently retrospective, but past knowledge is a small slice of what matters. Current knowledge assessments miss critical factors relevant to learning that include motivations to learn, responses to feedback and change, tacit understandings, and abilities to learn when no longer being told what to do.

Some readers might object that choice measurement is simply new packaging for knowledge assessment, because people's knowledge largely determines their choices. We agree that knowledge is *one* important determinant of choice, but this objection mistakes the purpose of assessment as being scientific rather than normative and practical. The scientific challenge would be to explain people's choices. However, education is first and foremost a practical matter, and as such, its lead construct of measurement should be the one closest to the realm of action. At the end of the day, whether a student has “good knowledge” will only be important to the degree that knowledge leads to good choices, so why not measure choices directly in educational assessments? In the meantime, scientific efforts can continue to see if a rational, knowledge-based account can provide a sufficient explanation of choice, which we highly doubt given the centrality of emotion in choice (see Damasio, 1994).

As we build our leverage to pry assessment from the grasp of knowledge, it may be useful to recognize that knowledge has not always been the measurement focus of assessment. Assessment in the United States has had many purposes, ranging from student tracking to individualized instruction to program evaluation to holding schools accountable (Haertel & Herman, 2005). The purposes and methods of assessment can change. Early on, assessment attempted to measure intelligence. This approach failed, in part, because it confused purportedly unchangeable individual differences with contextual sources of group variability including culture and socioeconomic status. Subsequent behaviorist approaches measured performance. These approaches emphasized the decomposition and mastery of observable skills, but they were training-oriented and too narrow to help evaluate whether students were being prepared for life outside of school. More recently, cognitive approaches have focused on assessing knowledge.

Knowledge assessments are an improvement over training and intelligence tests, because they are more flexible. Knowledge assessments assume change is possible, unlike intelligence tests. They also can examine sources of learner confusion and do not require performance on a very narrowly described set of trained tasks, unlike behavioral assessments. Despite the relative value of knowledge-based assessments, the construct of knowledge has limitations that have hampered further advances. For example, knowledge is often conceptualized as a sort of "mental text," so instructional metaphors often assume that teaching is something like transmitting the text from the mind of the instructor into the mind of the learner, much like the monks transcribed letters from one volume to the next. With choice as the central construct, it becomes harder to develop simplistic and potentially ineffective metaphors like this one.

There is also a deeper set of theoretical problems that make knowledge problematic in the context of assessment. They all stem from the isolation of knowledge: It is isolated from the bulk of social science research; it purports to describe a stable mental organization and is therefore isolated from change; as a description of a mental state, it is isolated from context; and perhaps worst of all, as an organization of information, it is isolated from the rest of the person. We detail these issues next.

Isolation from the Social Sciences

When considering the individual, most social sciences focus on choice rather than knowledge. In educationally relevant psychological research, choice is primarily treated as an independent variable: For example, does choice improve motivation (Iyengar & Lepper, 1999)? In the rest of the social sciences, choice is the dependent variable. Economics, for example, examines how financial matters drive choice and vice versa. Sociology looks at how patterns of association and structure influence choice. Management sciences examine "social selection"—how employees choose to configure their tasks and social relations. Political science and philosophy are intimately concerned with the balance of choice and necessity. In The Social Contract, Rousseau (1947/1762) puts freedom of choice as the fundamental issue: "The problem is to find a form of association… in which each, while uniting himself with all, may still obey himself alone, and remain as free as before."

This isolation of knowledge from other forms of scholarship comes at a loss to the field of assessment. For example, game theory (en.wikipedia.org/wiki/Game_theory), which examines choice behavior directly, could be a powerful source of ideas, but it has not been integrated into the discussion of assessment. Moreover, a focus on the individualized construct of knowledge makes it difficult to develop joint accounts of both individual and larger social change. Imagine that one wants to describe how changes to families' school access affects student learning. In current approaches, one would describe school access in terms of choice, while learning outcomes would be described in terms of knowledge. The ontological difference between choice and knowledge yields a black box between cause and effect. It would be preferable if school choices explained individual learning choices, so a single ontology could be used for both.

In some cases, scholars do use knowledge, or the lack thereof, to help explain the choices that people make (e.g., Tversky & Kahneman, 1974), but knowledge is only properly a means to an end. The goal in the social sciences is to account for human behavior, which is made manifest in choices. Treating knowledge as the central construct has left assessment as an isolated minority. Hot debates about constructivism mean little to the other social sciences, and we suspect that the appeal of constructivism for most educators who favor constructivist pedagogies has more to do with student agency and levels of activity than a theory of knowledge growth.

Isolation from Change and Learning

Another problem with knowledge is that it is a description of a stable mental structure. Assessment designers try to ensure that they are detecting stable knowledge and not a temporary effect by doing test-retest reliability measures. Knowledge is taken as an end or start state; it is not about change *per se*. Because the concept of knowledge is focused on descriptions of a state rather than mechanisms of state change, it is not about learning. The cognitive revolution has emphasized problem solving and, unfortunately, has largely left the learning emphasis of behaviorism behind. A perusal of the finest cognitive textbooks (e.g., Anderson, 2000) reveals scores of constructs that explain the knowledge organizations and processes that affect problem solving—schemas, priming, working memory, echoic buffers, and so on—but only a handful of constructs to explain learning, most of which emphasize memory encoding (association and compilation). 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 a posttest. But it would be more to the point to evaluate learning itself. As mentioned previously, most assessments take a sequestered problem solving (SPS) approach in which students are shielded from contaminating sources that might help them learn during the test and therefore invalidate the assessment. Tobias (2009) summarizes the knowledge researcher's anxiety when students have choices about learning, "Student choice may be a confounding variable in the discussion of instructional support and other forms of assistance." But SPS assessments of knowledge can misdiagnose the value of instruction and what students are prepared to learn.

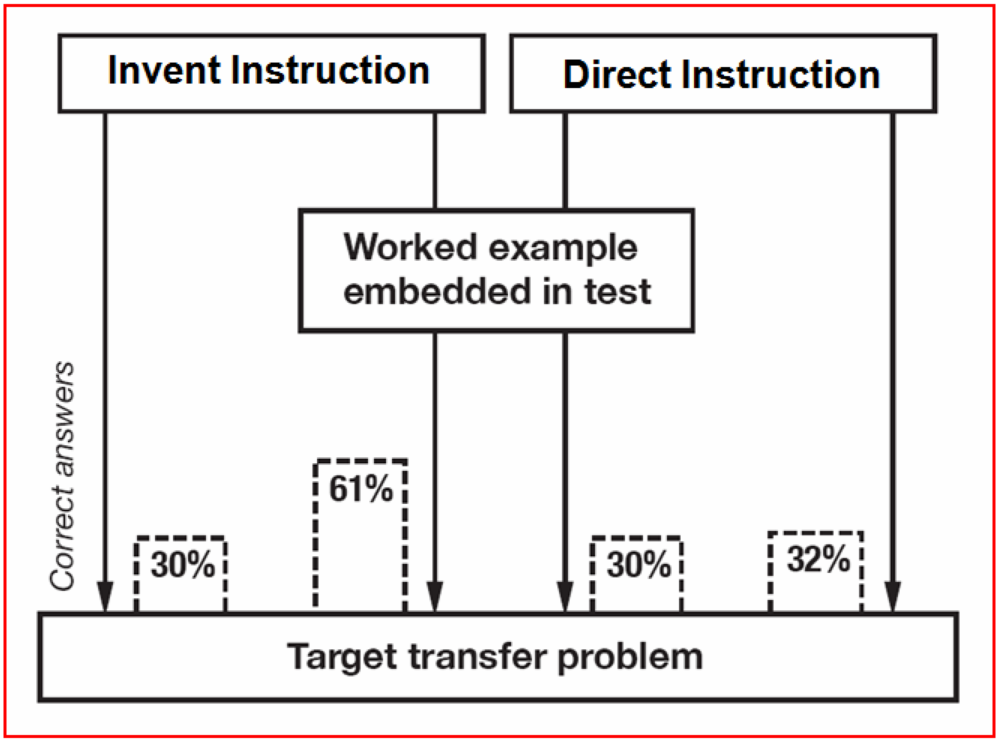

Figure 2. Improving the assessment of instruction by evaluating students' abilities to learn during a test. (Adapted from Schwartz and Martin, 2004).

In one study, Schwartz and Martin (2004) compared SPS assessments with Preparation for Future Learning (PFL) assessments. In a PFL assessment, students have an opportunity to learn during the assessment, which makes it a "dynamic assessment" (Feuerstein, 1979). In their study, students received one of two forms of instruction: direct instruction or invention. Students then took a long test, at the end of which there was a target problem. Neither form of instruction had covered this type of problem, which made it a difficult transfer problem. Half of the students were forced to attempt the target problem without any opportunities to learn during the test; this was an SPS assessment of knowledge. The other half of the students received a worked example in the middle of the test that was relevant to the target transfer problem at the end of the test. This was a PFL assessment, because it included an opportunity to learn from the worked example during the test. Figure 2 shows performance on the target problem. Students from both forms of instruction did roughly the same on the SPS version of the test. The PFL version of the test, however, showed that the invention instruction had prepared students to learn better than the direct instruction. Thus, in this case, assessments of stable knowledge were not as sensitive to the effects of instruction as were dynamic assessments of students' abilities to change and learn.

Isolation from Context

Yet another problem with knowledge, as it is typically formulated, is that knowledge is isolated from context. Knowledge is a description of the mental contents of an individual. In one extreme formulation, conceptual knowledge is thought to be improved by an increase in the abstractness of the mental representations. The logic is that more abstract knowledge can apply to a broader set of situations, because it is not tied to any single situation. According to this line of thinking, abstractness develops by a process of subtraction, so that less and less of the original context of learning appears in the knowledge. In fact, some scholars have proposed that knowledge should be taught as abstractly as possible to shortcut the deleterious effects of context specific representations (Kaminski, Sloutsky, & Heckler, 2008). Gibson and Gibson (1955) highlighted the irony of the abstraction-as-subtraction perspective. They pointed out that by this account, learning leads one farther from the world rather than closer to it, which seems absurd given that experts are much more able to perceive contextual information than are novices (e.g., a wine taster; also see Beiderman & Shiffrar, 1987).

At the core of many knowledge accounts is the idea that knowledge is a highly structured, internal representation or copy of experience—a well-organized picture or text in the head, to put it coarsely. There are alternatives to mental representation for describing competent performance. Plato, for example, proposed that understanding is like the sun: It illuminates the world rather than copying it. As a second analogy, take the case of a radio. The radio does not have a copy of the music it plays. Instead, it resonates to the context of the radio signals. If a radio could learn, it would not do so by constructing knowledge of the content it plays. Instead, it would get better at tuning more channels, separating one signal from another—in other words, learning would equate with better sensitivities for picking up and responding to contextual information.

In knowledge-based assessments, context does not receive as much attention as it should. Most of education uses "supply-side" assessments, which test students on what was supplied by the curriculum. Because supply-side assessments are confined to the curriculum context, they run the risk of producing a self-tightening knot. If students do poorly on the assessment, instruction will increasingly start to look like the assessment itself–-educators will teach to the test.

An alternative is a "demand-side" assessment. Here, the assessment is tethered to the demands of a future context rather than the past curriculum. The Programme for International Student Assessment (PISA) is a demand-side international assessment that is being increasingly used by policy makers (www.pisa.oecd.org). The items on the PISA reflect demands of the work world, but despite its merits in considering the context of application, the PISA still suffers from a lack of attention to the context created by the assessment. The PISA has a framework that specifies knowledge competencies (e.g., multi-step problem solving) and domains of application (e.g., math), and there is a good deal of heated negotiation among nations and scholars about these matters. However, construction of the specific items on the test—the contextual vehicle of the assessment—is farmed out to "item makers" and taken as non-problematic. Multiple-choice, true/false, or free-response items are all acceptable as long as they are efficient, reliable indicators of competency within the domain. If the PISA used choice as its main construct, then context could not be an afterthought, because choice does not exist independently of the decision-making context.

Isolation from the Rest of a Person

The final reason to relax the hold of knowledge on assessment is that knowledge is isolated from the individual. In the pages of *Cognitive Science*, a premier interdisciplinary journal focusing on the construct of knowledge, it is hard to find, across its history, more than a handful of articles on motivation, emotion, or identity. The fact that *Cognitive Science* partitions human performance by considering "cold knowledge" and excluding "hot affect" makes some sense. There is scientific value to analytically separating systems that nevertheless work together in nature. The study of knowledge as a separate construct has led to great advances in psychology, philosophy, and computer science, to mention just a few areas of success. However, the primary goal of assessment is the improvement of learning. To achieve this goal, it is important to include all sources of information about an individual's learning, not just cognitive markers of "cold" thinking.

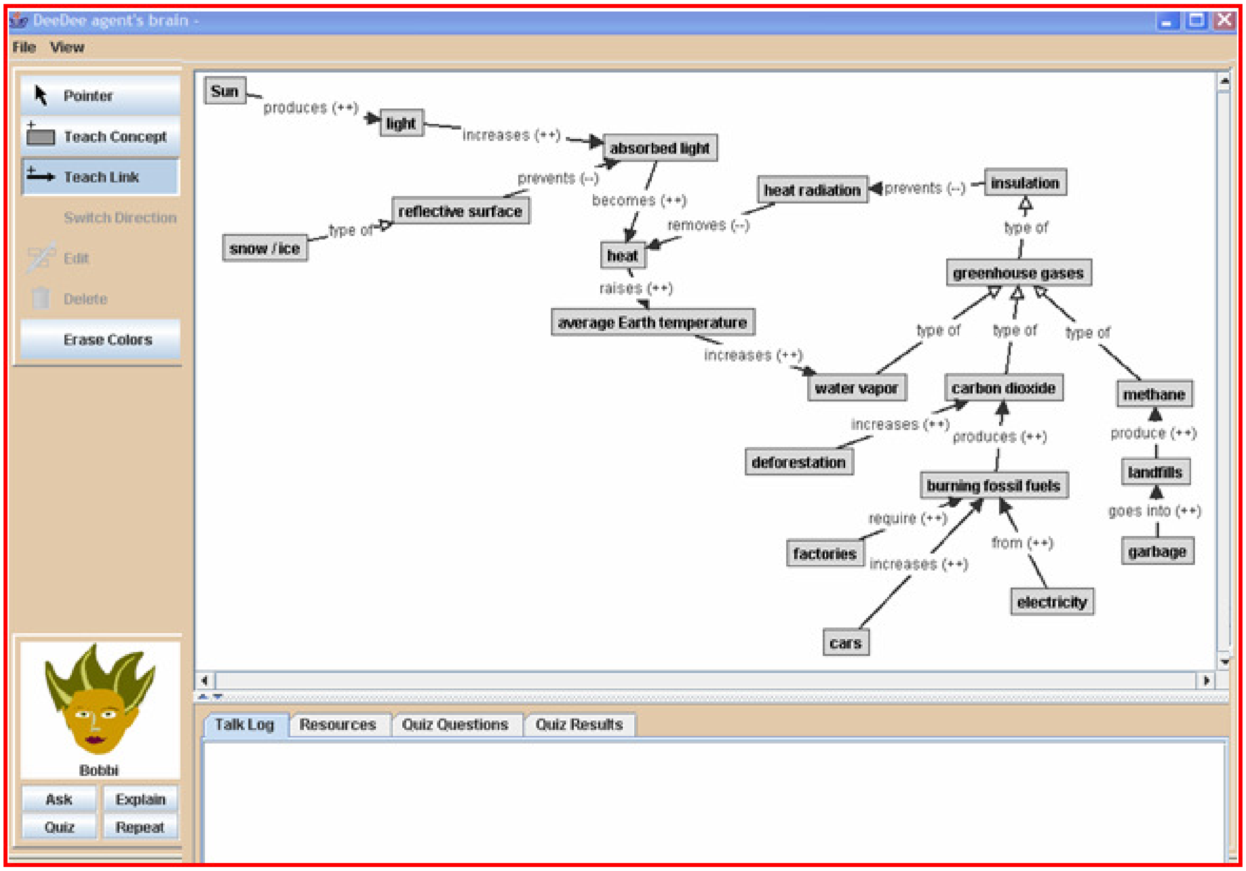

Figure 3. Students created an intelligent concept map that could answer questions. The character in the lower-left corner either represented the agent students were teaching, or it represented themselves. (Adapted from Chase, Chin, Oppezzo, & Schwartz, in press.)

Consider the case of low-achieving students. A knowledge assessment points out that they do not have strong knowledge. But ideally, an assessment would help predict what choices would lead to better learning and what contexts would help promote those choices. Chase, Chin, Oppezzo, and Schwartz (in press) conducted a study that examined learner choices. Students worked with an intelligent software environment called a Teachable Agent. In this environment students made digital concept maps under the guise of creating the Teachable Agent's brain. The digital maps were interactive, and the maps could use simple artificial intelligence techniques to answer causal questions by chaining through their links and nodes. There was also a chat feature, so students could write to each other in class, and there was a content-relevant game the students could play together online. In the Teaching condition, the students believed that by creating a concept map they were teaching a computer character (known as an agent) to answer questions. In the Self condition, the students believed they were simply making a concept map to help themselves learn (there was no cover story of the map being an agent's brain). Figure 3 shows the main teaching interface of the software.

On a posttest of learning, the Teaching condition outperformed the Self condition. When separating students based on their prior achievement, low-achieving students in the Teaching condition performed as well as the high-achieving students in the Self condition, and they did much better than the low-achieving students in the Self condition. The logs from their use of the software indicate what happened. The low-achieving Teaching students did well because they chose to spend more time working on their maps by reading relevant resources and then editing their maps' links and nodes. The low-achieving students in the Self condition spent more of their time chatting and playing the available game. It is hard to explain these differences by appealing to the low-achieving students' knowledge, either beforehand or after. Instead, the key assessment involved examining students' choices of whether and how to learn.

One might wonder what psychological states led to the choices. This is a good question that the researchers subsequently addressed, but as we mentioned previously, it is a scientific question indifferent to the assessment of learning choices. How far can assessments go by focusing on choice, without positing knowledge to explain those choices? This is an open question. The purpose of this paper is to catalyze exploration of this question so that the field of assessment can move beyond knowledge assessments that are so isolated they cannot even evaluate student motivations in the context of learning.

Choice-Based Assessments are Feasible

Thus far, we have made the argument that new technologies support new forms of interactive instruction and assessment that align with the normative mission of education. We have also argued that assessments organized around the knowledge construct are too far removed from the realm of action that education cares about, and that knowledge has inherent theoretical limitations for assessment. We now turn to the question of whether choice-based assessments are viable.

Industry has already embraced the methodology of analyzing people's choice behaviors on the Internet (Tancer, 2008). Simple analyses can determine, for example, whether people look for more complex dinner recipes on weekends than on weekdays. These analyses mine data from hundreds of thousands of users and search for relatively simple patterns of choice that have little to do with learning. To make choice assessments viable for learning, it is important to be able to work with smaller data sets, like those generated by a single user, and to analyze patterns among collections of choices rather than single clicks. To make the case that this is possible, we present some modest demonstration studies.

In the first example, we show that analyzing choices can provide access to critical aspects of learning that precede full-blown knowledge. Most assessments focus on student knowledge by asking what schema, concept, or skill set must they have in mind. But if students do not have any of the relevant knowledge, it is difficult to make much headway beyond saying they do not have the knowledge or by pointing to the "bad" knowledge or misconceptions that might be getting in the way. In the language of assessment, an item that students do not have the knowledge to answer yields very little useful information about the students' ability levels. Examining choices provides more information.

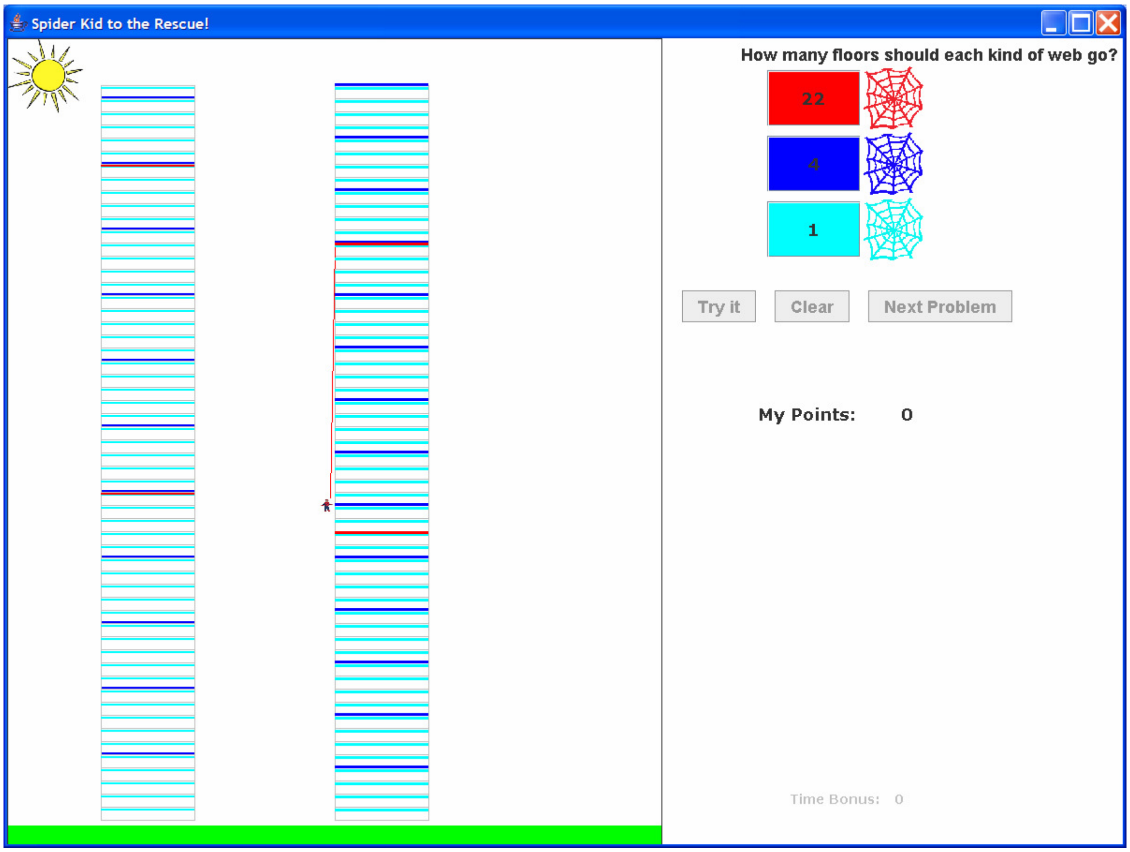

Figure 4. Screenshot of SpiderKid game. As part of a game to teach grouping, children had to learn how to make the right-hand ladder look like the left-hand ladder so they could rescue a cat in a later phase. They entered values for how far different colored webs should shoot. The little dot to the side of the right-hand ladder represents SpiderKid making red webs using the value the student entered.

Blair (2009) examined children's learning in a simulation environment intended to teach mathematical grouping. Figure 4 shows that students entered sets of numerical values in the upper-right corner. A character named SpiderKid used these values to shoot webs different distances to make something like a ladder. Students' task was to make SpiderKid's ladder look exactly like a target ladder (because later, SpiderKid would have to rescue cats from the target ladder by reaching over). The target ladder reflected the recursive structure of numerical bases (e.g., base ten).

The children were not very good at making a matching ladder. By analyzing the students' choices in response to their mistakes, Blair discovered that the children were failing because they did not perceive the feedback provided by the environment. They did not see the differences between the ladder they had created and the target ladder. The problem was not what students were doing with the information in their heads; the problem was that the information never got in there in the first place. By looking at their choices, Blair identified a stable learning progression in the children's abilities to perceive increasing amounts of structure in the feedback: (a) right/wrong information; (b) too high or too low information, (c) way too high/low or slightly too high/low information; and finally (d) precise distance and direction of the discrepancy. Students who moved through the complete progression learned the matching recursive structure, whereas students who stalled at earlier levels never did. Thus, in this case, analyzing learner choices supported assessments of what students could perceive and how this evolved, which is different from assessing what the students knew.

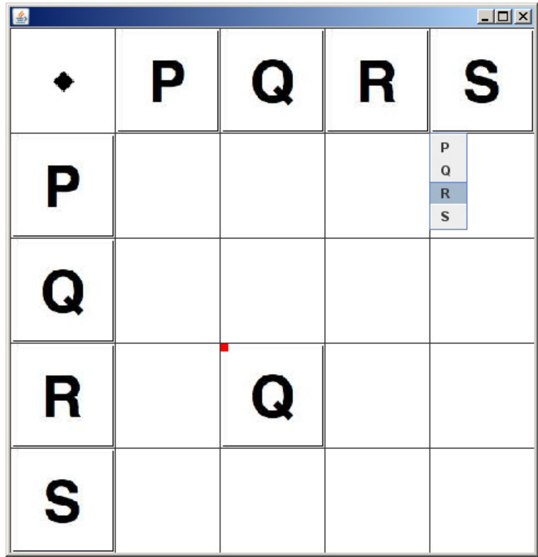

Figure 5. Task used to compare choice and knowledge assessments.

In our second example, Rafferty (2007) directly compared choice-based and knowledge-based assessments. Participants worked in the environment shown in Figure 5, which is like a multiplication table except that the diamond represents a mystery operator that uses letters instead of digits.[1] To determine the function of the mystery operator, participants guessed what belonged in each cell. They knew there was a lawful relation, and their task was to discover what it was. When participants clicked on a cell to guess its entry, their click provided choice data (i.e., which cell they chose). When they then used a pull-down menu to guess what letter belonged in the cell, their selection provided knowledge data, with more correct selections indicating higher knowledge. After completing this task, participants received a transfer task, in which they had to complete a new table that was governed by a different relation. The main

[1] The mystery operator is a form of modular multiplication, with P as 1, Q as 2, and so on. P is thus the "identity" element: P crossed with any letter yields that letter. R crossed with a letter yields "three times" that letter in modular or "clock" counting: for example, R crossed with Q yields Q because Q is 2 and R is 3, and $3 \times 2 = 6$, and 6 mod 4 = 2, which is Q.

research question of the study was whether knowledge or choices on the first task better predicted performance on the second task.

To analyze the choice data, machine learning algorithms looked for instances in which participants seemed to be trying to track down patterns. For example, if participants clicked on several cells in the P column, these choices suggested that they were systematically uncovering the identity relation. On the other hand, if participants clicked haphazardly from cell to cell, then the participant's choices indicated that they were not on a productive path. The analysis of the knowledge data was simply how many correct answers the participant produced.

Both knowledge and choice separately predicted performance on the second task, but choice was a significantly better predictor. Moreover, in a third condition, participants were not allowed to make choices about which cells to click. Instead, the system determined which cells they had to try. In this case, participants did more poorly on the second task. Thus, choice was better for learning and for assessment. This, of course, is exactly the vision for choice-based assessments, whose format encourages and tracks important forms of learning.

Rich digital environments can provide many more choices than the simple grid in the preceding example, and with more choices come more possible patterns of choice. The ability to identify informative patterns of choice will depend on advancements in data mining and machine learning techniques, which are proliferating quickly. One new machine learning technique (Li & Biswas, 2002) looks for Hidden Markov Models (HMM; Rabiner, 1989). Automated HMM analysis finds recurrent patterns of choices. Deriving an HMM is analogous to performing a factor analysis to reveal underlying structure among variables, except that in the case of HMM's the structure is underlying interaction patterns and transitions among those patterns rather than inter-correlations. These patterns of interaction are called “hidden” because they do not correspond to any specific choice or transition between states, but rather larger patterns of choice. We used HMM to analyze learner choices using the Teachable Agents software described earlier. The goal was to see whether students learn patterns of choice and whether those patterns predict future learning (Jeong et al., 2008).

When teaching their agents in the Teachable Agent environment, students can choose from among seven activities at any given time, and they can return to or leave an activity at any time. (This study predated the inclusion of the chat and on-line game features.) Figure 6 shows the result of an HMM analysis of participants’ choices. It revealed three major interaction patterns, as well as the probabilities of transitioning from one pattern to another. The top panel shows HMM analyses on two different implementations of the Teachable Agent software. The left side represents the choice behavior of children who were given corrective feedback about the quality of the map they were making (e.g., “The correct answer is that Algae decreases Oxygen”). The right side represents the choice behavior of children who were given tips on what activities they might choose to do given a mistake in their map (e.g., “Ask the agent a question, so you can see how it is figuring out the answer”). The specific meaning of the interaction patterns is not relevant to this discussion; what is important to notice is that the two treatments show strikingly different choice patterns.

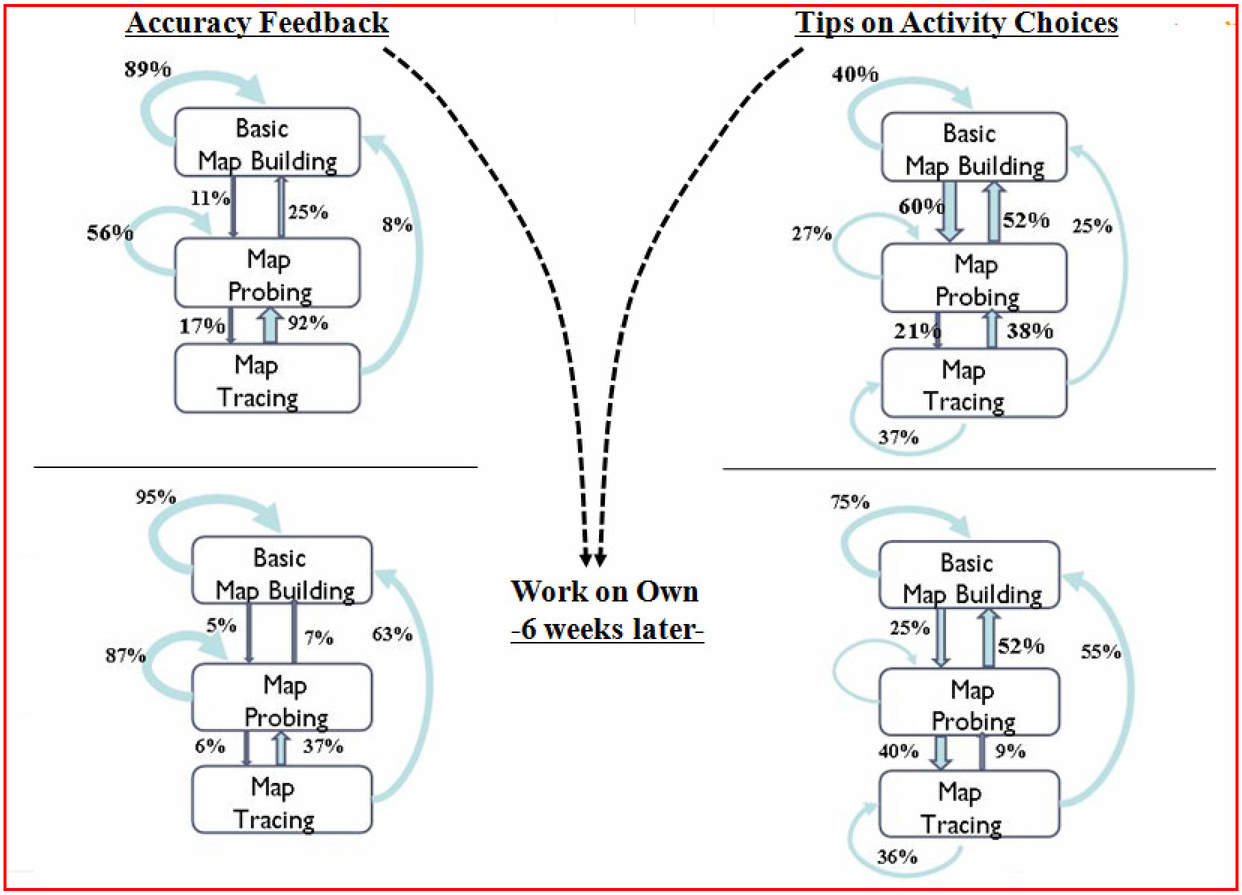

Figure 6. Analyses of choice patterns at initial learning and transfer.

The bottom panel shows the children's learning choices six weeks later, when students from both conditions learned a new topic. At this time, the corrective feedback and tip feature were turned off, so the children were using identical software to learn the new topic. As may be seen, the choice patterns from initial learning continued, even when children were working on their own using identical environments. Thus, the HMM choice analysis was able to distinguish and track the effects of the two treatments. Moreover, students who received tips on making learning choices learned more both in the initial learning period and six weeks later (as measured by standard paper-and-pencil tests of knowledge). There are many exciting directions in which future work on HMM analysis will go, but again, for our purposes the story is twofold: choice is a productive construct for assessment, and learning environments that support choice-making lead to better learning.

Towards Choice-Adaptive Learning Environments

The endgame in our vision of choice assessments is the creation of new types of digital learning environments that can both assess learning choices and support their development. We call these choice-adaptive learning environments. The term "adaptive" refers to both the learner and the environment—but not in the sense of computer-adaptive testing (in which computerized assessments can more efficiently hone in on a student's level of knowledge by constantly re-calibrating question difficulty based on the student's performance in the assessment). Instead, a choice-adaptive environment would adapt to students' choices to help guide them to better choices. At the same time, students would adapt by making choices among the available options, and thus learn to become more flexible and effective in learning. Instruction and assessment would be seamlessly coupled, providing important feedback to learners, teachers, and policy makers without putting a halt to instruction to administer assessments.

To move toward this endgame, it is important to remember that the choices we care about are learning-relevant choices; not all choices are diagnostic from the perspective of learning. Ideally, the problem of deciding what constitutes a learning-relevant choice could be addressed empirically by determining which choices, and at what level of detail, correlate with various learning outcomes. A data-driven answer would help alleviate some of the problems associated with the social construction of what constitutes a choice. Hazel Markus (personal communication) described a study where participants were led to the door of a room in which they were to fill out a short survey. The room had five substations, in each of which were five colored pencils, five sheets of paper that held the same questionnaire, and five candies. Once the participants were done, the experimenter asked how many choices they had made. East Indians said they made about one or two choices (evidently choosing among colored pencils was not a "meaningful" choice). Americans on average said they made four. One woman said she had made nine choices. For example, she had picked up one of the candies, set it down on second thought, and picked up another candy, which to her constituted three separate choices. If a choice-adaptive environment can empirically reveal which choices matter for learning, it may be possible to avoid getting buried in different conceptions of what constitutes a meaningful choice point.

It is also important to remember that some learning choices are better—more adaptive and productive—than others. This raises the difficult paradox that to help students learn to choose for themselves, it will be necessary to shape their choices, which would mean they were never "really" choosing. Callan (in press) describes a version of the paradox in his essay on choosing to be a Catholic. The dilemma is that deep Catholic faith requires that one choose to be a Catholic, but the only way one can truly choose Catholicism is if one has sufficient knowledge to understand what is being chosen. The question is: Should parents deny their children the choice of not being Catholic so that the children can gain the requisite knowledge for a future free choice on becoming Catholic? These are deep theoretical waters. But choice-adaptive learning environments are not the only educational endeavors that float on them. Education is normative, and forces larger than the learner conspire to deem what is worth learning—whether this involves knowledge or choice. The question in this paper is the practical one of whether focusing on choice will lead to better educational outcomes than will continuing to focus on knowledge. To answer that question empirically, it will be important to develop models of instruction and assessment that allow us to put the question to the test.

Designs for Enabling and Assessing Student Choice

A good choice-adaptive environment will need a specific structure that increases the chances of tracking diagnostic choices. The goal is not to unleash learners in an environment as expansive as the World Wide Web. It would be intractable to evaluate choice patterns in such an environment given the near-infinite space of possible trajectories. Instead, the goal should be a specifically designed environment for learning and assessment. One nice example comes from the IMMEX Project (Stevens & Thadani, 2007). The IMMEX Project hosts an online problem-solving environment in which students solve complex problems using various multimedia resources. Each problem set consists of several related problems, all drawing on a shared set of resources including experimental results, reference materials, and expert (or peer) advice. Students are free to use the resources however they see fit: They may choose to look repeatedly at some resources and ignore others, and they may explore the resources in any order. The students' use of resources (choices) and their correctness on questions are fed into machine learning algorithms to produce student problem-solving models. Student performances can be characterized in terms of strategies such as guessing (quick, incorrect results with little resource use), perseverating (combing through resources without achieving correct results), plodding (inefficient use of resources leading to correct results), and expert performance (using only the most useful resources to achieve correct results). The IMMEX project provides a strong example of how it is possible to capture and catalog learning choices in an open environment.

Here, we present some sketches of a more game-inspired environment. To our knowledge, this environment does not exist yet, but the rough sketch will help us present several considerations for the design of choice-adaptive learning environments. Throughout the discussion, the reader should recognize that it would only take small modifications to tilt the environment towards classroom lessons or high-stakes assessment, because in the current proposal, both are about learning choices.

The environment in our example is a virtual carnival filled with booths, places for redeeming tickets for prizes, and other characters. It is designed to connote choice to students and to include competing demands on students' attention. The carnival will have an overarching goal structure to help students make purposeful and cumulative decisions about which booths to visit. This goal structure will involve winning in "boss level" booths so students can earn the right to visit more "challenging" parts of the carnival. To find out about each boss booth, students will receive the week's carnival challenge when they enter. One week's challenge might be: "Step right up and test your skill at making a sealed tank where plants and animals can survive. The challenge is that no water or air can go in or out of the tank!" The exact resources that will be available to them in the boss booth will be unknown, so they will need to be well prepared. When students enter the boss booth, the boss will present some variation of the problem in a simulation (e.g., turtles are the animal this time), and students will need to make the sealed tank so that everything survives in the simulation. If they succeed, they will be able to "level up" to explore other areas of the carnival.

The choices in the system appear in the types of booths and activities the children will be able to explore to prepare for their boss challenge. To continue our boss-booth challenge example, one region of the carnival may be the Photosynthesis Arcade, another region might be Respiration Lane, and another might be Cotton-Candy Corner (not all areas will be equally useful for the challenge!). Students will need to choose what to learn. Learning only one thing, or learning only about an unrelated topic, before trying their luck in the boss booth would be diagnostic of their learning trajectory.

Within each region, there will be a number of booths. Several of the booths will cover similar content, but in different ways. For example, one booth in the Photosynthesis Arcade might involve playing a board game about the stages of photosynthesis, whereas another might be a reading contest on photosynthesis, and yet another may be some more constructive type of game (e.g., build a working cell). Students will also need to choose how to learn. A fact-driven booth might help with some aspects of the boss challenge, but an inquiry-based booth might help with others. The design of the system will avoid the assumption that there is a single perfect trajectory through a learning space. After all, such an assumption has not been optimal in schools thus far. Different sequences may suit different learners' needs in ways that instructional designers and learning theory have yet to predict (Murata, 2004).

The carnival will also include "tickets" (points) that students can win for prizes and so forth. Filling the environment with competing motivational and learning structures is very important. If every choice leads to the desired learning outcome or perfect motivations, then the learner's choices are not nearly as interesting, realistic, or diagnostic. Therefore, it is important to include booths that are fun but not particularly substantive—for example, an arcade game that includes drill and practice, but with nothing available to increase the understanding needed for the boss challenge. Do some children say in this booth too long at the expense of other booths?

Some booths need to be hard (but perhaps attractive because of high prize ticket potential); the things children choose to do after a failure will be very diagnostic about their learning trajectory. Do they quit? Do they perseverate in the same booth? Or do they go away to another booth to learn more and then return afterwards? We will also include activities that have no learning at all, so students can also choose when to learn and when to simply unwind with some mindless fun. As any teacher or parent can attest, this is an extremely important choice pattern to recognize.

Another important feature in the environment will be the presence of other students and computer-driven characters. Computer-driven characters might approach the student and offer some facts or tips, but not all computer characters will be equally reliable about all topics (e.g., carnies might know exactly where Respiration Lane is but might not be ideal sources of information about how fish breathe). Students will thus have the opportunity to evaluate the quality of characters' advice, as well as that of fellow students and even teachers. Students need to choose from whom to learn.

Given a good choice environment with suitable choice alternatives and good analytic techniques for extracting and categorizing patterns of choices, it will be necessary to tie choice patterns to learning outcomes. For example, if students perseverate on a cycle of failing at the boss booth and doing an arcade-style drill-and-practice booth, it would be nice to confirm that this is maladaptive for learning.

A “within-game” approach would define patterns as adaptive-for-learning if they lead to success in the boss challenge, because the boss challenge is the criterion of successful learning within the game. However, there is a large establishment built around external assessments that are separate from instruction. To make headway in convincing this establishment, it will be necessary to validate the diagnosticity of the choice patterns against standard types of knowledge assessments. For example, we might compare students who do well and poorly on a standard knowledge test. We could then map backwards to see what types of choice patterns predict the differences in knowledge performance. If these differences in choices adequately predict differences in boss-level success and also on other knowledge assessments, then the choice patterns would be providing evidence that they could *replace* the knowledge assessments. Eventually, it might be possible to remove these external knowledge assessments entirely given enough evidence that assessments integrated into the environment are working—and some of that evidence already exists. For example, in one study, Schwartz et al. (in press) found that automated assessments integrated into a learning environment could predict performance on different knowledge tests of the same content (e.g., multiple-choice format, short-answer format) better than those knowledge tests could predict each other.

Designs for Instructional Decision Making

On the assumption that it will be possible to track different patterns of learning choices and tie these to learning outcomes, there is the question of how to guide students to make more effective choices. There is very little evidence on this question. We do know the types of questions that foster deeper understanding given a single learning resource, such as a text passage (Chi, de Leeuw, Chiu, & LaVancher, 1994; Graesser & Person, 1994), but that is different from suggesting to people what resources to use and for what purpose. When should the environment implicitly shape choices by, say, making an over-visited booth closed for repairs? When should the environment be a bit more explicit, such as by having a computer character stroll by the student's character to exclaim, "Hey, everybody is having a lot of success at the chlorophyll booth"? And when should the system model choice-relevant thinking by having a student's own character think aloud, e.g., "We seem to be having trouble understanding oxygen; maybe we should head over to the fish tank to watch that 3 p.m. demonstration of how a candle affects a mouse in a jar"?

To get a handle on this type of question, there is a need for new types of research. One useful line of work might examine how gaming experts guide novices, which forms of guidance are more and less helpful, and whether different forms of guidance are better or worse for different game contexts. For example, teaching a novice how to play a twitch-response game with an obvious goal and simple controls might be done best with a brief explanation and lots of opportunity to practice. Teaching a novice how to play an open-ended strategy game with complex rules and a complicated interface, on the other hand, might require a more expansive explanation prior to diving into game play. A complex simulation that targets "enduring" understandings of key physics concepts may best be taught with structured walk-through to familiarize the novice with basic controls and other low-level technical details, along with a period of monitoring during which the expert might guide the novice's attention to salient features of the simulation. These, of course, are speculations. Because most instruction has been based on a pre-determined linear sequence with impoverished learning resources, there is little evidence on how to guide learner's choices productively in an open learning space without undermining the opportunity to learn how to make good learning choices.

The magnitude of the investigations that will be necessary can appear daunting. When high-stakes assessments are built, each new item is vetted extensively, often by administering it to thousands of participants. Yet here we are, proposing the design of a new style of assessment and an accompanying framework for learner guidance without the benefit of extant empirical literature! Formidable though this endeavor will surely be, it is possible to make headway without having to build a full digital choice environment (such as our carnival) at the outset. Reasonable first steps could use a "booth" made in a small room in the real world. The room could have a dictionary, encyclopedias, hands-on experimentation activities, computer simulations, the Internet, and so on. It would be possible to track learner macro-choices by sight and explore what types of suggestions seem to drive better learning choices in this simple context.

There are, however, clear benefits to using digital technologies. Some of these already exist and could be used to good effect with minimal effort, such as websites that permit the rapid acquisition of user data for a small fee (e.g., Mechanical Turk; www.mturk.com). On these websites, visitors receive a few pennies for completing an online activity, and it is not hard to get thousands of user tests in a single evening. A common practice is for designers or researchers to post different versions of the same advertisement to see which position of a button or message leads to more "click-throughs" from users to a target site. It is not hard to imagine that a slice of a choice-adaptive learning environment could be posted; it would be easy to compare, for example, how three different types of guidance affect whether users choose to go read more resources on the respiratory needs of living things. Alternatively, one might work on refining the internal structure of a booth or even the booth poster so that students have a sense of what they will do and learn inside. The point here is that small, tractable steps can start the ball rolling toward choice-adaptive learning environments—and the prospect of leveraging new technologies to develop rapid methods for continual improvement of assessments and instruction toward that goal has a good deal of appeal.

21st Century Assessments

Of late, there has been a good deal of discussion about the skills and competencies needed for the 21st century. Proposals range from increased creativity to improved social skills, as well as many others. The catalyst for generating lists of 21st century skills comes from a realization that times have changed and will continue to do so. There is a vision of a future filled with rapid changes in work, communication, global interdependence, technology, and ideally learning. In this future, individuals' abilities to adapt, as well as to produce the rapid changes in the first place, will largely be a function of their abilities to make effective learning choices. Therefore, we think that "increasing abilities to make good learning choices" would make a fine contribution to the list of desired 21st century skills. *Learning choices* is inherently a dynamic construct that suits the realization that people will need to continue to learn and adapt. This construct also emphasizes the significance of individual agency. The goals of education, in addition to contributing to society at large, should contribute to individuals' abilities to effectively execute "productive agency" (Schwartz, 1999) to realize and achieve their own goals.

There are many subcategories that would fall under the umbrella of learning choices. Seeking formative feedback would be one of the most important learning choices, and it requires action in the world rather than a characterization of knowledge in the head. Seeking out formative assessment opportunities would be right behind the ability to make productive choices about when and what to learn. When cast at this abstract level, it is critical to recognize that increasing one's abilities to make good learning choices is not the same thing as simply learning to learn. The literature on learning to learn referred to domain-general study skills (e.g., Olivier & Bolwer, 1996). In all of our examples, the choice assessments were designed to detect the effects of prior, domain-specific learning experiences. Whether providing and emphasizing learning choices will develop domain-general learning abilities is an empirical question that is yet to be tested, because there are so few examples where learning choices are the norm within education.

The list of 21st century skills and competencies needs to be actionable if it is meant to do anything other than sort students by their abilities. The skills have to be amenable to instruction and assessment. Thus, there is some value to addressing the question of what can be taught and assessed. The answer to this question depends on the nature of instruction and assessment and, more generally, designed learning experiences. Given new digital possibilities, the nature of instruction and assessment will likely change in the 21st century, and with it, what people will be able to learn. Perhaps more than a new list of skills and competencies, what is needed are new theories of learning appropriate to a future that demands high adaptation rather than routine efficiency. Moving to choice as an assessment of competence rather than using knowledge or skills may provide one good trajectory for transforming learning theory to suit the 21st century.

Acknowledgments

The development of this paper and its ideas were supported by a grant from the MacArthur Foundation to James Gee.

References

Anderson, J. R. (2000). Cognitive psychology and its implications (5th ed.). New York: Worth.

Appelbaum, E., & Batt, R. (1994). The New American Workplace. NY: ILR Press.

Benner, C. (2002). Work in the new economy: Flexible labor markets in Silicon Valley. Oxford: Blackwell Publishing.

Biederman, I., & Shiffrar, M. M. (1987). Sexing day-old chicks: A case study and expert systems analysis of a difficult perceptual learning task. Journal of Experimental Psychology: Learning, Memory, and Cognition, 13, 640 – 645.

Blair, K. (2009). The neglected importance of feedback perception in learning: An analysis of children and adults' update of quantitative feedback in a mathematics simulation environment. Unpublished dissertation. Stanford University, CA.

Bransford, J. D. & Schwartz, D. L. (1999). Rethinking transfer: A simple proposal with multiple implications. In A. Iran-Nejad & P. D. Pearson (Eds.), Review of Research in Education, 24, 61–100. Washington, D.C.: American Educational Research Association.

Callan, E. (in press). Why bring the kids into this? McLaughlin and Anscombe on religious understanding and upbringing. Faith in education: A tribute to Terence MacLaughlin. London: London Institute of Education Publications.

Chase, C., Chin, D.B., Oppezzo, M., & Schwartz, D.L. (In press). Teachable agents and the protégé effect: Increasing the effort towards learning. Journal of Science Education and Technology.

Chi, M.T. H., de Leeuw, N., Chiu, M-H., & LaVancher, C. (1994). Eliciting self-explanations improves understanding. Cognitive Science, 18, 439-477.

Damasio, A. R. (1994). Descartes Error: Emotion, Reason and the Human Brain. NY, NY: G.P. Putnam's Sons

Feuerstein, R. (1979). The dynamic assessment of retarded performers: The learning potential assessment device, theory, instruments, and techniques. Baltimore, MD: University Park Press.

Gee, J.P. (2003). What video games have to teach us about learning and literacy. New York: Palgrave/Macmillan.

Gibson, J. J., & Gibson, E. J. (1955). Perceptual learning: Differentiation or enrichment. Psychological Review, 62, 32 – 51.

Graesser, A. C., & Person, N. K. (1994). Question asking during tutoring. American Educational Research Journal, 31 (1), 104-137.

Haertel, E. H., & Herman, J. L. (2005). A historical perspective on validity arguments for accountability testing. In J. L. Herman & E. H. Haertel (Eds.), Uses and misuses of data in accountability testing. Yearbook of the National Society for the Study of Education (Vol. 104, Part I, pp. 1 – 34). Boston, MA: Blackwell Publishing.

Hatano, G. & Inagaki, K. (1986). Two courses of expertise. In H. Stevenson, H. Azuma, & K. Hakuta (Eds), Child development and education in Japan (pp. 262-272). NY: Freeman.

Iyengar, S. S., & Lepper, M. R. (1999). Rethinking the value of choice: A cultural perspective on intrinsic motivation. Journal of Personality and Social Psychology, 76, 349-366.

Jeong, H, Gupta, A., Roscoe, R., Wagster, J., Biswas, G. & Schwartz, D. L. (2008). Using hidden Markov models to characterize student behaviors in learning-by-teaching environments. Proceedings of the 9th International Conference on Intelligent Tutoring Systems, 614-625.

Kaminski, J. A., Sloutsky, V. M., & Heckler, A. F. (2008). Learning theory: The advantage of abstract examples in learning math. Science, 320, 454 – 455.

Koedinger, K.R. and Anderson, J.R. (1997). Intelligent Tutoring Goes To School in the Big City. International Journal of Artificial Intelligence in Education, 8, 30-43.

Li, C., & Biswas, G. (2002). A Bayesian approach for learning hidden Markov models from data. Scientific Programming, 10, 201 – 219.

Murata, A. (2004). Paths to learning ten-structured understanding of teen sums: Addition solution methods of Japanese Grade 1 students. Cognition and Instruction, 22, 185 – 218.

Olivier, C., & Bolwer, R. F. (1996). Learning to learn. NY, NY: Fireside.

Rabiner, L.R. (1989). A tutorial on hidden Markov models and selected applications in speech recognition. Proc. IEEE, 77, 257 – 286.

Rafferty, A. N. (2007). Using FACT to Challenge Assumptions: Frequency, Accuracy, Choice, and Timing in Machine Learning. Unpublished Masters Thesis. Stanford University, Stanford, CA.

Reeves, B. & Read, L. (in press). Total Engagement. Cambridge, MA: Harvard Business Press.

Resnick, D.P., & Resnick, L.B. (1994). Performance assessment and the multiple functions of educational measurement. In M. Kane & R. Mitchell (Eds.), Implementing performance assessment: Promises, problems and challenges (pp. 23 - 38). Mahwah, NJ: Erlbaum.

Roediger, H. L., III, & Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. Psychological Science, 17, 249-255.

Rousseau, J.J. (1947/1762). The social contract, Or principles of political right (C. Frankel, Trans.). New York: Hafner.

Schwartz, D. L. (1999). The productive agency that drives collaborative learning. In P. Dillenbourg (Ed.), Collaborative learning: Cognitive and computational approaches (pp. 197-218). NY: Elsevier Science.

Schwartz, D. L., & Martin, T. (2004). Inventing to prepare for learning: The hidden efficiency of original student production in statistics instruction. Cognition & Instruction, 22, 129 – 184.

Schwartz, D. L., Bransford, J. D., & Sears, D. (2005). Efficiency and innovation in transfer. In J. P. Mestre (Ed.), Transfer of learning from a modern multidisciplinary perspective (pp. 1 – 52). Greenwich, CT: Information Age Publishing.

Schwartz, D. L., Chase, C., Wagster, J., Okita, S., Roscoe, R., Chin, D., & Biswas, G. (in press). Interactive metacognition: Monitoring and regulating a teachable agent. To appear in D. J. Hacker, J. Dunlosky, and A. C. Graesser (Eds.), Handbook of Metacognition in Education.

Stevens, R.H., & Thadani, V. (2007). Quantifying student's scientific problem solving efficiency and effectiveness. Tech., Inst., Cognition and Learning, 5, 325 – 337.

Tancer, B. (2008). Click: What millions of people are doing online and why it matters. New York: Hyperion.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. Science, 185, 1124 – 1131.

Vygotsky, L. S. (1934/1987). The collected works of L. S. Vygotsky. (Eds). R. Rieber & A. Carton. NY: Plenum.