Research Report

Monitoring and Fostering Learning Through Games and Embedded Assessments

Valerie J. Shute

Matthew Ventura

Malcolm Bauer

Diego Zapata-Rivera

December 2008

ETS RR-08-69

Listening. Learning. Leading.®

Monitoring and Fostering Learning Through Games and Embedded Assessments

Valerie J. Shute[1]
Florida State University, Tallahassee, FL

Matthew Ventura, Malcolm Bauer, and Diego Zapata-Rivera
ETS, Princeton, NJ

December 2008

As part of its nonprofit mission, ETS conducts and disseminates the results of research to advance quality and equity in education and assessment for the benefit of ETS's constituents and the field.

ETS Research Reports provide preliminary and limited dissemination of ETS research prior to publication. To obtain a PDF or a print copy of a report, please visit:

http://www.ets.org/research/contact.html

Abstract

To reveal what is being learned during the gaming experience, this report proposes an approach for embedding assessments in immersive games, drawing on recent advances in assessment design. Key to this approach are formative assessment to guide instructional experiences and evidence-centered design to systematically analyze the assessment argument (including the claims about the learner and the evidence that supports or fails to support those claims). Elements of this approach that have been applied in a nongame setting are shown and ideas are discussed for applying the approach to an existing immersive game setting. Finally, the report offers suggestions for extending and applying this approach for existing games and the design of new ones.

Key words: Games, evidence-centered design, formative assessment

Acknowledgments

We thank Gary Bente, Eric Hansen, Irv Katz, Jody Underwood, and Dan Eignor for their ideas and editorial suggestions.

Table of Contents

List of Figures

Page

What children do for fun and what they are required to do in school constitutes a wide gap. Generally, children are motivated by fun activities (e.g., interactive, entertainment games), while they might not be by schoolwork. Numerous studies have shown that student engagement is strongly associated with academic achievement (e.g., Fredricks, Blumenfeld, & Paris, 2004; Fredricks & Eccles, 2006; Finn & Rock, 1997). One solution for potentially increasing learning, particularly for disengaged students who are not performing as well as they could, is to combine games with schoolwork.

This report describes a strategy with a two-stage approach to address the methodological problems[2] involved in combining games with schoolwork. The focus of this report is the first stage, which details a method for using games to extract information from students during game play that is relevant to learning. This could help validate the claim that, during the course of play, important knowledge and skills are learned. Success in the first stage will inform the second, which entails adapting existing games or designing new ones to monitor and support students' learning skills relevant to schoolwork.

The goal of this report is to present a methodological approach for extracting data relating to valued educational constructs while sustaining the students' engagement. The main assumptions here are that: (a) learning by doing improves learning processes and outcomes, (b) different types of learning can be verified and measured during game play, (c) a student's strengths and weaknesses can be capitalized on and bolstered, respectively, and (d) formative feedback can be used to further support student learning. An additional goal is to help students to consider knowledge and skills as important currencies in gaming. Ultimately (i.e., within Stage 2 of the research—outside of the scope of this report), the data obtained from the stealth assessment can be used to inform changes to the gaming environment to support student learning and also to inform the creation of new games. This report first defines serious games and embedded (or stealth) formative assessment and then illustrates how they can be joined through evidence-centered design (ECD; Mislevy, Steinberg, & Almond, 2003) and Bayesian networks (e.g., Pearl, 1988) to monitor and support learning. ECD permits embedding assessments in the gaming environment, enabling unobtrusive collection and analysis of data. This is shown in two contexts: (a) an ECD-based simulation developed for training network administrators in the Cisco Networking Academy Program (CNAP; Bauer, Williamson, Mislevy, & Behrens, 2003), and (b) a fairly well-known immersive game, *The Elder Scrolls IV: Oblivion* (Bethesda

Softworks, 2006), which was used in this study to elicit evidence about current and emergent cognitive and noncognitive attributes.

Definitions

Serious Games

Virtual environments that have to goal of educating or training are called *serious games*. Such diverse groups as the U.S. military and the National Association of Home Builders invest in games that represent and instruct their particular content and views (Squire, 2006). As players become immersed in game-playing activities, serious games impart their contents. An example of a serious game is *America's Army*, a game used by the U.S. Armyfor recruitment. The game shows players what it's like to be a soldier in the U.S. Army.

Another way to understand serious games is in contrast to more typical digital games that have no explicit goals about being educational or informational—such as *Dance Dance Revolution* (Konomi Corporation, 2008) and *Diner Dash* (Gamelab, 2008). The *raison d'être* of such casual games is to entertain. In contrast, and according to Carey (2006), serious games (as well as educational simulations, like physics or chemistry simulations) represent a unique product category with functional requirements that are different from casual games. Two key features of serious games are: educational and immersive. Casual games are typically not viewed as educational, but they can be immersive.

Players may experience immersion within a virtual world because of features such as interactive stories that provide context and clear goal structures for problem-solving in the game environment. Researchers have noted that features that are common to all intrinsically motivating environments include elements of challenge, control, and fantasy to pique curiosity and engage attention (Lepper & Malone, 1987; Malone, 1981; Rieber, 1996). These characteristics all work together to induce what is commonly called *flow*, defined as the state of optimal experience, where a person is so engaged in the activity that self-consciousness disappears, sense of time is lost, and the person engages in complex, goal-directed activity not for external rewards, but simply for the exhilaration of doing (Csikszentmihalyi, 1990).

The aim here is to identify what players do and learn within immersive games, specifically immersive games that are *not* explicitly educational. While these games are not by definition serious games, the purpose of this report is to describe how learning and assessment can be accomplished in immersive games that have the *potential* for being educational.

Immersive games are the focus here because they have the greatest potential for inducing and sustaining flow (i.e., finding the perfect spot between too hard and too easy; see Csikszentmihalyi, 1990 for more). Along the same lines, Pausch, Gold, Skelly, and Thiel (1994) described the essence of video game design as: (a) presenting a goal, (b) providing clear-cut feedback to the user as to their progress towards the goal, and (c) constantly adjusting the game's challenges to a level slightly beyond the current abilities of the player. Similarly, Rieber (1996) contended that challenge must be matched to the player's current skill or ability level. The degree to which there is a mismatch, frustration or boredom may ensue.

Embedding assessments within such immersive games permits a player's current level to be monitored on valued competencies and then that information to be used as the basis for adjusting game features, such as the difficulty of challenges. This is intended to maximize both the *flow* and *grow* - of learning. Integrating the flow state of immersive games with learning theories has tremendous potential to enhance students' learning—both in the short- and long-term (e.g., Gee, 2003; Lieberman, 2006; Squire & Jenkins, 2003). The idea is to exploit animation and immersive characteristics of game environments to create the flow needed to keep the students engaged in solving progressively more complex learning tasks. In other words, this study used the flow to facilitate the grow in terms of students' acquisition of valued proficiencies.

As more and more researchers are pointing out (e.g., Cannon-Bowers, 2006; de Freitas & Liver, 2006; Squire, 2006), there is currently a shortage of experimental studies examining learning through game play, despite the fact that games represent a very rich venue for conducting learning research. For practical purposes, and in line with the ideas presented in this report (i.e., to leverage immersive games to support learning), exactly what players are taking away needs to be ascertained from games such as *Grand Theft Auto IV* (Rockstar Games, 2008) and *Civilization IV* (Firaxis Games, 2008). Gee (2003), Lieberman (2006), and others in the field firmly believe that a lot of important learning and development is going on within these games. But are these educationally valuable skills and strategies? As mentioned, many immersive games are intrinsically motivating, likely because they employ such features as challenge, control, and fantasy, as well as opportunities for social interaction, competition, and collaborative play (Malone, 1981; Malone & Lepper, 1987).

Immersive games can potentially have adverse effects, such as players acquiring undesirable attitudes or learning mal-adaptive social behaviors. This occurs due to the freedom enabled in immersive games.

The next section covers embedded formative assessments (FAs), which have the potential to improve student learning directly (e.g., via feedback on personal progress) or indirectly (e.g., through modifications of the learning or gaming environment). In this context, *embedded* refers to assessments that are unobtrusively inserted into the curriculum (or game). Their *formative* purpose is to obtain useful and accurate information about student progress, on which the teacher, instructional environment, and/or the student can act.

Embedded Formative Assessment

> If we think of our children as plants... summative assessment of the plants is the process of simply measuring them. The measurements might be interesting to compare and analyze, but, in themselves, they do not affect the growth of the plants. On the other hand, formative assessment is the garden equivalent of feeding and watering the plants - directly affecting their growth. Clarke (2001, p. 2)

When teachers or computer-based instructional systems know how students are progressing and where they are having problems, they can use that information to make real-time instructional adjustments such as re-teaching, trying alternative instructional approaches, altering the difficulty level of tasks or assignments, or offering more opportunities for practice. This is, broadly speaking, formative assessment (Black and Wiliam, 1998a), and it has been shown to improve student achievement (Black and Wiliam, 1998b; Shute, Hansen, & Almond, 2007).

In addition to providing teachers with evidence about how their students are learning so that they can revise instruction appropriately, formative assessments (FAs) may directly involve students in the learning process, such as by providing feedback that will help students gain insight about how to improve. Feedback in FA should generally guide students toward obtaining their goal(s). The most helpful feedback provides specific comments to students about errors and suggestions for improvement. It also encourages students to focus their attention thoughtfully on the task rather than on simply getting the right answer (Bangert-Drowns, Kulik, Kulik, & Morgan, 1991; Shute, 2007a). This type of feedback may be particularly helpful to lower-achieving

students because it emphasizes that students can improve as a result of effort rather than be doomed to low achievement due to some presumed lack of innate ability (e.g., Hoska, 1993).

A more indirect way of helping students learn via FA includes instructional adjustments that are based on assessment results (Stiggins, 2002). Different types of FA data can be used by the teacher or instructional environment to support learning, such as diagnostic information relating to levels of student understanding, and readiness information indicating who is ready or not to begin a new lesson or unit. FAs can also provide teachers or computer-based learning environments with instructional support based on individual student (or classroom) data. Examples of instructional support include: (a) recommendations about how to use FA information to alter instruction (e.g., speed up, slow down, give concrete examples), and (b) prescriptions for what to do next, links to web-based lessons and other resources, and so on.

Conjoining Games and Embedded Assessments

New directions in educational and psychological measurement allow more accurate estimations of students' competencies, and new technologies permit one to administer formative assessments during the learning process, extract ongoing, multi-faceted information from a learner, and react in immediate and helpful ways, as needed. When embedded assessments are so seamlessly woven into the fabric of the learning environment that they are virtually invisible, this is called *stealth assessment*. Stealth assessment can be accomplished via automated scoring and machine-based reasoning techniques to infer things that would be too hard for humans (e.g., estimating values of evidence-based competencies across a network of skills).

One big question is not about collecting this rich digital data stream, but rather, how to make sense of what can potentially become a deluge of information. Another major question concerns the best way to communicate student-performance information in a way that can be used to easily inform instruction and/or enhance learning.

Our solution to the issue of making sense of data and thereby fostering student learning within gaming environments is to extend and apply evidence-centered design (ECD; e.g., Mislevy, Steinberg, & Almond, 2003). This provides (a) a way of reasoning about assessment design, and (b) a way of reasoning about student performance whether in gaming or other learning environments.

The Methodology

There are several problems that must be overcome to incorporate assessment in serious games. Bauer, Williamson, Mislevy, and Behrens (2003) addressed many of these same issues with respect to incorporating assessment within interactive simulations in general. Here several of the issues are outlined and an example of how they may be addressed using ECD.

There are many factors that may influence learning in games and simulations. Are immersive games more engaging than more typical venues such as lectures, textbooks, and even serious games? If so, does simply providing a more engaging environment (and hence increasing time on task) produce increased learning outcomes? Can one provide richer learning experiences and new venues for learning that could not be explored otherwise? Consider, for instance, the prospect of learning by playing out what-if scenarios in history, such as through the games *Civilization III* or *Revolution* (for more scenarios, see Squire & Jenkins, 2003).

Two good reviews of studies that have been conducted with games' effects on learning outcomes include the dissertation of Blunt (2006) and a recent chapter by Lieberman (2006). However, compared to other types of instructional environments, there are currently too few experimental studies examining the range of effects of immersive environments and simulations on learning. For instance, Cannon-Bowers (2006) recently challenged the efficacy of game-based learning. Furthermore, of the evaluation studies that have been conducted, the results of games and simulations effects on learning are mixed. For example, Kulik (2002) reported that a recent meta-analysis of six studies of classroom use of simulations found only modest learning effects and two of the six studies could not find any increase in learning at all. In addition, research on the use of simulations to enhance students' understanding of physics has also yielded mixed results (e.g., Ranney, 1988).

In playing games, students naturally produce rich sequences of actions while performing complex tasks, drawing upon the very skills to be assessed (e.g., critical thinking, problem-solving). Evidence needed to assess the skills is thus provided by the students' interactions with the game itself—the *processes* of play, which may be contrasted with the *product(s)* of an activity, as is the norm within educational settings. Making use of this stream of evidence to assess skills and abilities presents problems for traditional measurement models used in assessment. First, in traditional tests the answer to each question is seen as an independent data point. In contrast, the individual actions within a sequence of interactions in a simulation or

game are often highly dependent on one another. For instance, what one does in a flight simulator at one point in time affects subsequent actions later on. Second, in traditional tests, questions are often designed to get at one particular piece of knowledge. Answering the question correctly is evidence that one knows a certain fact (i.e. one question—one fact). By analyzing students' responses to all of the questions, each providing evidence about students' understanding of a specific fact or concept, teachers or instructional environments can get a picture of what students are likely to know and not know overall. Because typically a range of skills and abilities needs to be assessed from evidence coming from students' interactions within a game or simulation, methods for analyzing the *sequence* of behaviors to infer these abilities are not as obvious.

ECD is a method that can address these problems and enable the development of robust and valid simulation- or game-based learning systems.

Evidence-Centered Design

A game that includes stealth assessment must elicit behavior that bears evidence about key skills and knowledge, and it must additionally provide principled interpretations of that evidence in terms that suit the purpose of the assessment. Figure 1 sketches the basic structures of an evidence-centered approach to assessment design (Mislevy, Steinberg, & Almond, 2003).

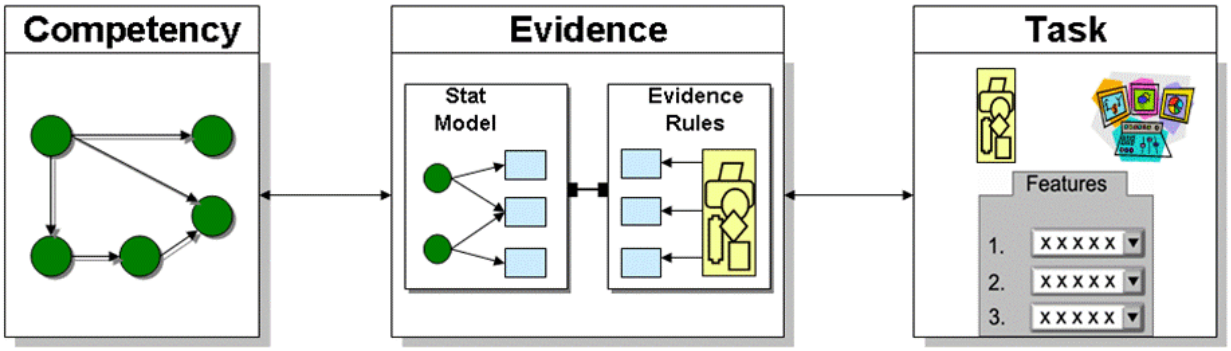

Figure 1. **The central models of an evidence-centered assessment design.**

Working out these variables and models and their interrelationships is a way to answer a series of questions posed by Sam Messick (1994) that get at the very heart of assessment design:

- *What complex of knowledge, skills, or other attributes should be assessed?* A given assessment is meant to support inferences for some purpose, such as a licensing decision, provision of diagnostic feedback, guidance for further instruction, or some combination. Variables in the competency model (CM) describe the knowledge, skills, and abilities on which the inferences are to be based. The term *student model* is often used to denote a student-instantiated version of the CM. That is, values in the student model express the assessor's current belief about a student's level on variables within the CM.
- *What behaviors or performances should reveal those constructs?* An evidence model expresses how the student's interactions with, and responses to a given problem constitute evidence about student-model variables. Observables describe features of specific task performances.
- *What tasks or situations should elicit those behaviors?* Task-model variables describe features of situations that will be used to elicit performance. A task model provides a framework for characterizing and constructing situations with which a student will interact to provide evidence about targeted aspects of knowledge.

In games with stealth assessment, the student model will accumulate and represent belief about the targeted aspects of skill, expressed as probability distributions for student-model variables (Almond & Mislevy, 1999). Evidence models will identify what the student says or does that can provide evidence about those skills (Steinberg & Gitomer, 1996) and express in a psychometric model how the evidence depends on the competency-model variables (Mislevy, 1994). Task models will express situations that can evoke required evidence.

An Example of Embedding Assessment in a Simulation

Bauer et al. (2003) described a simulation and assessment system developed for the Cisco Networking Academy Program (CNAP). Based on the needs of CNAP, an online simulation-based training system with stealth assessment was designed and developed. The system uses realistic scenarios to set the stage for authentic design, configuration and troubleshooting tasks that are provided via Flash simulations and remote access to actual computer networks. The system is used by students to practice networking skills, and students receive detailed feedback on their performance on each problem. The system also accumulates evidence, via stealth

assessment and gleaned from students' performances across tasks, to estimate their overall skills and abilities. The simulation environment was structured to support learning, based on accepted psychological principles including active construction of knowledge, use of multiple representations, performance on realistic complex tasks, and support for abstraction and reflection.

This section describes the competency, evidence, and task models within the interactive simulation and assessment design to provide a concrete example of how the ECD methodology works. The CM in Figure 2 represents the constellation of knowledge, skills, and abilities that are important for success as a student of Cisco's networking academy. The CM was generally developed to support the claims that instructors would like to make about the skills their students have. It was specifically developed on the basis of a cognitive task analysis, a pre-existing job-task analysis of computer networking professionals, and judgments of subject matter experts. The CM was structured to reflect the dependencies among competencies in the domain.

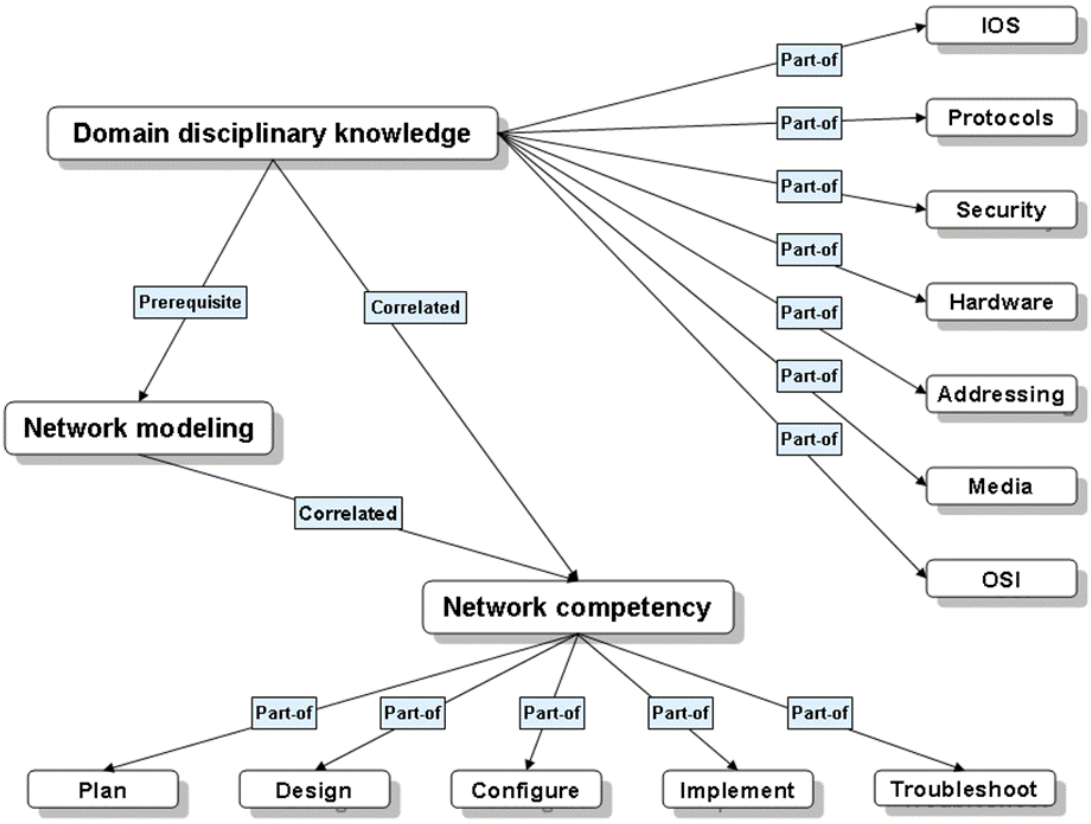

Figure 2. **The competency model (conceptualization).**

As shown in Figure 2, the CM is composed of a number of variables representing aspects of knowledge, skill, and ability. The Domain Disciplinary Knowledge variable represents the declarative knowledge of network components and operation. There are a number of elements of declarative knowledge that are part of Domain Disciplinary Knowledge, such as addressing schemes, hardware components of a network, media, protocols, etc. The Network Competency variable represents the overall networking ability including the sub-skills of planning, designing, configuring, implementing, and troubleshooting a network. As each of these network activities requires some declarative knowledge in order to conduct the procedures required to perform these tasks, there is a modeled relationship between the declarative knowledge represented in Domain Disciplinary Knowledge and the procedural knowledge and skills required for Network Competency. The Network Modeling variable is the ability of the student to explain and predict the behavior of a network. Experts identified this skill as a key to the highest levels of skill in Network Competency; hence the two variables have a link between them. The ability to produce a model of a network requires Domain Disciplinary Knowledge, which is therefore represented as a prerequisite of Network Modeling ability.

The evidence model describes what specific behaviors or *observables* are indicative of different levels of skill in the CM. On the basis of the results from a cognitive task analysis, the statistical portion of the evidence model is constructed by positing CM variables to be parents of observables, which are meant to bear evidence about their (inherently unobservable) values. Table 1 presents an outline of several evidence model observables used to update the CM variables for Design, Implement, and Troubleshoot. The italicized composite variables are included in probabilistic models (i.e., Bayes net objects; see Koller & Pfeffer, 1997) as observable variables. Their values are summaries of the nonitalicized features listed below them, along the lines of Clauser et al. (1995).

For each of these features, an algorithm was written to score the student's work product to identify, evaluate, and summarize the quality of the work product in that aspect. For example, in Table 1, under the heading *Troubleshoot*, the sequence-of-targets observable provides evidence of students' fault-locating behaviors. The log files of students' command sequences are parsed to determine the search pattern. That is, data are examined to see if the student (a) immediately visits the device on which there is a fault, (b) systematically searches devices, rarely (or never) returning to a previously-visited network device, or (c) unsystematically ping-pongs

among the devices, visiting many again and again. The different patterns are associated with different levels of competency.

Table 1

Examples of Observables in the Evidence Model

Design	Implement	Troubleshoot
Correctness of outcome	*Correctness of outcome*	*Correctness of outcome*
Functionality of design	*Correctness of procedure*	Error identification
Core requirements	Efficiency of procedure	Error over-identification
Peripheral requirements	Help usage	*Correctness of procedure*
	IOS syntax	Efficiency of procedure
	Volume of actions	Help usage
	Procedural sequence	IOS syntax
		Volume of actions
		Procedural sequence
		Sequence of actions
		Sequence of targets

All of the observables from a given scenario are modeled as conditionally dependent (Mislevy, Steinberg, Breyer, Almond, & Johnson, 2002).[3] These observables are used to update the student model and provide summary feedback to students and teachers. The features of the student work products on which the observables are based also contain more detailed information about students' performance on the task on which they are currently working, and used in providing task-level feedback. Hence the same evidence that is accumulated to make estimates of students' knowledge, skills, and abilities is also used, in a more detailed and timely manner for instruction in the form of task-level feedback. To illustrate, the following represents actual task-level feedback given to a student after attempting to solve a difficult design task (*Create Network Diagram*):

- Check your diagram. You have forgotten a networking device or placed a networking device in the wrong location.
- Check your diagram. You are missing a connection between two networking devices.

- You have configured an incorrect IP address or you have left off an IP address.

The question now is whether this type of stealth assessment approach, employed in a simulation as described above, can similarly be used within immersive gaming environments. This question is examined in a case study involving a popular immersive game called *Oblivion*.

Application of the ECD Approach Using a Highly Immersive Game

Over the past 15 years, the gaming market has exploded due mainly to advances in software and computer technology. With the advent of this new technology, sophisticated graphics engines can now display breathtaking graphics of landscapes, humans, and other real world and fantasy environments. Additionally, advances in artificial intelligence have enabled challenging environments that require players to adopt dynamic strategies for success. Finally, millions of dollars now get invested in creating complex plots and problems requiring hours of time to solve. All of these components set the stage for highly immersive game play.

The purpose of this case study is to test the viability of our approach within an existing immersive game and to identify knowledge, skills, and abilities that may be learned during game play. Gee (2003) has asserted that the secret of an immersive game as an engaging teaching device is not its 3D graphics but its underlying architecture. Each level seeks to be hard enough to be just doable. Similarly, cognitive psychologists (e.g., Falmagne et al., 2003; Vygotsky, 1987) have long argued that the best instruction hovers at the boundary of a student's competence.

The case study that follows describes the typical game play of *Elder Scrolls IV: Oblivion* (Bethesda Softworks, 2006). This popular game is a first person role-playing game set in a 3D medieval world. The user can choose to be one of many characters (e.g., knight, mage, elf), each of whom possesses various strengths and weaknesses. Each character also has (or can obtain) a variety of weapons, spells, and tools. The primary goal of the game is to gain rank and complete various quests in a massive land full of castles, caves, virtual characters, monsters, and animals. There are multiple mini quests along the way, and a major quest that results in winning the game. Players have the freedom to complete quests in any order they choose. Quests may include locating a person to obtain information, eliminating a creature, retrieving a missing item, or finding and figuring out a clue for future quests.

Character Skill Modification (Persistence)

There are many character skills to improve in *Oblivion*, and each skill improvement is frequency based, evidenced by the number of successful actions in relation to the particular skill. For instance, successfully hitting creatures with a sword in combat will increase the skill of *blade* over time. Additionally, successfully convincing someone to talk to you will increase the skill of *speechcraft*, which defines the probability that a stranger will respond to you in conversation in the future.

To improve these skills and thus gain rank requires many hours of game play, and many hours of game play implies *persistence*. This involves sticking with some activity both in the face of success and failure. Each time a player successfully engages some activity, the frequency and hence probability of subsequent success in the future is increased. In education, the attributes of persistence and self discipline have been shown to significantly predict students' academic achievement—both in the near- and far-term (e.g. Dweck, 1996; Duckworth & Seligman, 2005).

Quest Completion (Problem-Solving)

There are over one hundred quests in *Oblivion*. The key challenge in these quests is to stay alive and to defeat creatures that try to harm you. For instance, during the course of game play, a player can contract vampirism while exploring caves around the land. In order to find a cure for vampirism, one must find a witch who will then provide information regarding key ingredients needed to make a potion for a cure. Each key ingredient is then marked on the map which is used by the player to travel around in order to obtain the ingredients. Since the player has vampirism, many new obstacles enter into the quest. For example, as a vampire, one cannot travel during the day without dying (with certain exceptions), and the level for the attribute charisma decreases, which leads to difficulty in conversing with people, and so on.

Problem-solving (which can range from simple to complex) plays a key role in quests since the player has to figure out what to do and how to do it (e.g., locate pertinent information that will provide clues to carry out a current quest). In the case of contracting vampirism, one must determine how and where to obtain information concerning a cure. In addition to problem-solving skills, the player's background (or folklore) knowledge is often helpful (e.g., knowing about likely places to find useful information, such as within chapels, from mages, etc.). This knowledge may be acquired over time with the game, or transferred from other games of this type.

In education, problem-solving is often viewed as the most important cognitive activity in everyday and professional contexts (e.g., Hiebert et al., 1996; Jonassen, 2000; Reiser, 2004). However, learning to solve complex problems is too seldom required (or rewarded) in formal educational settings. As with persistence, the assessment and support of problem-solving skills are vitally important to improve students' long-term learning potential.

Combat (Attention and Multitasking)

Combat scenarios represent one way to keep the user engaged in game play. In *Oblivion*, combat requires the user to attend to several factors: health, magic level, fatigue, enemy maneuvering, enemy health, and escape plan. Like many games in general, and combat games in particular, concentration and attention play key roles in success. Additionally, there are many heuristics that can be used to more easily defeat particular creatures. The player must be aware of which creatures pose a serious threat (i.e., those who inflict massive amounts of health damage) and which ones can be easily defeated. In many cases, retreat is an option which enables a more strategic combat plan for difficult creatures.

In education, the central role of attention in learning has been clearly demonstrated for decades (e.g., Kruschke, 2001; Nosofsky, 1986; Trabasso & Bower, 1968). One of the main benefits of gaming environments is that they tend to capture and sustain attention. Thus attention represents another educationally valuable variable.

Other Learning Components

Reading. Since much of *Oblivion* involves interaction with other people, reading and listening skills are essential to success in quests. Additionally, there are many books that give clues to quests and recipes for potions.

Creativity. There are many ways to solve a quest or defeat enemies in *Oblivion*. This freedom allows players to be creative in how they advance in the game. For example, if the player needs to obtain an object to aid in a quest, one can steal the object, buy the object, or persuade someone to relinquish the object. Each choice has various advantages and disadvantages.

Figure 3 illustrates some possible educationally-relevant competencies that might be assessed during game play in *Oblivion*. This CM, with its cognitive and noncognitive variables, should be viewed as illustrative only.

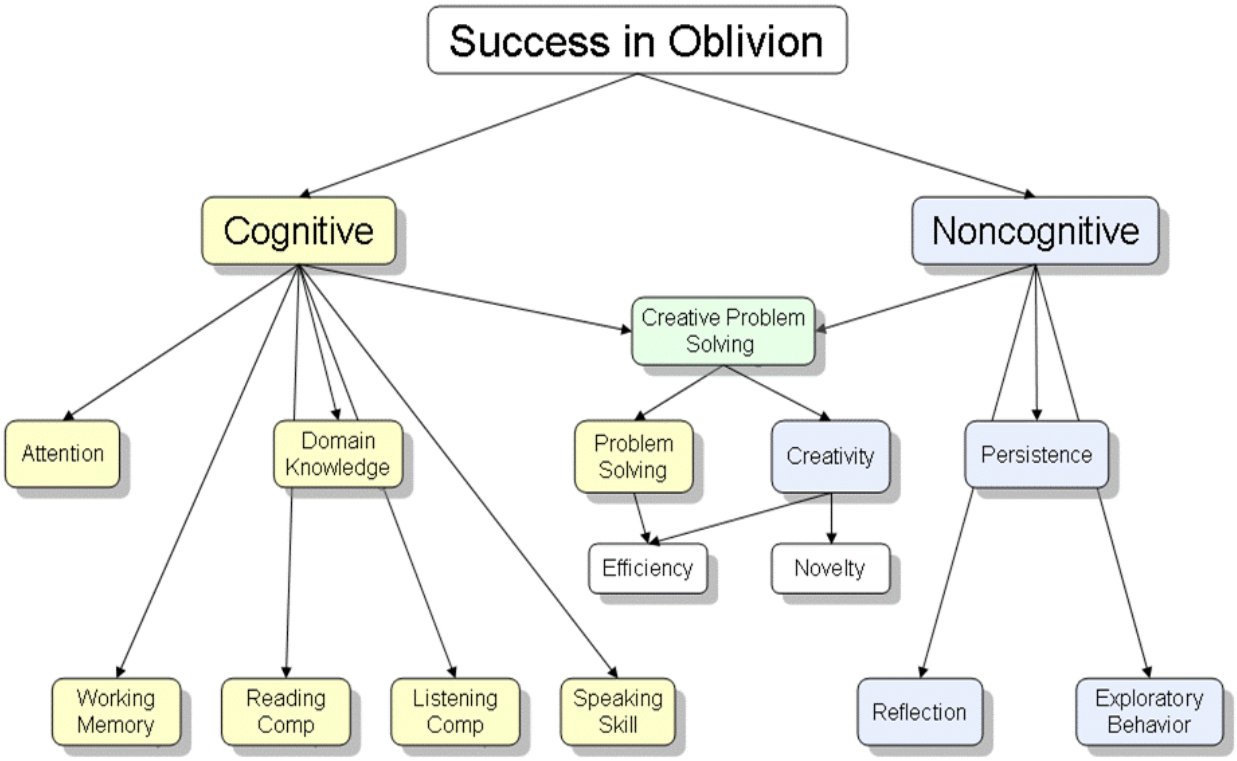

Figure 3. **Illustration of a competency model for success in the game *Oblivion*.**

To show how stealth assessments can be created for one of the competencies cited above using an ECD approach, the next section focuses on the attribute labeled *creative problem-solving*.

Illustrating the Stealth Assessment Idea

Creative problem-solving can be viewed as the aggregate of two abilities: creativity and problem-solving. Creativity is a mental process involving the generation of new ideas or concepts, or new associations between existing ideas or concepts. The products of creative thought are usually considered to have both originality (novelty) and appropriateness (relevance). However, while creativity has been studied from many different perspectives (e.g., cognitive science, artificial intelligence, philosophy, history, design research, social psychology, management, and so on), there is no single, authoritative definition of creativity, nor is there any standardized measurement technique. Problem-solving generally refers to higher-order cognitive processes invoked to advance from an initial state to a goal state. And like creativity, problem-solving has been studied extensively (see Newell & Simon, 1972), in areas as diverse as mathematics, political science, writing, and game playing.

When these two constructs are put together, creative problem-solving (CPS) can be defined as the mental process of *creating a solution to a problem*. It is a special form of problem-solving in which the solution is independently created rather than learned with assistance. CPS always involves creativity, but creativity often does not involve creative problem-solving (e.g., in the arts). Creativity requires novelty as a characteristic of what is created, but does not necessarily imply that what is created has value or relevance. Thus to qualify as CPS, the solution must be relevant and clearly solve the stated problem (Sternberg, 2006). Solving school-assigned homework problems does not involve creative problem-solving because such problems usually have well-known solutions.

Conceptual Framework for Creative Problem-Solving

Whereas creativity can be seen in the products, it can also be considered in terms of processes. For example, Weisberg (1986) proposed that creativity can be defined by the novel use of tools to solve problems. Given the importance of relevance in CPS, creative contributions should be defined in some context (Sternberg, 1999). If an individual's CPS ability is judged within a context, then it will help to understand how the context interacts with how the person is judged. In particular, what are the types of creative contributions a person can make within a given context? Most theories of creativity concentrate on attributes of the individual, but to the extent that creativity is in the *interaction* of person and context, one would need as well to concentrate on the attributes of the person and his or her work relative to the environmental context—like the gaming environment.

Based on the work of Sternberg (1999), this report adopts a notion of CPS that is measured within a context—as defined through a particular scenario or quest within a game. By focusing our definition of creativity to problem-solving, one can assess novel and *efficient* contributions toward goals. Figure 4 shows a fragment of the ECD models for this CPS variable. Notice that competency model and evidence model are the same terms used in our previous ECD example, but this report uses the term *action model* instead of task model. Action model reflects dynamic modeling of students' *action sequences*. These action sequences form the basis for drawing evidence and inferences and may be compared to simpler task responses as with typical assessments. Finally, note that *scene* is used to define a particular quest in the game.

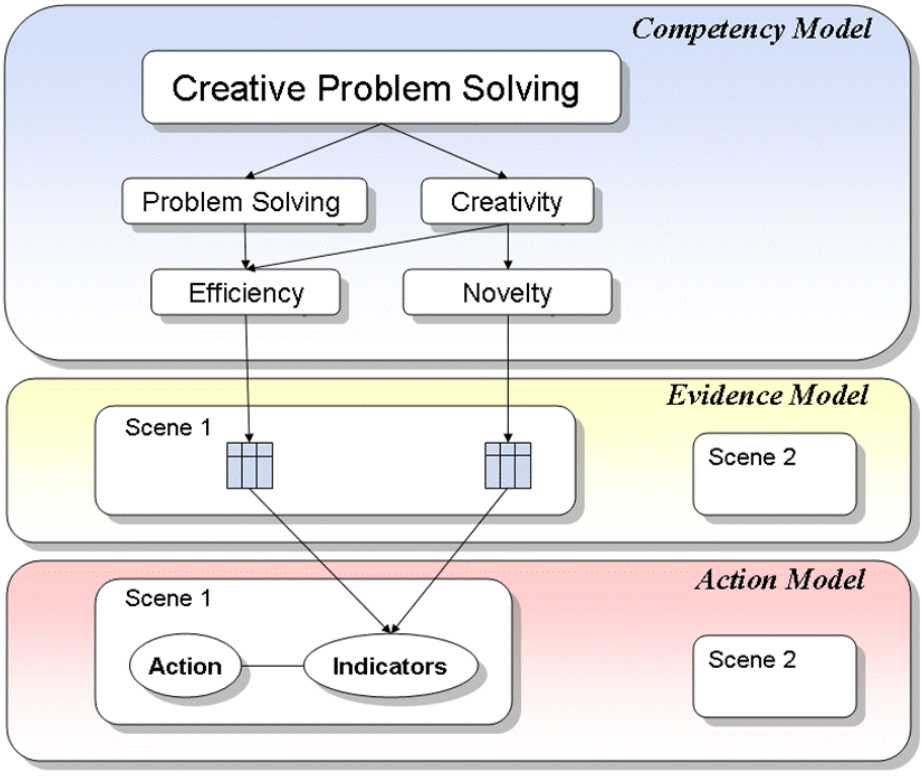

Figure 4. **Evidence-centered design (ECD) models (conceptualization) applied to games.**

Competency model: As shown earlier in Figure 3, problem-solving and creativity are joined to form the creative problem-solving competency. Efficiency is shown as informing both problem-solving and creativity, but novelty only informs creativity in this model. Novelty is defined in relation to choosing less common (i.e., low frequency) actions in the solution of problems, while efficiency is defined in relation to the quantity and quality of steps taken toward a solution. Both novelty and efficiency are constrained by relevance. That is, the problem-solving space per scene is limited to only those actions explicitly linked or relevant to the particular problem or quest.

Evidence model: The evidence model defines the connections between specific observables and their underlying competencies—novelty and efficiency. These connections are represented as little distribution tables within Scene 1 of the evidence model in Figure 4. In particular, the evidence model includes: (a) scoring rules for extracting observables from students' game play indicators found in log files, (b) the observables (i.e., scored data), and (c) measurement rules for accumulating evidence from the observables, which are then used to update the student model variables. For simplicity, our illustration includes just two observables,

each informing either novelty or efficiency. Both of these, in turn, inform the CPS variable through intermediate variables (i.e., problem-solving and creativity). The degree to which variables differentially inform their parent nodes is represented in a Bayes net (discussed in the next section, and illustrated in Figure 5).

Action model: The action model is similar to the task model in ECD, but it has been modified in this study for use in existing games to define particular sequences of interactions from which to extract the observables. Interactions consist of actions and their specific indicators. An action represents anything a player does within the context of solving a particular problem (contained within a scene), such as crossing a river and exploring a cave. Each action that a player takes to solve a given problem may be characterized along two dimensions: novelty and efficiency, illustrated in more detail in the next section. A list of indicators is explicitly linked to each action. These are the things that can be directly measured and reside within the player's log file.

For players in immersive gaming environments such as *Oblivion*, their performance can be monitored across many and varied problems and quests in terms of particular constructs. To assess the latent construct of creative problem-solving, indicators of actions can be defined for, say, efficiency and novelty, which are ultimately combined into a general estimate of creative problem-solving.

Creative Problem-Solving Instantiation

To illustrate how this methodology would actually work inside of a game (*Oblivion*), each of the ECD models (competency, evidence, and action) were implemented using a Bayesian network approach.

Consider the problem of attempting to cross a raging river full of dangerous fish in *Oblivion*. Table 2 contains a sample list of actions one can take to solve this problem, as well as the indicators that may be learned from real student data, or elicited from experts. For the system to learn the indicators from real data, estimates of *novelty* may be defined in terms of the frequency of use across all players. For instance, swimming across the river is depicted as a high-frequency, common solution, thus associated with a low *novelty weight*. An estimate of *efficiency* may be defined in terms of the probability of successfully solving a problem given a set of actions. To illustrate, swimming across the river is associated with a low efficiency weight because of the extra time needed to evade the piranha-like fish that live there. On the other hand,

digging a tunnel under the river to get to the other side is judged as highly novel, but less efficient than, say, freezing the water and simply sliding across; the latter being highly novel and highly efficient. The indicator values shown in Table 2 were obtained from two *Oblivion* experts, and they range from 0 to 1. Higher numbers relate to greater levels of both novelty and efficiency.

Table 2

Example of Action Model With Indicators for Novelty and Efficiency

Action	Novelty	Efficiency
Swim across river filled with dangerous fish	$n = 0.12$	$e = 0.22$
Levitate over the river	$n = 0.33$	$e = 0.70$
Freeze the water with a spell and slide across	$n = 0.76$	$e = 0.80$
Find a bridge over the river	$n = 0.66$	$e = 0.24$
Dig a tunnel under the river	$n = 0.78$	$e = 0.20$

Actions can be captured in real-time as the player interacts with the game, and associated indicators can be used to provide evidence for the appropriate competencies. Again, this is accomplished via our evidence model. Figure 5 shows a Bayesian model (using Netica software) linking evidence indicators (i.e., *ObservedEfficiency* and *ObservedNovelty*) to various competencies. Note that Figure 5 represents an instantiation of our ECD conceptual framework (see Figure 4). That is, the upper five nodes (boxes) show a fragment of our competency model for CPS. The bottom two nodes represent a simple evidence model linking actions to competencies via their associated probability distributions. Each node has two or more discrete states (e.g., low and high). Marginal probabilities are presented for each state. The lower two evidence-model nodes represent continuous variables that have been discretized into four states, ranging from 0 to 1, that will be used to model the actions depicted in Table 2. The same Bayesian model can be used to illustrate a variety of actions in the game.

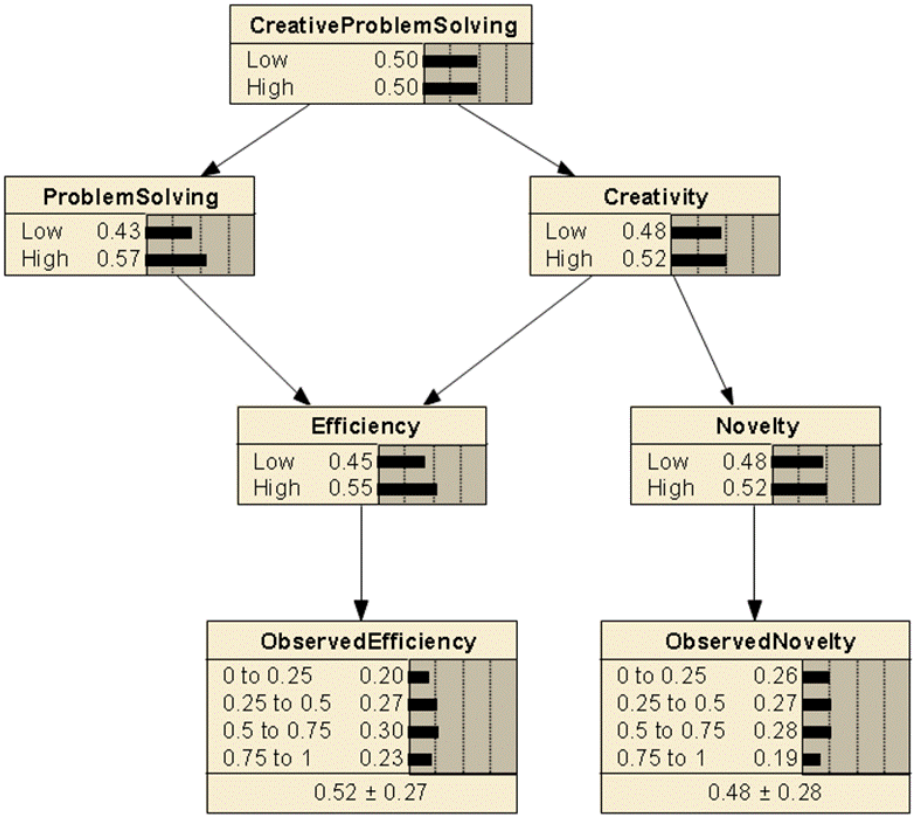

Figure 5. **Bayesian model used to instantiate our evidence-centered design (ECD)-based conceptual framework.**

Prior and conditional probabilities can be elicited from experts and refined using players' data. In our case, conditional probability tables for *ObservedEfficiency* and *ObservedNovelty* have been initialized based on a normal distribution whose parameters can be eventually adjusted using real data. Means and standard deviations are shown at the bottom of each observable box.

The general model in Figure 5 can be used to illustrate, through various actions, how the Bayesian model integrates evidence from particular cases. First, suppose a player chose to cross the river by digging a tunnel under it. As noted earlier, this represents an action that is classified as low in efficiency (e = 0.20; linked to the lowest of four discrete states for ObservedEfficiency) and high in novelty (n = .78; linked to the highest state for ObservedNovelty). This evidence is added to the model shown in Figure 5 and propagated throughout the CM producing a new model with updated marginal probabilities for competency nodes and observed states for evidence nodes presented in Figure 6. Some of the marginal probability values are shown below while the full range of probability values are shown in Figure 6.

- Pr(Efficiency = High | evidence) = 0.14
- Pr(Novelty = High | evidence) = 0.98
- Pr(Creativity = High | evidence) = 0.89
- Pr(ProblemSolving = High | evidence) = 0.36
- Pr(CPS = High | evidence) = 0.40

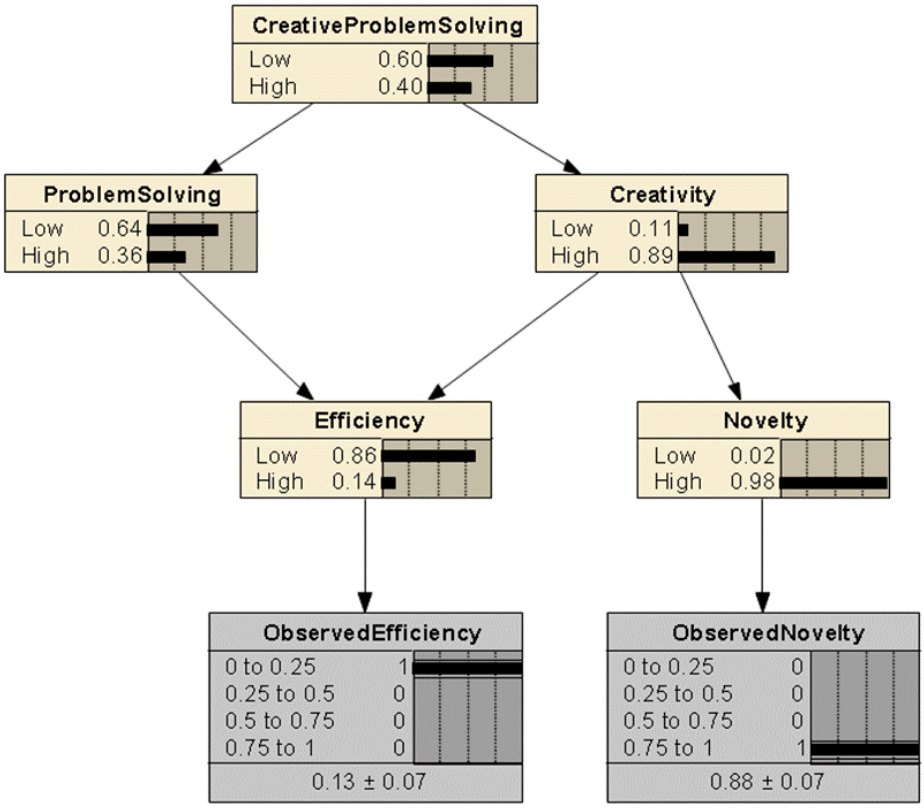

Figure 6. **Bayes model depicting marginal probabilities after observing a low efficiency and high novelty action such as crossing the river by digging a tunnel under it.**

Even though the player evidenced very high novelty in her solution, the parent node of CPS is still inferring that she is more low than high on this attribute—illustrating that efficiency is a more valued competency than novelty, based on the way the CM was set up.

Our second case is shown in Figure 7 where a player has successfully used a magical spell to freeze the river and slide across it. This action is associated with high efficiency and high novelty levels, resulting in the following marginal probability values:

- Pr(Efficiency = High | evidence) = 0.98
- Pr(Novelty = High | evidence) = 0.99
- Pr(Creativity = High | evidence) = 0.97
- Pr(ProblemSolving = High | evidence) = 0.88
- Pr(CPS = High | evidence) = 0.82

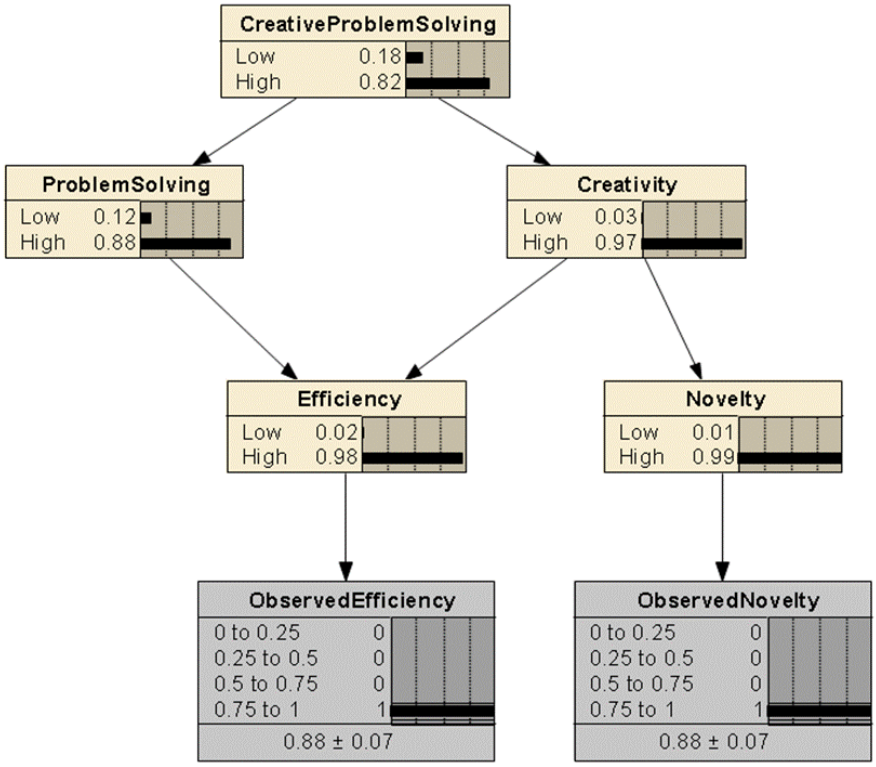

Figure 7. **Bayes model depicting marginal probabilities after observing a high efficiency and high novelty action such as freezing the river and sliding across it.**

These two cases illustrate that different actions taken within *Oblivion* can be used to infer quite different levels of CPS, which could be used to inform teaching and learning—the grow part of the story, and described as part of the next steps.

Next Steps

By extending the example described in this report, it is possible to build actual (as opposed to illustrative) ECD models for the various competencies shown in Figure 3 that (a) are presumed to have educational value, and (b) may be monitored via stealth assessment during game play with *Oblivion*. The justification for modeling creative problem-solving is generally critical to success in many real world settings (e.g., school, business, the military). Stealth assessment within serious games offers the opportunity to inform and support a wider variety of knowledge, skills, and thinking needed for the 21st century.

Additionally, there are numerous and valuable constructs that cannot be measured except in complex immersive games like *Oblivion*. For instance, many of the novel problem-solving tasks that have been studied in the past (e.g., Tower of Hanoi) do not have the external validity found in immersive games. In *Oblivion*, the task of finding objects in the environment matches obstacles one would find in searching for objects in the real world (i.e., using focused attention coupled with heuristic search strategies). Data collected by measuring progress in these types of problems yields a richer source of information that can be used in formative feedback to ultimately improve learning. While the learning that can occur in this stealth assessment approach has not yet been mapped, the concept of dynamic feedback in game play lays the initial groundwork for such a framework. More work is needed to decide how dynamically changing the game play itself can best accommodate the proficiency levels of players. Currently *Oblivion* enemies do become more difficult to defeat as the player gains rank (i.e., an approach to keep the game from actually getting easier) but no one has yet investigated how these changes in game difficulty can actually lead to increased learning of valued constructs. Its obvious extensions to learning can be investigated by developing a framework of dynamic stealth assessment.

Finally, the ideas presented in this report can be applied in future studies to another game to show proof-of-concept and generalizability of the approach. If that exercise is successful, the next step would be to use players' data (log files) to inform decisions concerning the adaptation of game play—such as increasing or decreasing challenges, introducing new characters, and so on. Ideally, and in subsequent projects, ECD would be employed to design games from scratch, in conjunction with game designers. This is because the fit between many current immersive games and education is not very good—particularly given objectionable content in many games, such as violence and sex. If the essential elements in games that induce flow can be identified,

learning indicators can be efficiently and effectively culled from series of actions, and the information used to support learning, one can design valid (and more suitable) immersive games. Squire et al. (2005) have begun the process of identifying such design features and analyzing emergent learning.

Conclusions

Students in the United States, particularly those who are disadvantaged, are not learning adequately (Shute, 2007b). For instance, students in 25 out of the 30 most developed countries in the world outperformed U.S. students in mathematics problem-solving (Lemke et al., 2004). U.S. students' problem-solving skills have to be strengthened so that the United States can compete effectively at international and national levels. Noting that *top* U.S. students do compare favorably (or at least comparably) to their non-U.S. counterparts, Kirsch et al. (2007) contended that student engagement can help close the achievement gap.

In this report, an ECD-inspired approach was presented to address these educational challenges by harnessing the potential of immersive games. This approach comprises these steps:

- Specifying educationally valuable competencies believed to contribute toward successful game play
- Defining evidence models that link game behaviors to the competencies
- Updating the student model at regular intervals

In this approach, ECD models were retrofitted to an existing game. This has certain implications, including the need to gather valid assessment information without interfering with the flow of the game (i.e., the engaging features). The study used Bayesian models to monitor actions, integrate evidence on players' performance, and update the student model in relation to emerging competencies. Bayesian models can also support learning through the generation of progress reports for various educational stakeholders (e.g., teachers, students, parents). For example, teachers could use the reports to recommend specific activities, while students could use them to improve a particular skill.

The system can also use data about student competencies to select new gaming experiences. More challenging quests, for example, could be offered to students who have high CPS abilities. Current estimates of student competencies that are based on assessment data

handled by the Bayesian nets also can be integrated into the game and displayed as progress indicators. Players would see how their competencies change based on their game performance. Games such as *Oblivion* already have status bars that represent the player's current levels of such things as magic, health, and fatigue. By clicking one of these bars, which are in the lower left part of the screen, players see more information on a particular variable, such as spells and potions they have. Competency bars that represent attributes such CPS could be added to these games. Users would click a competency bar to see such states as efficiency, novelty, and problem-solving. If any competency bar gets too low, the student has to do something to increase the value. Allowing players to see game- or learning-related aspects of their state could enhance their metacognitive processes and help them gain more awareness of personal attributes. The literature calls these types of models *open student models*, and shows how they support knowledge awareness, reflection, and learning (Bull & Pain, 1995; Hartley & Mitrovic, 2002; Kay 1998; Zapata-Rivera & Greer, 2004; Zapata-Rivera et al., 2007).

To conclude, within the storyline of a well-designed game, learning occurs naturally. Seamlessly aligning the lesson with the story, however, is not trivial (Rieber, 1996). The approach for addressing this problem in this study entails first examining existing games to help determine what kind of activities support learning and then using that knowledge to develop new games. While these new games should be just as engaging as their predecessors, they would be based on research from three fields: (a) artificial intelligence, (b) cognitive science, (c) educational measurement. These games would also include assessments to monitor students' cognitive and noncognitive abilities accurately over time and to support learning by adjusting the game environment. This report presented the first methodological step towards harnessing student engagement induced by flow to promote learning of valuable and life-long skills.

References

Almond, R. G., & Mislevy, R. J. (1999). Graphical models and computerized adaptive testing. *Applied Psychological Measurement, 23*(3), 223–237.

Bangert-Drowns, R. L., Kulik, C. C., Kulik, J. A., & Morgan, M. T. (1991). The instructional effect of feedback in test-like events. *Review of Educational Research, 61*(2), 213–238.

Bauer, M., Williamson, D., Mislevy, R., & Behrens, J. (2003). Using evidence-centered design to develop advanced simulation-based assessment and training. In G. Richards (Ed.), *Proceedings of world conference on e-learning in corporate, government, healthcare, and higher education 2003* (pp. 1495–502). Chesapeake, VA: AACE.

Bethesda Softworks. (2006). *Elder schools VI: Oblivion.* Retrieved June 7, 2007, from http://www.bethsoft.com/games/games_oblivion.html

Black, P., & Wiliam, D. (1998a). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice, 5*(1), 7–71.

Black, P., & Wiliam, D. (1998b). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan, 80*(2), 139–148.

Blunt, R. D. (2006). *A causal-comparative exploration of the relationship between game-based learning and academic achievement: Teaching management with video games.* Unpublished doctoral dissertation, Minneapolis, MN: Walden University.

Bull, S., & Pain, H. (1995). "Did I say what I think I said, and do you agree with me?": Inspecting and questioning the student model." In J. Greer (Ed.), *Proceedings of the 7th world conference on artificial intelligence and education* (pp. 501–508). Charlottesville, VA: AACE.

Cannon-Bowers, J. (2006, March). *The state of gaming and simulation.* Paper presented at the training 2006 conference and expo, Orlando, FL.

Carey, R. (2006). *Serious game engine shootout: A comparative analysis of technology for serious game development*. Retrieved March 23, 2007, from http://seriousgamessource.com/features/feature_022107_shootout_1.php

Clauser, B. E., Subhiyah, R., Nungester, R. J., Ripkey, D., Clyman, S. G., & McKinley, D. (1995). Scoring a performance-based assessment by modeling the judgments of experts. *Journal of Educational Measurement, 32,* 397–415.

Csikszentmihalyi, M. (1990). *Flow: The psychology of optical experience*. New York: Harper Perrennial.

de Freitas, S., & Oliver, M. (2006). How can exploratory learning with games and simulations within the curriculum be most effectively evaluated? *Computers and Education, 46*, 249–264

Duckworth, A. L., & Seligman, M. E. P. (2005). Self-discipline outdoes IQ in predicting academic performance of adolescents. *Psychological Science, 16*(12), 939–944.

Dweck, C. S. (1996). Implicit theories as organizers of goals and behavior. In P. M. Gollwitzer & J. A. Bargh (Eds.), *The psychology of action: Linking cognition and motivation to behavior* (pp. 69–90). New York: Guilford Press.

Falmagne, J.-C., Cosyn, E., Doignon, J.-P., & Thiery, N. (2003). The assessment of knowledge, in theory and in practice. In R. Missaoui & J. Schmidt (Eds.), *Lecture notes in computer science: Vol. 3874: 4th international conference on formal concept analysis* (pp. 61–79). New York: Springer-Verlag.

Finn, J. D., & Rock, D. A. (1997). Academic success among students at risk for school failure. *Journal of Applied Psychology*, *82*(2), 221–234.

Firaxis Games. (2008). *Civilization IV*. Retrieved February 2, 2008, from http://www.firaxis.com/

Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School engagement: Potential of the concept, state of the evidence. *Review of Educational Research, 74*(1), 59–109.

Fredricks, J. A., & Eccles, J. S. (2006). Is extra participation associated with beneficial outcomes? Concurrent and longitudinal relations. *Developmental Psychology, 42*(2), 698—713.

Gamelab. (2008). *Diner dash*. Retrieved February 5, 2008, from, http://www.gamelab.com/game/diner_dash.

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

Hartley, D., & Mitrovic, A. (2002). Supporting learning by opening the student model. In S. A. Cerri, G. Gouardères, & F. Paraguaçu (Eds.), *Lecture notes in computer science: Vol. 2363. Proceedings of the 6th international conference on intelligent tutoring systems* (pp. 453-462). New York: Springer-Verlag.

Hiebert, J., Carpenter, T. P., Fennema, E., Fuson, K., Human, P., Murray, H., et al. (1996). Problem-solving as a basis for reform in curriculum and instruction: The case of mathematics. *Educational Researcher, 25*(4), 12–21.

Hoska, D. M. (1993). Motivating learners through CBI feedback: Developing a positive learner perspective. In J. V. Dempsey & G. C. Sales (Eds.), *Interactive instruction and feedback* (pp. 105–132). Englewood Cliffs, NJ: Educational Technology.

Jonassen, D. H. (2000). Toward a design theory of problem-solving. *Educational Technology, Research and Development, 48*(4), 63–85.

Kay, J. (1998). *A scrutable user modelling shell for user-adapted interaction.* Unpublished doctoral dissertation, University of Sydney, Sydney, Australia.

Kirsch, I., Braun, H., Yamamoto, K., & Sum, A. (2007). America's perfect storm: *Three forces changing our nation's future* (ETS Policy Information Report). Princeton, NJ: Educational Testing Service.

Koller, D., & Pfeffer, A. (1997). Object-oriented Bayesian networks. In D. Geiger & P. P. Shenoy (Eds.), *Proceedings of the thirteenth annual conference on uncertainty in artificial intelligence* (pp. 302–313), Providence, RI: Morgan Kaufmann.

Konami Corporation. (2008). *Dance Dance Revolution.* Retrieved February, 19, 2008, from http://www.konami.com/Konami/ctl3810/cp20102/si1727630/cl1/dance_dance_revolution_universe_with_dance_pad.

Kruschke, J. K. (2001). Toward a unified model of attention in associative learning. *Journal of Mathematical Psychology, 45*, 812–863.

Kulik, J. A. (2002). *School mathematics and science program benefit from instructional technology* (InfoBrief No. NSF-03-301). Arlington, VA: The National Science Foundation, Science Resources Statistics.

Lemke, M., Sen, A., Pahlke, E., Partelow, L., Miller, D., Williams, T., et al. (2004). *International outcomes of learning in mathematics literacy and problem-solving: PISA 2003 results from the U.S. Perspective*. Retrieved May 4, 2007, from http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2005003

Lepper, M. R., & Malone, T. W. (1987). Intrinsic motivation and instructional effectiveness in computer-based education. In R. E. Snow & M. J. Farr (Eds.), *Aptitude, learning, and*

instruction: Vol. 3. Conative and affective process analyses (pp. 255-286). Hillsdale, NJ: Lawrence Erlbaum.

Lieberman, D.A. (2006). What can we learn from playing interactive games? In P. Vorderer & J. Bryant (Eds.), *Playing video games: Motives, responses, and consequences*. Mahwah, NJ: Lawrence Erlbaum.

Malone, T. W. (1981). Towards a theory of intrinsically motivating instruction. *Cognitive Science, 4*, 333–369.

Messick, S. (1994). The interplay of evidence and consequences in the validation of performance assessments. *Education Researcher, 32*(2), 13–23.

Mislevy, R. J. (1994). Evidence and inference in educational assessment. *Psychometrika, 59*, 439–483

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

Mislevy, R. J., Steinberg, L., S., Breyer, F. J., Almond, R. G., & Johnson, L. (2002). Making sense of data from complex assessments. *Applied Measurement in Education, 15*(4), 363–389.

Newell, A., & Simon, H. A. (1972). *Human problem-solving*. Englewood Cliffs, NJ: Prentice-Hall.

Nosofsky, R. M. (1986). Attention, similarity and the identification-categorization relationship. *Journal of Experimental Psychology: General*, *115*, 39–57.

Pausch, R., Gold, R., Skelly, T., & Thiel, D. (1994). What HCI designers can learn from video game designers. In C. Plaisant (Ed.), *Conference companion on human factors in computing systems* (pp. 177–178). New York: ACM Press.

Pearl, J. (1988). *Probabilistic reasoning in intelligent systems: Networks of plausible inference*. San Mateo, CA: Kaufmann.

Ranney, M. (1988). Changing naive conceptions of motion. *Dissertation Abstracts International, 49*, 1975B.

Reiber, L. (1996). Seriously considering play: Designing interactive learning environments based on the blending of microworlds, simulations, and games. *Education and Technology Research & Development, 44,* 42–58.

Reiser, B. J. (2004). Scaffolding complex learning: The mechanisms of structuring and problematizing student work. *Journal of the Learning Sciences, 13*(3), 273–304.

Rockstar Games. (2008). *Grand theft auto IV*. Retrieved February 18, 2008, from http://www.rockstargames.com/IV/

Shute, V. J. (2007a). *Focus on formative feedback* (ETS Research Rep. No. RR-07-11). Princeton, NJ: ETS.

Shute, V. J. (2007b). Tensions, trends, tools, and technologies: Time for an educational sea change. In C. A. Dwyer (Ed.), *The future of assessment: Shaping teaching and learning* (pp. 139-187). New York: Lawrence Erlbaum Associates, Taylor & Francis Group.

Shute, V. J., Hansen, E. G., & Almond, R. G. (2007). *An assessment for learning system called ACED: Designing for learning effectiveness and accessibility* (ETS Research Rep. No. RR-07-26). Princeton, NJ: ETS.

Simon, H. A. (1995). The information-processing theory of mind. *American Psychologist, 50,* 507–508.

Squire, K. D. (2006). From content to context: Videogames as designed experience. *Educational Researcher, 35*(8), 19–29.

Squire, K. D., & Jenkins, H. (2003). Harnessing the power of games in education. *Insight, 3*(5), 7–33.

Squire, K. D., Giovanetto, L., Devane, B., & Durga, S. (2005). From users to designers: Building a self-organizing game-based learning environment. *TechTrends: Linking Research & Practice to Improve Learning,49* (5), 34–43.

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

Sternberg, R. J. (1999). A propulsion model of types of creative contributions. *Review of General Psychology*, *3,* 83–100.

Sternberg, R. J. (2006). Creating a vision of creativity: The first 25 years. *Psychology of Aesthetics, Creativity, and the Arts, S*(1), 2–12.

Stiggins, R. J. (2002). Assessment crisis: The absence of assessment FOR learning. *Phi Delta Kappan Professional Journal*, *83*(10), 758–765.

Trabasso, T., & Bower, G. H. (1968). *Attention in learning.* New York: Wiley.

Vygotsky, L. S. (1987). *The collected works of L. S. Vygotsky.* New York: Plenum.

Weisberg, R. W. (1986). *Creativity: Genius and other myths.* New York: Freeman.

Zapata-Rivera, D., & Greer, J. E. (2004). Interacting with inspectable Bayesian models. *International Journal of Artificial Intelligence in Education, 14*, 127–163.

Zapata-Rivera, D., Vanwinkle, W., Shute, V. J., Underwood, J. S., & Bauer, M. (2007). English ABLE. In R. Luckin, K. Koedinger, & J. Greer (Eds.), *Artificial intelligence in education—Building technology rich learning contexts that work* (pp. 323–330). Amsterdam: IOS Press.

Notes

[1] This report was written while Valerie Shute was an employee of ETS.

[2] Other obstacles exist with regard to using serious games in education. These have been summarized and elaborated in the recent *Summit on Educational Games*, 2006 [http://www.fas.org/gamesummit], hosted by the American Federation of Scientists. Those issues, however, are beyond the scope of this report.

[3] Because all observables come from the same scenario (i.e., "task") there are a number of ways the context and activities can create dependencies among the observables. They are not known to be independent and they share a context, so we assume there is some degree of conditional dependence.