The Right Kind of GATE:
Computer games and the future of assessment

David Williamson Shaffer
and
James Paul Gee[1]

Abstract

In the past we have referred to games as good "learning engines." Here we argue that games are good learning engines because they are first good *assessment engines*, and that games require the kind of thinking that we need in the 21st Century. They use actual learning as the basis for assessment. They test not only current knowledge and skills, but also preparation for future learning. They measure 21st Century skills like collaboration, innovation, production, and design by tracking many different kinds of information about a student, over time. Thus we suggest that the road to better schools starts by making the tests in school more like the games that students are already playing out of school.

Introduction

We have argued in previous work that computer games are good for learning (Gee, 2003, 2007; Shaffer, 2007). Computer games can create virtual worlds where players solve simulations of real-world problems and in the process learn real-world skills, knowledge, and values. Parents and teachers can use commercial games to stimulate discussions of important social, intellectual, and academic subjects. But most of all, places where people learn—schools, corporate training centers, summer camps, or living rooms—should become more game-like.

Good games focus on problem solving. They provide a good mix of practice and guidance. They use language and introduce complex concepts when they are needed—and thus when those tools can best be used and understood. They provide extensive time on task, but

[1] David Williamson Shaffer is Professor of Learning Science at the University of Wisconsin-Madison. James Paul Gee is the Mary Lou Fulton Presidential Professor of Learning Studies at Arizona State University. They are equal coauthors with names listed in reverse alphabetical order by middle name.

players are motivated to spend that time because games provide a sequence of challenges that gradually increase in difficulty, so players are constantly working at the cutting (and most exciting) edge of their abilities.[2] These are all things that any good learning environment should have, and commercial games have to use them for commercial reasons: if players couldn't learn to play commercial games—and eventually master them—no one would buy them (Gee, 2003, 2007).

These good principles for learning are even more important in the 21st Century, where students need to learn to work with others and with digital tools to produce and not just to consume knowledge. They need 21st century skills like innovation, critical thinking, and systems thinking. Digital technologies—including games—are letting young people produce products and knowledge and to participate in learning communities as never before. Through the Internet young people are becoming amateurs with professional level skills in areas like digital storytelling, fan fiction, graphic arts, machninma, game design, digital photography, robotics, and so on (Leadbeater & Miller, 2004).

Computers have changed learning. More and more they are being used by privileged families to accelerate their children's skills in literacy, history, civics, and math, science, and technology. But the same changes haven't happened in schools.

The problem with testing, now

The reason we have not yet entered the 21st century in our classrooms is simple: assessment.

[2] Experiences of this kind, that provide experiences that are neither too hard (and thus frustrating) or too easy (and thus boring) induce what psychologists and game designers refer to as a "flow state" of high excitement and focus, which, of course, explains why good games (like good books or good movies) can be so compelling. (Csikszentmihalyi, 1996; Gee, 2007).

Our standardized tests, coupled with our accountability policies, force teachers to teach to out-of-date tests. The curriculum is based on reading from text books and listening to teachers talking on drill and practice, rather than on doing, exploring, and developing deep understanding of complex topics and issues. Classes focus on facts and formulas that learners need to pass standardized tests, even though when people learn that way they have a very hard time applying what they "know" to solve real problems (Gardner, 1991; Gee, 2004).

Part of the problem is that in our schools right now, learning and assessment are quite separate things. A teacher teaches for weeks and months, but the judgment of how well student and teacher have done is made on one day, on a test that knows nothing about the development of the learner. It is an assessment that captures one small snapshot in time of what a student can do. Based on that small slice of time, students, teachers, schools, and neighborhoods are graded. Decisions are made that affect funding, careers, and futures.

If we are going to succeed in introducing the new ways of learning that computers make possible, first we have to radically transform assessment. Only when teachers, parents, educators, and policy makers look at testing and assessment in new ways will they look at learning in new ways. We've been designing *games for learning* when we should have been designing *games for testing*.

What's in a GATE?

The idea of designing games for testing is less radical—far less radical, in fact—than it sounds. To see why, let us examine some of the features that any Good Assessment for Twentyfirstcentury Education (GATE) would have to possess.[3] We argue that three fundamental

[3] We're describing this briefly here, but this summary of principles for 21st Century Assessment is based on a series of discussions funded by the Macarthur Foundation in their 21st Century Learning and Assessment Project (PI:

properties of assessment need to change in the 21st Century: ***what*** is assessed, ***how*** the assessment takes place, and the ***purpose*** of assessment in the first place.

What we assess

The single biggest problem with standardized tests today is that they are built around facts and information in and for themselves, rather than around problem solving. When students master facts and information in isolation they often can write them down on a test, but they cannot reliably use them in the real world (Chi, Feltovich, & Glaser, 1981; Gardner, 1991; Gee, 2007). Any GATE would have to be built around central problems in an academic domain (like algebra, civics, or biology) or a real-world profession (like medicine, engineering, or law)—any place where central concepts in 21st century thinking are put to work in solving complex real-world problems.

In assessing students' problem solving skills, a GATE would also have to assess 21st Century skills. There are now lists of such skills (Partnership for 21st Century Skills, 2004), often including things like innovation, collaboration, civic engagement, critical thinking, system thinking, technical skills, ability to produce with digital media, and so on, but nearly all of those lists includes, at a minimum, collaboration, innovation, production, and design.

Moreover, a GATE would not just tell us what students know and can do *now*. Knowledge and skills change and transform themselves quickly in the modern world, so a GATE would have to provide information about how instruction has helped students be ready to learn more *later on*: that is, how well prepared students are to learn more in the same or a related area in the future. A GATE needs to include resources that let students learn during the test, so we can

---

James Paul Gee) over several years that have brought together some of the best current research on the problem of assessment in the new century.

assess what Bransford and Schwartz (Bransford & Schwartz, 1999) and others have described as *preparation for future learning*. Recent work has shown that choices students make on problem solving can tell us a great deal about their ability to learn new material later on, and this makes sense: certain choices in a domain show that someone understands problem solving at a basic level well enough to succeed at higher levels (Schwartz & Arena, 2009). So a GATE would assess whether learners can make good choices and understand the consequences of their choices.

In other words, a GATE would *test whether students make the kind of choices that experts do in a domain as they work with other people to solve complex problems of innovation, production, and design*.

How we assess

In order to test whether students are making good decisions while problem solving, a GATE will have to track multiple variables. Learning in any domain is a complex phenomenon. For example, successful reading for content (say, in social studies) requires at a minimum skills in decoding text, domain vocabulary, and interpretive skills. These different abilities have to work together in sophisticated ways. Learners with a problem in reading do not all have the same problem, and often the problem is an interaction between two or more different issues. So a GATE has to be able to track how a student's decisions and actions are related to his or her overall development—and thus needs to clearly explain its theory of how the domain being learned works, and how learning and instruction works best.

Since decisions and actions unfold over time, a GATE would also have to be developmental: it would provide information relevant to students' learning and growth at different points. However, it is difficult to measure how a student's decisions and actions in are

related to his or her overall development based on "one-off" measurement events, like our current tests. Instead, we need measurements that show what students can do over time and tell us about the course of their development and how it can be improved. Any GATE should tell us about the different paths that students can take to mastering a domain, and also tell us where any student is on one of those paths.

In order to do this, a GATE needs to integrate assessment with learning. Digital media make it possible to collect large volumes of information, and to organize that information in real time. In a world where we can collect copious information and visualize it in different ways, the distinction between formative and summative assessment begins to disappear. We will be using much the same information to help learners and to judge the success of programs, processes, and practices for learning. To accomplish this, a GATE should be part of the learning space. That is, students shouldn't have to "step outside" for separate assessment events. When diagnostic learning tasks continually assess the development of learners, we get a portrait of problem solving decisions in real time. We can provide feedback to customize learning, and we can probe the strengths and weaknesses of students' thinking.

Why we assess

All of which points to what is perhaps the most significant point: we need to rethink why we assess students in the first place. In the world that the No Child Left Behind testing regime has created, assessment is largely about punishing teachers. But continuous assessments of multiple data sources about complex problem solving and 21st century skills would be more about giving teachers, parents, students, and other stakeholder's useful and actionable information. Any GATE should yield information—and organize that information—in ways that

help teachers, parents, students, administrators, or policy makers to take appropriate actions to improve instruction and learning.

Most important in that regard, any GATE has to deal with one of the deepest problems with our current assessments: they treat all learners as if they have had the same opportunity to learn—the same experiences relevant to learning—and judge them all alike. They are oriented to the "status" of each learner in the sense that we assess whether each learner has reached a level of performance labeled something like "proficient". We do not measure growth, only whether or not a student has mastered some set of skills, no matter how small or long a trip this was for the child. And we hold teachers accountable in such terms, regardless of the progress their students have made, making judgments only about how many of their students have reached a certain level.

This is clearly problematic. On a reading test, for example, some children have experienced the content of the passages on the test in other books they have read, media they have watched, or on trips to the zoo or other educational locations. So they can answer some of the questions on the test—whether a tiger is larger than an elephant, for example—based on background knowledge, while other children cannot. Some children have heard parents use school-based vocabulary many times—Latinate words like "process" and "establish" for instance—while others have not. A GATE would have to take into account the fact that these children have not had the same experiences and provide missing resources before or during the assessment. After all, what we care about is less whether a child has passed point X in their development than what they are capable of doing next. We care about what they are capable of doing in a world where they have to solve complex problems with collaborators and technologies, not whether they remember the relative sizes of two species of mammals.

Imagine, if you will, two students. One comes from a family that has provided many learning experiences outside of school. This student starts the school year working at the 11th grade level and finishes at the 12th grade level. A second student comes from a family that does not (and perhaps can not) provide many enrichment experiences. She starts the school year working at the 7th grade level and finishes at the 11th grade level. The first child is a year "ahead" of the second in terms of her "level" of performance. However, with the appropriate resources, the second child made up 4 grade levels in a single year. The question is: which of these students would you want to hire? The choice seems obvious if what we care about is how well students take advantage of opportunities and use resources, rather than simply measuring what resources were already made available to them

What we need

In other words, we need to break the mold in our schools and introduce a new paradigm for teaching and learning, focused on real world problem solving and 21st century skills like innovation, critical thinking, and systems thinking. To break out of the old paradigm of teaching to standardized tests of basic facts and skills we need new assessments that:

1. Change what we test by focusing on complex problem solving, 21st Century skills like collaboration, innovation, production, and design, and evaluating students' preparation for future learning;

2. Change how the assessment takes place by tracking many different kinds of information about a student, over time, and integrate assessment with learning; and

3. Change the purpose of assessment from sorting students and punishing "underperforming" teachers and schools to providing students, administrators, parents and teachers with feedback they can use to make decisions that support good learning, and that account for the fact that different students have had different opportunities to learn inside school and out.

That is surely a very different view of assessment than we currently have. However, to get an assessment system for the 21st Century, we *don't* have to reinvent the wheel.

Games as assessments

We don't have to reinvent the wheel because games are already an exemplary platform for assessment. They have much to teach us about 21st century assessment, and they can lead us to design a transformative assessment system that has the potential to usher in a new paradigm for teaching and learning.

Too often today designers of learning in and out of school first think about how the learning ought to work—that is, what the curriculum and pedagogy will be—and then worry about how to assess the learning. To be clear, we include ourselves (or at least our past selves!) in that group.

But games take just the opposite approach. They worry first about how to test and challenge a player in an effective way. The learning design then follows from the assessment.

Consider, for example, some of the key properties of games, and how they create the conditions of a GATE:

1. ***Games are built around problem solving***, and on the choices and actions players take to solve problems. So players have to use facts, information, and other representations (like graphs, diagrams, maps, and models) in the context of making consequential decisions.

2. ***Games inherently require and assess a set of 21st century skills***. Modern computer games require players to solve problems collaboratively with other people. In a game like *World of WarCraft*, a team of five players constitutes what modern workplaces call a "cross-functional" team, composed of people with deep and special expertise in different areas who can understand and integrate with each of the other team member's specialties. Games place a premium on a player's ability to create, innovate, and produce. Players are pushed to find their own solutions to challenging *boss levels*,[4] and often share these solutions with other players on fan forums. Many games today come with the software by which the game was made, so players can modify (*mod*) the game, designing their own levels and scenarios, becoming designers as well as players. Finally, players have to figure out and model the rule system of a game in order to use it. In a game like *Civilization*, the player must map out a complex set of relationships among variables within a civilization and across civilizations. In turn, the player must use this model based reasoning and systems thinking to his or her advantage in the game.

[4] "Boss battles" at the end of a level in a game are often used to assess whether the player has mastered the skills of the level just finished, and whether he or she is prepared for learning the more demanding challenge of the next level.

3. ***Games assess whether a player is ready for future challenges.*** Boss levels do not just assess what a player knows and can do—that is, they don't just measure a player's mastery of the previous level. They also are designed to see if players are prepared for the greater challenges ahead. Good boss levels test whether the player is ready and prepared to learn, and learn well, on the next level.

4. ***Games collect information about players on many dimensions.*** They track multiple variables and relate them back to players in clear and actionable ways. In a game like *Civilization*, the game keeps track of how players deal with problems across time: issues in the economy, industry, technology, military, environment, religion, diplomacy, and governance of a civilization. The game tracks how the player's decisions and actions in all these spheres are related to his or her overall development and success.

5. ***Games track information across time.*** Games are designed in terms of levels. Each level demands players have mastered the skills on an earlier level, and demands that they learn new skills on the new level. That is, levels are deliberately designed to model the development of the player as the game proceeds.

6. ***Games integrate learning and assessment.*** In a game, learning and assessment are, in many ways, inseparable, and it is often hard to tell where one ends and the other begins. Every action a player takes and every choice a player makes has consequences. The player is given feedback about what worked and what did not. The player's actions and choices across a game as a whole are tracked and the player is

informed in various ways as to how he or she is progressing. Results are always apparent. But such information does not only help, mentor, and develop learners. All the information the game does or could track and give back to the player as helpful feedback is also just the sort of information that could give us a deeply nuanced evaluation of the player and his or her learning.

7. ***Games provide information that players can use to get better at the game.*** The information a game gives a player, level by level—or when the player gets graphs and diagrams in a real-time-strategy game like *Rise of Nations*—is not used primarily to sort the player against other players[5]; rather the information is meant to be acted on, and so is presented in ways that make it actionable. It is the sort of information that allows players—and would allow people who wanted to mentor them—to make decisions about what to do next to get better, have more success, and develop. When a player finishes a level of an action game like *Darksiders*, the player knows whether he or she should repeat the level to get better, practice certain skills with more care in the next level, or try a new approach to the game. When a player gets feedback from a real-time-strategy game like *Age of Empires*, he or she knows what went well and what went poorly in the last session of play, and has ideas about specific things to try next.

8. ***Games have to be equitable.*** To market a game successfully, game designers need to make games so that poor people and rich people, minorities and non-minorities, and

[5] Some websites do use that information to create player rankings; but games can be played without reference to those external resources.

players with little experience and players with lots of experience can play them. After all, the game industry is a business, and it cannot afford to cater only to the best players (although it cannot afford to lose them either). Games have traditionally not done a very good job at inviting girls and women in, but this is being remedied. The majority of players of the best selling game of all time, *The Sims*, are girls and women. The game industry is well aware that how much experience a player has already had with games or games of a certain type will predict a good deal about how well that player plays a new game of the same type. So games take this into consideration and offer different resources and different rewards for different sorts of players. Games resource players with less experience differently than they do more experienced players. They offer tutorials, advice and hints, lower difficulty levels, the ability to replay levels and so on. Sometimes they adjust the difficulty of the game on the fly, making it easier or harder based on how well the player is doing moment by moment.

Deep down, in other words, games do not just "have good assessments built into them." Games are nothing *but* good assessment. The player is always being tested, given feedback, and challenged to get better. Good game design starts with the question: How will the player be tested? The design follows from that: How can we help the player pass the test? How will we know the player has passed the test? If the player can pass one test, what's the next test he or she should be able to pass on the way to mastery? How do we know the test is fair?[6] These questions

[6] *Fair* is a term of art for players, meaning that the test/challenge has not been artificially constructed to help the player lose rather than win.

lead games to incorporate good learning designs precisely because they have first incorporated good assessment designs.

Good games achieve good learning because they do not set out, first and foremost, to teach. They set out to assess, and their approach to assessment leads to good teaching and learning.

Three possible solutions

The fact that games are based, fundamentally, on the kind of assessments we need to promote 21st Century learning has three immediate—and very dramatic—consequences, for games and for learning.

Adopt game principles

The first, and perhaps most evident, is that designers of 21st Century assessments can learn a lot from games. Too much of the work currently being done on digital tools for assessment takes the same old standardized tests as a model: finding ways to make them cheaper, to use question banks more effectively, to make them more time-efficient by skipping questions a student is likely to get right, to make it harder to cheat, and so on.

Games offer a radically different example for assessment designers to build from: a kind of working model of what a 21st Century test can look like. Designers can look for guidance to see how games offer hints or provide just-in-time resources to struggling students. They can look at how feedback is presented in games to help students and teachers use assessment as a constructive tool—that is, how to present feedback that can actually help students learn from the test. They can look at how games capture and use multiple sources of data over long periods of time to get information about a student's work. And perhaps most of all, they can look at games

to see an example of how to present students with complex problems that require collaboration, systems thinking, and creativity to solve.

In other words, games can provide educators with an example of assessments that are *standardized*—in the sense that every player who opens a box or logs onto a game's website gets the *same game*—but that is about more than basic facts and basic skills.

But there is a more transformative consequence of the fact that games are good examples of 21st Century assessment: Namely, that we can—and should—use games instead of traditional tests to assess what students know.

Using existing games as assessments

The simplest way to use games as assessments is to have students play existing games and use their performance in the games (or perhaps their ability to explain what they did in the game and why) as a test of how well they understand a domain of knowledge.

For example, consider the game *Civilization*.[7] *Civilization* is a strategy game in which players build an empire starting from a Stone Age settlement. They make strategic decisions to invest in technological development or agriculture, and to use a combination of trade, diplomacy, religious conversion, and warfare with their neighbors. The game is based on historically accurate information about advances in technology, religion, warfare, and the arts, and takes a materialist-determinist approach to history, like the one presented by Jared Diamond in his Pulitzer Prize-winning book *Guns, Germs, and Steel* (Diamond, 2005). To do well in the game, players have to understand how geographical location, ease of trade, and access to raw materials create conditions for the successful growth of a civilization—and they have to be able to

[7] Our description of the educational potential of *Civilization* is based, in part, on the work of Squire, who has studied the use of the game in classrooms and after school clubs. (Squire, 2004)

demonstrate that understanding in action. The game provides a wealth of information about how well a player has done in building his or her civilization and about the strengths and weaknesses of the strategies the players has chosen.

The game is realistic enough that advanced players develop better strategies by reading up on world history. So one can imagine a teacher asking students to play *Civilization* not to learn history (although that would surely be a good outcome), but to test how well they understand history. This might involve not just asking students to play the game and produce their scores, but also to provide an annotated explanation of what they did during the game and why—which is what advanced players in the game already do.[8] The fact that games like *Civilization* can be modified by players means that teachers or curriculum developers could produce scenarios customized to a particular content area, and also that students could be asked to design scenarios as part of their assessment.

There are, of course, two obvious drawbacks to such an approach. The first (and less significant) is that some work would need to be done to adapt a commercial game to serve as an assessment instrument, including a significant investment in determining the reliability of the measures used in the game, methods for ensuring that the test scenarios are not distributed in advance, and other criteria for assessments that would have to be met. In other words, a commercial game could be the core of an assessment tool, but the tool would have to be built to use the game in that way. But that is only to say that more research and development would be needed before a commercial game could be part of an assessment system.

[8] This approach of creating annotated walk-throughs of game play is used on the fan website Apolyton University, where advanced players share scenarios organized into courses on strategy games (including *Civilization*) and history more generally.

The more significant issue is that, because the demands of the commercial marketplace differ in some ways from our assessment needs, there are not necessarily commercial games out there that meet every testing need.

Fortunately, that presents a possible solution: Develop a game system, game engine, and approach to educational gaming that can serve as a framework for creating assessment games.

Designing games for assessment

In what follows, we describe one system of developing games that can be used as 21st Century assessments, and we do so by describing one particular game. It is a game we have written about elsewhere (Shaffer, 2007; Shaffer & Gee, 2005), and we present it as an example of the *kind of* assessment system we need to create. But we want to emphasize that there are other examples that we could have chosen as well.

We argued above that 21st Century assessments have to be built around central problems in an academic domain or a real-world profession. The profession of urban planning is a good example of what we mean.

Urban planning is a domain of practice traditionally taught at the postsecondary level, but it is the kind of innovative and creative thinking that students need in the 21st Century. Work in urban planning calls for some of the same skills and knowledge that are in the National Science Education Standards: (National Research Council, 1995) things like understanding systems, order, and organization; evolution and equilibrium; and form and function in natural systems. Land use models that urban planners work with combine geographic features and other information into interactive visual models of complex systems. They show how land use decisions affect key environmental, economic, and social indicators: pollution, tax revenue, and acreage of wildlife habitat and so on. These models show the interaction between ecological and

social systems in a local community that let planners explore, propose, and justify solutions to complex ecological and economic issues.

In the game *Urban Science,* players work as urban planners to create proposals for the development of the north side of Madison, Wisconsin, an area adjacent to a large wetland known as Cherokee Marsh. This development project raises a number of economic and ecological issues around wetland ecology and conservation. Not surprisingly, while working on plans for development near the Cherokee Marsh, players of *Urban Science* have to investigate, analyze, understand, and communicate about scientific issues: local species, their life cycle, and their habitat; the role of wetlands in the local ecological system; and specific pollutants, their sources, and their impacts.

To be successful in the game, players have to use and develop skills and knowledge from state science and environmental science standards. They have to learn and use concepts in ecology, such as systems thinking and sustainability. They have to value civic thinking, and use technology and scientific understanding to develop innovative solutions to real problems facing the city. They have to solve complex problems using the mathematics, communications, and science skills of urban planners (Bagley & Shaffer, 2009; Beckett & Shaffer, 2005; Shaffer, 2007).

Previous studies of *Urban Science* have focused on whether players developed these kinds of 21st Century skills from playing the game. In one study, for example, middle school students used knowledge, skills, and values from ecology and urban planning more after playing the game (Bagley & Shaffer, 2009). Other studies have looked at whether games like *Urban Science* develop skills, interests, and motivation that can help players do better in science and other school subjects. For example, because players communicate with adult mentors in games

like *Urban Science*, some become more comfortable talking with their teachers and talking in class.(Shaffer, 2007)

What would it take, then, to use a game like *Urban Science* to assess whether students are learning anything useful in their classes. To do that, we would need to be able to measure the kind of 21st Century thinking that is happening in the game, and show that the game can collect and report information that will help students, teachers, parents, and others decide whether teaching has been effective and where individual students still need help.

Research in the last two decades that suggests that learning to solve complex problems comes from being part of a community of practice: a group of people who share similar ways of solving problems (Hutchins, 1995; Lave & Wenger, 1991; Shaffer, 2006). A community of practice shares a common body of knowledge and set of skills, but also a system of values that determines when and how those skills and that knowledge should be employed, and a set of processes through which such decisions are made. And, of course, such a community also has a shared identity. In previous work we have described this collection of skills, knowledge, identity, values and epistemology of a community as its *epistemic frame* (Shaffer, 2007).

Thus, we can look at what urban planners say and do in their work, find the relevant skills, knowledge, identity, values and epistemology, and create a model of the way planners think about problems: a model that describes what it means "to solve problems the way a planner does." And we can do the same thing in the game Urban Science:

1. Look at what players say and do in their work in the game;

2. Find the relevant skills, knowledge, identity, values and epistemology from urban planning;

3. Create a model of the way the players think about problems in the game; and

4. Compare that to how real planners think.

Consider, for example, a player in *Urban Science*. Let's call her Sarah.

We could determine the epistemic frame of urban planning that Sarah has at any point in the game. We could also determine the frame of the group or groups that Sarah has worked with. In fact, we could determine the frame of all of the parts of the game (including other players) that Sarah saw while playing. We could, from studies of real urban planners (or by having real planners play the game), construct a *reference frame* of how real planners solve problems.

A mathematical modeling technique, *epistemic network analysis* (ENA), can measure the similarities and differences between these ways of thinking—that is, between these epistemic frames (Shaffer, et al., 2009). So we might ask, for example: How close did Sarah's frame get to the reference frame of real planners? That is, we could measure how much Sarah learned to think like a planner. We could see the path over time of Sarah's frame development in the game, and compare it to experts, or other players. We could ask whether Sarah's frame was more likely to become like a real planner's if the frame of the players she worked with had frames that became more like a planner's. That is, we could quantify one of the most elusive concepts in education, *opportunity to learn* (Darling-Hammond, 2006), by looking at how players do in the game compared to the context in which they were being tested.

In other words, we could use *Urban Science* to show conclusively how well Sarah was thinking like a planner—and give feedback about what areas she still needed to work on. In this sense, we call *Urban Science* an *epistemic game*: that is, a game based on the way of thinking (the epistemic frame) of some important community in the real world (Shaffer, 2007). In a similar way, we could imagine creating a whole host of games that could test how well students are able to think like journalists, architects, mathematicians, historians, engineers, physicists, doctors, biologists, and so on. And these epistemic games could test whether students can solve complex problems using 21st Century skills.

Conclusion

We want to reiterate that epistemic games like *Urban Science* are just one example of how learning games can—and should—be used as assessment games. Games support and require the kind of learning that we need in the 21st Century—and so they have the potential to usher in the new paradigm that we need to develop the skills students need. It is in this sense that we suggest that instead of building games for learning we should be tackling the more difficult—but more fundamental—problem of assessment.

There is a good reason why the field of educational computer games (and educational technology more generally) has been looking at learning: Those of us who study educational games needed time to experiment with and to understand the kind of learning that computers make possible. But we all know what happens when we try to bring good games for learning into schools. Parents and teachers rightly ask: Will this help my children do well on the tests they need to pass?

What would happen, however, if students were tested—and schools judged—not by how well they perform on our current tests, but on whether they could solve real-world problems the way real-world professionals do.

Assessments drive the learning in which a system will engage. Today's standardized assessments, coupled with a punitive accountability model, encourage and support a skill-and-drill system of learning that does not lead to problem solving, innovation, or preparation for future learning.

We are arguing that games can be an assessment system, too. That system would also drive teaching and learning, but it would drive schools where we want them and need them to be: toward a new paradigm.

There is a great deal of research and policy work today concerned with how to change our standardized testing system and how to build deeper, more authentic forms of assessment. All of that work suggests that any change will be contentious and difficult to achieve. (Abell & Lederman, 2007) But we believe that one way to achieve a system wide change is to provide assessment and accountability tools that demand better forms of learning—and that properly designed games can do a better job of showing us where learners are in a course of development towards mastery than can any standard paper-and-pencil test.

They can do this because games use actual learning as the basis for assessment: their assessments are built on problem solving and facing challenges. They test not only current knowledge and skills, but also preparation for future learning. They measure 21st Century skills like collaboration, innovation, production, and design by tracking many different kinds of information about a student, over time. They can account for the fact that different students have had different opportunities to learn inside school and out. And they can provide students,

administrators, parents and teachers with feedback they can use to make decisions that support good learning.

A game like *Urban Science* can be a completely new kind of standardized test. It is standardized, in the sense that every player can have an experience designed in advance. It is a test in the sense that it can return a score or scores that indicates how well a player has done in the game. But it is a test that measures not the basic facts and basic skills of our current testing regime, but the kind of thinking that we value in the 21st Century.

The road to better schools starts by making the tests in school more like the games that students are already playing out of school.

Acknowledgements

This work was funded in part by the Macarthur Foundation and by the National Science Foundation through grants REC-0347000, DUE-091934, DRL-0918409, and DRL-0946372. The opinions, findings, and conclusions do not reflect the views of the funding agencies, cooperating institutions, or other individuals.

References

Abell, S. K., & Lederman, N. G. (2007). *Research on Science Education*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Bagley, E. A. S., & Shaffer, D. W. (2009). When people get in the way: Promoting civic thinking through epistemic gameplay. *International Journal of Gaming and Computer-Mediated Simulations, 1*(1), 36-52.

Beckett, K. L., & Shaffer, D. W. (2005). Augmented by reality: The pedagogical praxis of urban planning as a pathway to ecological thinking. *Journal of Educational Computing Research, 33*(1), 31-52.

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* (Vol. 24, pp. 61-101). Washington D.C.: American Educational Research Association.

Chi, M., Feltovich, P., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science, 5*(2), 121-152.

Csikszentmihalyi, M. (1996). *Creativity: Flow and the Psychology of Discovery and Invention*. New York: Harper Perennial.

Darling-Hammond, L. (2006). Securing the Right to Learn: Policy and Practice for Powerful Teaching and Learning *Educational Researcher, Vol. 35, No. 7, 13-24 (2006), 35*(7), 13-24.

Diamond, J. M. (2005). *Guns, germs, and steel : the fates of human societies*. New York: Norton.
Gardner, H. (1991). *The Unschooled Mind: How Children Think and How Schools Should Teach*. New York: Basic Books.
Gee, J. P. (2003). *What Video Games Have to Teach Us About Learning and Literacy*. New York: Palgrave/Macmillan.
Gee, J. P. (2004). *Situated Language and Learning: A Critique of Traditional Scooling*. London: Routledge.
Gee, J. P. (2007). *Good Video Games and Good Learning: Collected Essays on Video Games, Learning and Literacy*. New York: Peter Lang.
Hutchins, E. (1995). *Cognition in the wild*. Cambridge, MA: MIT Press.
Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. Cambridge, MA: Cambridge University Press.
Leadbeater, C., & Miller, P. (2004). *The Pro-Am Revolution: How Enthusiasts are Changing our Society and Economy*. London: Demos.
National Research Council (1995). *National Science Education Standards*. Washington, DC: National Academy Press.
Partnership for 21st Century Skills (2004). *21st Century Readiness for Every Child*. Tucson, AZ.
Schwartz, D. L., & Arena, D. (2009). Choice-based assessments for the digital age. Stanford University.
Shaffer, D. W. (2006). Epistemic Frames for Epistemic Games. *Computers and Education, 46*(3), 223.
Shaffer, D. W. (2007). *How Computer Games Help Children Learn*. New York: Palgrave.
Shaffer, D. W., & Gee, J. P. (2005). *Before every child is left behind: How epistemic games can solve the coming crisis in education* (Working Paper): University of Wisconsin-Madison, Wisconsin Center of Education Research.
Shaffer, D. W., Hatfield, D., Svarovsky, G., Nash, P., Nulty, A., Bagley, E., et al. (2009). Epistemic Network Analysis: A prototype for 21st Century assessment of learning. *The International Journal of Learning and Media, 1*(2), 33-53.
Squire, K. D. (2004). Sid Meier's Civilization III. *Simulations and Gaming, 35*(1).

.