

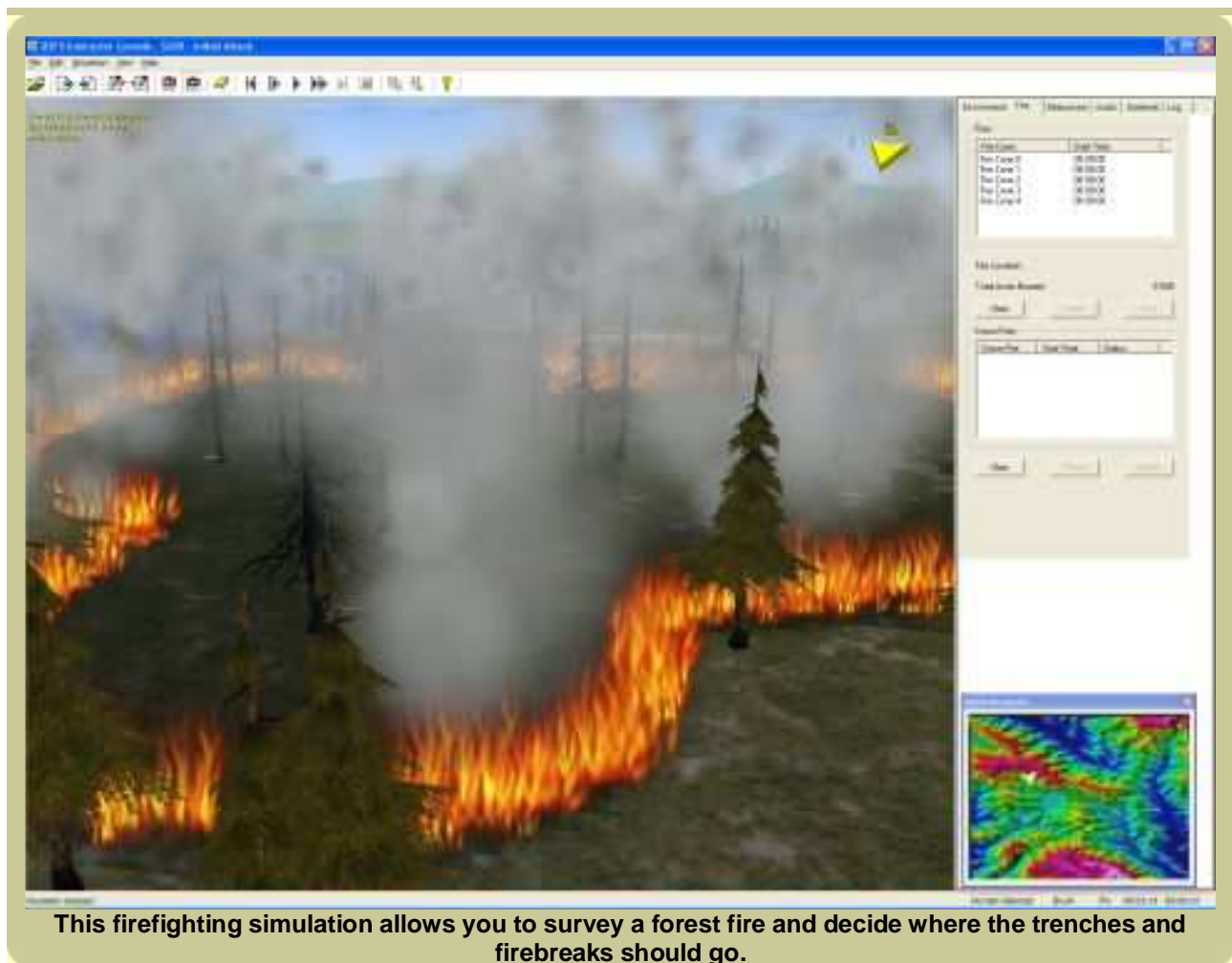
## Proof of Learning: Assessment in Serious Games

By Sande Chen and David Michael

Gamasutra

October 19, 2005

In Houston, Texas, a new hire steps onto a simulated offshore drilling platform and rehearses safety protocols. In Washington, D.C., a firefighter surveys a digital raging forest fire and chooses locations for trenches and firebreaks. A soldier in Iraq prepares for an upcoming mission using a detailed simulation of the urban battlefield. And a high school student in Portland, Oregon, manages the political campaign of Abe Lincoln as he tries to beat out Rudy Giuliani in the presidential elections of 2008.



**This firefighting simulation allows you to survey a forest fire and decide where the trenches and firebreaks should go.**

Games and game technology are poised to transform the way we educate and train students at all levels. Education and information, skill training, even political and religious beliefs can be communicated via video games. But these games and repurposed game technology, collectively called "serious games," have yet to be fully embraced by educators.

It's not enough to declare that "games teach" and leave it at that. Teachers aren't going to hand out a game to a bunch of students and simply trust that the students have learned the material.

Serious games, like every other tool of education, must be able to show that the necessary learning has occurred. Specifically, games that teach also need to be games that test. Fortunately, serious games can build on both the long history of traditional assessment methods and the interactive nature of video games to provide testing and proof of learning.

### A Quick Note

Assessment is a huge topic. In order to fit as much in as we can, we use the following simplifying terms:

- **Teacher** - The person in charge of the training, whether in a school, corporate training program, or military training facility. In most cases, "teacher" can be freely swapped for "trainer" or even "drill sergeant."
- **Student** - The person being taught. A "student" can be a "trainee," "recruit," or "middle manager hoping to further his career."

### A Tradition of Testing

Education is not merely the presentation of a subject to students. Assessment and testing is crucial in order to determine that the students have understood the material and can be expected to recall and use the material appropriately.

For millennia, teachers have used pop quizzes, recitals, competitions, verbal examinations, and a variety of other testing methods to see how well their students have learned the material. Teach and test, teach and test, the cycle repeats itself over and over throughout the process of education. For the teacher, the student, and any other interested parties, the purpose of this continual testing is to demonstrate proof of learning. Examples of why such proof is necessary are:

- Student advancement from level of education to another.
- National and international comparison of students.
- Demonstration that the student has completed a particular training program.

As they move into classrooms around the world, on computers and even video game consoles, serious games will continue this tradition of testing.

### Beyond MCQs

If you mention "computers" and "testing" in the same sentence, the first things most people think of are long sequences of multiple-choice questions (MCQs), and specially designed answer cards filled in with No. 2 pencils. Because computers can quickly and accurately grade MCQs, those types of questions have become the foundation of almost all modern testing. This makes MCQs the obvious first choice, and often the easiest choice, for assessment in serious games. MCQs are not always the *best* choice, though. While MCQs can accurately gauge memorization and retention of a set of facts, they are hardly the best way to gauge whether the student is following a process correctly. This is a notable shortcoming because some disciplines, such as advanced math, are more about the processes used to reach the answer and less about the answer itself. Multiple choice math tests can only provide a list of possible answers and have no easy mechanism for determining whether the student figured the answer out properly or merely guessed well. Another issue with MCQs is that outside of a few isolated examples such as *Trivial Pursuit* and *Who Wants to be a Millionaire*, they have little or nothing in common with video games. While a review of any collection of edutainment software reveals that MCQs can be easily tacked on to a video game, doing so does not take advantage of any of the features that make

serious games compelling: engagement of the player, self-motivated progress through the material, and fun.

Serious games represent an opportunity to move past this simplistic, narrowly focused type of testing. In fact, they can do so by combining other forms of traditional assessment with methods modern video games now use on a regular basis. Together, it's possible to create more complex and complete types of assessment than have ever been available before.

### **Assessment in Entertainment Games**

The major difference between regular video games and serious games, of course, is that serious games have education as a primary goal while video games focus on entertainment. Despite this fundamental difference, however, even video games designed for nothing more serious than hour upon hour of mindless entertainment have a learning objective, at least at the beginning: teach the player how to play the game. These games also employ pass/fail mechanisms no less rigorous than many college entrance exams. This may come as a surprise to many game developers. James Paul Gee, though, the author of *What Video Games Have to Teach Us About Learning and Literacy*, argues in his book that the best video game designs demonstrate sound educational technique. Specifically, many games designers (whether intentionally or otherwise) build complex learning and progression into their games. In the game development industry, we call these "tutorials." Tutorials present the player with the basics of how to control and interact with the game and then test the player on this information with a series of levels or missions. Tutorial missions often introduce only a few new game features or play elements at a time to avoid overwhelming the player. By the time the player has completed these first few missions, he or she has "learned" the essentials of the game and can be bombarded with ever greater in-game challenges. This process even continues past the tutorial, as later levels and missions in the game become more and more difficult. Another form of assessment in entertainment games is scoring. Many games even offer comparisons between players with high score lists. These high scores can be a source of bragging rights for the player, but, more importantly, the scoring system teaches the player what is important within the game. A positive score indicates a good choice, a negative score a bad choice, and no score at all indicates that the attached action is probably unimportant. Though few classrooms stress the level of competition seen in most video games, the similarity to the posted test grades is unmistakable. In the same way, the education strategy of "teaching to the test" clearly identifies to the student what is important to learn and what can be ignored just like in-game scores do in entertainment games. Jim Brazell, consulting analyst at the Digital Media Collaboratory (DMC) in the IC<sup>2</sup> Institute at the University of Texas at Austin, talks about another type of assessment method that stems from video games. "I believe that the most serious game is the game of game construction," says Brazell, who advocates the use of game development itself as a learning tool. His reasoning is that the only way a designer can make an effective game that simulates a particular phenomenon or teaches particular information is if the designer already understands the phenomenon or information. Further, the creation of such a game has the potential to lead to new knowledge and new ways to do things through emergent behavior. As the methods and tools of game development become more accessible, perhaps this new kind of "using games in education" could take its place alongside other serious games.

So, rather than only translating traditional testing methods like MCQs into serious games, designers of serious games can also build on the methods that have worked in mainstream video games. That isn't to say that game designers already know everything there is to know about testing and other pedagogical methods. Nor are we saying that traditional testing methods have no place in a game environment. Instead, both game designers and educational professionals need to work together in developing serious games as a new teaching tool.

### **Assessment Challenges**

Both the medium of serious games itself and its newness create certain challenges that can make assessment difficult:

- With less emphasis on rote memorization of facts, the assessment obtained from traditional methods may not accurately reflect the learning gained from serious games.
- Open-ended simulations can support a wide range of possible solutions. Which one is more correct?
- When teaching abstract skills such as teamwork and leadership, how do you measure learning and/or improvements?
- What is "cheating" in the context of serious games?

Despite their success using educational methods such as tutorials, game designers and developers must recognize their own limits when it comes to serious games. "Figuring out if somebody learned something is a very difficult task," says Jonathan Ferguson, Interaction Designer at the EduMetrics Institute in Provo, Utah. So difficult that there is an entire field of study devoted to it - psychometrics. Psychometrics is the field of study concerned with measuring mental capabilities. It has evolved over the past two centuries and has been used to measure such disparate and seemingly immeasurable capacities such as personality, individual attitudes and beliefs, academic achievement, and quality of life.

According to Ferguson, too many people assume that any game will teach and be helpful regardless of the software's actual capability. The core questions to ponder, he says, are:

- How do you show that the students are learning what you claim they are learning?
- How do you know that what you are measuring is what you think you are measuring?

Cheating is another challenge faced by serious games. Cheat codes have a long, colorful history in video games, but they could compromise the learning experience in a serious games. Also, what kinds of interaction by students both internal and external to the game should be supported or discouraged? Students might be expected to work together and to provide insight into how well they and their classmates understood the material.

Finally, the "game" part of "serious games" presents a challenge for designers. Whether "fun" is a necessary or even desirable element of serious games has already become one of the perennial debates within the serious games community. A large part of the appeal of serious games is that they provide a familiar environment for the latest generation of students. Games are something these students relate to and understand. However, games that act too much like a classroom, with pop quizzes interrupting the player's experience can disrupt their appeal.

### Meeting the Challenges

Because serious games have such challenges, serious game developers have turned to more sophisticated assessment methods. Of note, there are three main types of assessment used in serious games:

- Completion Assessment - Did the player complete the lesson or pass the test?
- In-Process Assessment - How did the player choose his or her actions? Did he or she change their mind? If so, at what point? And so on.
- Teacher Evaluation - Based on observations of the student, does the teacher think the student now knows/understands the material?

The simplest form of assessment is completion assessment: Did the student complete the serious game? In traditional teaching, this is equivalent to asking, "Did the student get the right answer?"

Since many serious games are simulations, this simple criterion could be the first indicator that the student sufficiently understands the subject taught. Note that this is *not* the same as asking, "Did the student attend every lecture?" Because serious games require interaction by the students with the material, completing the game could signify more learning progress and comprehension than passively attending lectures in a typical classroom setting. Unfortunately, the mere criterion of successfully completing the game falls short on a number of fronts. Besides the possibility of students cheating or exploiting holes in the system (a time-honored tradition in video games, but considered in a less positive light in classroom settings), it's important to know whether the student learned the material in the game, or just learned the game and how to beat it. As the pedagogy of serious games evolves, assessment in serious games will come closer to this simple ideal. In the meantime, though, more is needed. In-process assessment is analogous to teacher observations of the student as the student performs the task or takes the test. In advanced math and science courses, for example, students are required to write out each step of the process they followed. Erasures are often disallowed in favor of drawing a line through incorrect steps and conclusions so that errors in the process can be more easily seen by the teachers. This is because the errors and corrections can be valuable indicators, sometimes more so than just giving the correct answer.

Serious games, or more specifically serious video games, offer logging and tracking potential that has seldom been available or even possible in traditional classrooms. Video games have long had logging features that allow players to replay their performance in the games. Modern games have even begun to learn from the player's actions within the game, adjusting storylines, strategies, monster strength, and other variables to adjust to what the player has done and is doing. Serious games can take advantage of these features. For instance, Offshore Safety Initiative, located in Houston, Texas, performs detailed logging in its safety simulation software, tracking such data as:

- Time required to complete the lesson;
- Number of mistakes made;
- Number of self-corrections made; and more.

Full in-process assessment of players, in which the serious game itself determines how well the player is learning, is still some time away. In the meantime, though, the information logged can be used to assist teachers with their assessments of the students. Teacher evaluation is a combination of both completion assessment and in-process assessment. Despite the predictions (or fears) of some, serious games aren't going to be replacing teachers anytime soon, and probably never. To that end, serious games should include tools to assist teachers in their evaluation of students. Such tools can include homework and assignment controls, grade tracking, reporting, and more. Like the process notes mentioned above, with detailed logging, properly presented, teachers can evaluate their students' mastery of the material. The more data that is available, the less subjective that evaluation needs to be.

Teacher evaluation can also include observation of the student in action. Multiplayer video games often include "observer modes" that could be used for this, both by the teacher and other students. Other possibilities exist too. In the firefighting simulations developed by Dynamic Animation Systems of Fairfax, Virginia, the National Fire Academy and the United States Department of Agriculture Forest Service only uses one main assessment tool: the instructor. Students submit their responses to the instructor, who feeds those responses into the simulation. Then, the students and the instructor watch how the situation progresses. In this case, the instructor is not looking for the one correct answer. Instead, the goal is to teach students how to quickly choose a good way to improve the situation and bring the fire under control.



**PIXELearning's Learning Beans.**

Good pedagogy and instructional value is paramount in serious games. While games with educational content have existed for a long time, too many have relied on simplistic or unproven metrics. "What's been missing is a pedagogy engine and an assessment engine," Brazell says. Both the DMC and the EduMetrics Institute advocate addressing assessment issues as an initial part of serious game design.

Some companies like PIXELearning, of Coventry, UK, are already devising such pedagogy and assessment engines into their products. PIXELearning utilizes its own proprietary engine, called Learning Beans, to integrate assessment methodologies into its game-based business simulations. Managing Director Kevin Corti says, "Entertainment game developers frequently encounter frustration when they are required to do this but it is a crucial aspect of games for learning purposes. A simple post-game multiple-choice questionnaire will not suffice."

"Assessment starts pre-game," Corti continues, "runs all the way through [the game] and continues after the game." An important feature of this built-in assessment is the way the game adapts to the player's behavior and gives the player the appropriate feedback. Players come to understand the connection between their in-game actions and the outcomes. Meanwhile, the teacher receives detailed assessment results to properly gauge the student's progress. In addition, the assessment engine leads the student through a series of qualitative questions such as "You just choose to do X. What was your basis for this decision? Why did you not choose Y?"

Thus, the teacher has a lot of information available to judge how well the student really does understand the material being taught.

All of this creates what Corti calls "authentic learning." Since the learning in the game is personally meaningful and relevant, the serious game provides the student with the opportunity to practice and apply skills needed in the real world.

## **Conclusion**

The future of serious games as an educational tool depends on their improved support for completion assessment, in-process assessment, and teacher evaluation. Designers and developers will need to reach beyond simple multiple-choice questions and incorporate the best of video game tutorials with sound educational and psychometric techniques.

Moreover, if game developers can show skeptical teachers that not only do serious games help teach the material better, but that the games can be easily integrated into existing lesson plans, those teachers are bound to lose their objections.

"[Serious games] will not grow as an industry unless the learning experience is definable, quantifiable and measurable," Corti says. "Assessment is the future of serious games."

*Copyright © 2003 CMP Media Inc.*