REFLECTIONS ON EMPIRICAL EVIDENCE ON GAMES AND LEARNING

James Paul Gee

Mary Lou Fulton Presidential Professor of Literacy Studies

Arizona State University

james.gee@asu.edu

Abbreviated Title: Reflections on Empirical Evidence

James Paul Gee
50 Mesquite Circle
Sedona, AZ 86336
(608) 345-5258
James.gee@asu.edu

Games and Learning

Over the last few years there has been a growing interest in video games and learning (Gee, 2007; Salen, 2007; Shaffer, 2007; Squire, 2006). Some researchers have been interested in the potential of commercial games for learning. Others have been more interested in so-called "serious games", games specially built to enhance learning in a variety of areas. There are now a good number of research and game-design projects across the globe devoted to non-entertainment-based learning games.

So far, in my view, the evidence for and against video games as a tool for learning is not deep, especially for more modern sorts of games (see Chapter 8 of this book for more nuanced overviews). This is not surprising, since the field of games and learning, at least in its current form, is new (though the area of simulations and learning is much older). Nonetheless, now that a good deal of work has been done making the argument that games are good for learning, it is surely time to engage in robust empirical projects to test this hypothesis.

There are, of course, a number of studies devoted to testing empirically whether games enhance learning (for a review see Chapter 8 of this book). In my view, the evidence collected so far is a mixed bag. However, I do not intend to evaluate or even comment on that evidence here. What I want to do here is argue that a good deal of work needs to be done before one can offer convincing and relevant evidence in regard to games and learning. In my view, much of this work has not yet been done. This work has implications for what needs to be stated clearly in any empirical study of games and learning if the study is to be fully relevant and meaningful.

To begin to see the work that must be done before one can offer convincing and relevant evidence on games and learning, let's start with the hypothesis: "Video games are good for learning" (Gee, 2003, 2004, 2007). Obviously this hypothesis is untestable as it stands. The hypothesis has no clear meaning until we specify what is meant by "video games", "learning", and "good". Defining each of these is not an easy matter and the definition of each involves a good deal of controversy.

Video Games

Let's start with "video games". There are people in Game Studies who argue that all video games (even all games) share some necessary and sufficient features and, thus, that "video game" is a unitary category (e.g., see Juul, 2005). I do not myself believe any such thing. But if you do believe this, and want to test whether games are good for learning, you need to offer an argument that "video games" are a general unitary category and tell us what this category is.

In my view, there are many different types of video games and the category of "video game" is not a unitary one with necessary and sufficient conditions (Gee, 2007). At best, the category of "video game" is what Wittgenstein (1958) called a "family resemblance" concept (and, ironically, Wittgenstein used the concept of a "game" as his prototype example of a family resemblance concept). In my view, different types of games are very likely to have different types of effects, including different types of learning effects.

There are casual games and games played by non-casual gamers. There are what I call "problem games" and "world games", though that distinction is not air tight (Gee,

2007). Problem games focus on solving a given problem or a single class of problems (e.g. *Tetris*, *Diner Dash*), while world games simulate a wider world within which the player must solve many different sorts of problems (e.g., *Half-Life*, *Rise of Nations*). There are different genres of video games, such as real-time-strategy games, first-person shooters, adventure games, sports games, god games, platform games, role-playing games, and others.

So far no one has given a widely agreed upon break down of the types of video games there are or could be. Even the genres of games listed on game sites differ from each other. So if researchers want to collect evidence that games are good for learning, and they believe (as I do) that there are quite different types of video games that do not fall into one unified category and which have different effects, then they must specify quite clearly what type of game they are testing and how this type relates to other types. This is hard to do in the absence of a good theory of types or genres.

Games of any type can be good or bad. Players can like them and devote much time, attention, and energy to them or not. A game can motivate players highly or not. It is obvious that gaining empirical evidence that a bad game does not work to enhance learning is not an interesting result. It is a "so what?" result. So what you found a bad game did not work well for learning? No one argues—or thinks—bad games are good for learning. Rather, people have claimed and believe that good games are good for learning.

So when researchers want to test a game—gain empirical evidence that a game, or a category of games, or games in general enhance learning—they must first show it is a good game. The problem, of course, is we so far have no good theories about what

makes games good and bad. We certainly know that gamers do not always agree whether a game is good or bad, although there are certainly games that are widely praised (e.g., *Half-Life*, *Civilization*, and *Castlevania: Symphony of the Night*). "Goodness" in games may involve things like motivation, the right level of challenge, interesting problems, a good story (though this is controversial), and surely involves good game mechanics. It surely involves other things as well. But, as I said, as of yet, we have few good theories here.

It is an interesting—and thus far unasked and unanswered—question as to whether the standards of goodness for commercial games and so-called serious games are or will be the same. If it turns out they are same, this means serious games being tested empirically for their learning effects must meet a pretty high standard. This is not a standard solely to do with cost (there are relatively inexpensive commercial games that are good games, for example, *Braid*, *Flower*, and *Shadow Complex*). In all likelihood, it has more to do with game mechanics, which are the heart and soul of game design.

There are manifestly different types of gamers and of learners. Thus, too, it is crucial for anyone seeking evidence that games are good for learning to specify clearly what category of gamers and learners (or categories, if there is more than one) they are making claims about in their work. Likewise, showing that a game, a type of game, or games as a whole work or don't work for one category of gamer or learner does not show they work or don't for others categories of gamers and learners.

I want to stress that we have to worry about both types (categories) of gamers and learners, since different people not only learn in different ways, they play games in different ways (e.g., not at all, casual gamer, regular gamer, hard core gamers). When we

get to categories of gamers like that developed by Richard Bartle (2004) for players of massive multiplayer games—Killers, Socializers, Achievers, and Explorers—we almost certainly have categories that are related to categories of types of learners, and, thus, the categories of gamers and learners become confounded. It is important to see that types of gamers goes well beyond how much time someone plays video games.

Finally, we come to, perhaps, the most important point and yet one that has played little role in the empirical research on games and learning. Some people who have made the claim that games are good for learning are not making a claim just about a piece of digital software (Gee 2004, 2007; Gee & Hayes, to appear). For me and other games-and-learning scholars (e.g., Barab, Zuiker, Warren, Hickey, Ingram-Goble, Kwon, Kouper, & Herring 2007,; Shaffer, 2007) it is important to distinguish between the game as software (let's call this the "little 'g' game") and the whole social system of interactions that players engage in inside (for multiplayer games) and around the game (this is sometimes called the "meta-game"). We can call the combination of the game (software) and the meta-game (social interactional system) the "big 'G' Game".

Games involve lots of social interaction. This is obvious for multiplayer games where gamers are playing competitively or cooperatively together (Taylor, 2006). But many single player games have multiplayer versions and even ones that do not often still inspire a good deal of discussion and sharing on web sites that contain discussion forums on various topics, strategy guides, and often opportunities to learn to "mod" (modify) or use mods of the game. Those of us who have made claims about good commercial games being good for learning were talking about games embedded in their social

systems, not by themselves or, worse, embedded in artificial or poorly designed social systems made up by researchers who don't game.

Some people who have made serious games for education—*Quest Atlantis* at Indiana University would be a good example (Barab, et al., 2007)—have seen themselves as implementing a game (the "little 'g' game") inside a social system (the meta-game) of interactions among gamers/learners that are meant to integrate with, complement, enhance, and expand the learning done in the game world. These researchers are not, in fact, claiming that games (little 'g') are good for learning. They are claiming that Games (big 'G': game + meta-game) are good for learning.

Thus, researchers seeking to test games empirically for learning need to specify clearly what was in the game (or games) they tested and what sort of meta-game (if any) was in and around this game or these games. In the case of serious games or commercial games put into an educational context, empirical researchers need to specify what was in the game they tested and what sort of social learning system was built (if any) in and around this game (where by a "learning system" here I mean a learning-enhanced meta-game). If researchers are testing only games and not Games (game + meta-game), then they need to say so and acknowledge that their research does not bear on claims being made by people about Games and not games.

We said above that it is pointless to test a bad game. The same is true of meta-games and Games. We need to show—particularly if we have found negative results—that the game, meta-game, and Game (their combination) we have tested are all good. Again, however, we have as yet few good theories about goodness in these respects.

To summarize the discussion so far: Researchers seeking evidence that games are good for learning need to say what "game" means. If they believe there is a unitary category of video games (or games more generally) they need to say what it is. If they believe "game" is a family resemblance concept, they need to say what types of games they used within some coherent view of game types. They need to show the games they used were good, again within some coherent view of what constitutes goodness in games. They need to specify what types of gamers and learners they are making claims about and any confounding between these two. Finally, they need to be clear about what was in their game and in their meta-game (if any) and how the two related to each other as a Game to enhance learning or not.

Learning

Let's now move on from "game" to "learning". There are, of course different and contested theories of how human learning works in general and how it works for specific purposes. My own view of learning is a sociocultural-situated-embodied-extended mind type of theory (Barsalou, 1999; Brown, Collins, & Dugid, 1989; Clark, 2008; Gee, 1990, 2004; Glenberg, 1997). Others hold different theories. But it does not matter what theory I hold. What matters is which theory a given piece of empirical research on games and learning is based on and is, thus, testing.

Let me say here that I take as a fundamental tenet of empirical science that science, at least mature science, does not test isolated claims but only theories from which the claims follow (can be deduced) (Gee, 2005; Quine, 1951). When evidence goes against a claim, we must think not just about the possibility of asserting the opposite

of the claim, but also about the possibility that some premise in our theory is wrong and the claim is correct. In a complicated theory this can be a difficult decision.

One reason why it is true that science tests theories and not facts or claims in and of themselves is that any claim only has meaning based on a given theory. We have seen this with the claim that "Good video games are good for learning". Different theories give different meanings to "video game", "good" (in both uses), and "learning". You aren't testing the claim by itself, but the theory that gives it meaning.

Another reason why it is true that science test theories and not facts or claims in and of themselves is that when evidence is gathered against a claim one always has the choice either to give up the claim or to change some other aspect of one's theory. For example, if you show that some video game that fits my theory's criteria of "good" does not work for learning, then I can give up my claim that "Good video games are good for learning" or I can change my criteria of "good" and claim your game is not really "good". This is a common move in mature sciences (because in many cases data are copious and cheap, but good theories are hard to come by, so we resist giving them up until new theories, and not just new claims, are available). What this move means is that we have to debate theories (e.g., what the criteria of "goodness" are) and not just facts or claims.

This view of science means that any researchers testing games for learning must state clearly what theory of learning they are testing. Further, they must, if they get negative results, engage in debates about theory construction (i.e., suggest an improved theory) and not just claim they have shown a specific claim to be false. Researchers need to be careful about (and engage in debate about) whether they have shown some criteria,

definition, or bridging principles in a theory to be inadequate or whether they have actually shown a substantive and important claim to be false.

Anyone who is claiming good games (really, as we have seen, Games) are good for learning is really claiming "Good games that implement theory of learning X are good for learning", where "X" is the theory of learning the person espouses. Showing that a game that implements theory of learning Y does not work for learning is not going to be relevant to people who have made the claim that "Games that implement theory of learning X are good for learning".

Let me give a specific example. I do not myself care about skill-and-drill learning done in meaningless contexts. I do not believe that humans learn at their best or learn the best things in that way. So if you show me that a game that implements such learning works or doesn't, while I may find it interesting, I do not find it relevant to any claims I have made.

I happen to believe that games might be a more motivating setting for decontextualized skill-and-drill than other settings (e.g., drill sheets in school). But I do not believe decontextualized skill-and-drill is how humans learn in deep and meaningful ways and I do not care about, or care to spread, such learning. What would be important and meaningful to me is research that showed that good games (actually, Games) that implemented sociocultural, situated, embodied, extended mind learning well worked or did not work well for conceptual learning that led to problem solving and not just rote learning. That is the only thing I have made claims about.

Of course different scholars hold different theories of learning and certainly not all agree with mine, nor need they. But the point is not that anybody needs to agree with

my theory of learning. The point is that researchers testing games for learning need to specify clearly what theory of learning their game implemented and they need to show the game implemented that theory well.

There is an additional problem with theories of learning that complicates empirical research on games and learning. Many theories of learning are complex and predict a complex set of interacting outcomes. Showing anything about one outcome in this set in isolation all by itself tells us nothing important about the theory being tested.

Let me give an oversimplified example to make the point. Say person A believes that human learning works best by cutting large tasks into a sequence of small tasks arranged in a linear sequence, each one of which is assessed before moving on to the next. And say person B believes that human learning works best when the sub-skills of a larger task are kept within the setting of that task and often practiced as an integrated part of the whole larger task. Further, they believe that sub-skills need to be assessed as they function in the whole. Person A and Person B hold different theories of learning, theories A and B (and it is possible that one theory works better for some tasks and the other theory works better for other tasks).

Now say that Person A with theory A does a piece of research comparing a game that implements theory A and a game that implements theory B. A finds that learning results on sub-skills were better in game A than in game B. A then goes on to claim that theory A is better than theory B. However, he really can claim no such thing.

Imagine B now shows that a game implementing theory B gets results in sub-skills that are, in the end, as good as, but slower than a game implementing theory A. Further, there is a gain over the game implementing theory A on how well these sub-

skills are integrated and on how innovative the learners could be with new combinations of the sub-skills. Now A's claim about sub-skills (that they are acquired better in a game implementing theory A than a game implementing theory B) does not look all that significant and does not tell the whole story.

The only interesting comparisons between a game implementing theory A and a game implementing theory B is on the whole set of interacting and integrated outcomes theories A and B predict. Testing one of the predicted outcomes (e.g., sub-skill acquisition) in isolation is uninteresting and irrelevant. Thus, researchers seeking to test games empirically for learning outcomes need not just to be clear about what theories of learning they are testing, they also need to be careful to test the whole set of interacting and integrated outcomes the theory predicts or at least to consider important interactions among different predicted outcomes (e.g., that sub-skills may be slower, but integration of sub-skills and innovation may be higher).

This claim about testing sets of interacting and integrated outcomes, not isolated components, is true of testing all theories of learning whether a game is involved or not. But adding a game to the mix does complicate matters a good deal. This is because theories of game design and of learning are confounded.

A person who believed theory A above would design a game that gamers call "linear" and "on rails" and, indeed, there are such games (including good ones) and this is one theory about how to design games. A person who believed theory B above would design a game that was less linear and more open-ended. And, indeed, there are such games (including good ones) and this is one theory of how to design games.

Just as different theories of learning predict different and complex sets of interacting and integrated outcomes, so, too, games designed in different was are likely to give rise to different and complex sets of interacting and integrating outcomes in regard to learning and other things. Thus, a researcher testing a game for learning outcomes needs to worry about the complex interacting and integrated outcomes of both a theory of learning and a theory of game design and, also, their interaction.

Let me say, here, finally that any extreme view of theory A would give rise to games that are more linear and "on rails" than even linear commercial games tend to be. Even linear commercial games integrate sub-skills (which are often segregated and practiced on each level) within actions and tasks that are not primarily focused on the sub-skills themselves. A game that implemented an extreme view of theory A would probably be a bad game—though it may be that non-game digital technologies could better implement an extreme view of theory A.

Two Types of Games

Above I pointed out that there are many different types of video games. One distinction one can make between types of games, a distinction I also mentioned above, is between what I call "problem games" and "world games", though that distinction is not air tight. Problem games focus on solving a given problem or a single class of problems (e.g. *Tetris*, *Diner Dash*), while world games simulate a wider world within which the player must solve many different sorts of problems (e.g., *Half-Life*, *Rise of Nations*).

One could imagine making a problem sort of game that allowed players/learners to understand a specific problem or class of problems in physics (e.g., problems connected to conservation laws). Such an understanding would mean knowing how to solve problems, not just pass paper-and-pencil tests about them. However, in my view, no one has made such a "serious" game that is as good as, say, the commercial game *Portal*, which allows players to experience and use principles like the conservation of momentum, though it does not teach any physics. If one really wants to test claims about games being good for learning, we need physics games as good as *Portal*.

One could imagine a game where players/learners learned both how government and civil society works and how to participate actively in society to bring about change. One could imagine a game where players/learners came to understand and be able to intervene in the complex and interacting properties giving rise to global warming and its natural, social, economic, and political effects. However, no one, in my view, has made such a game that is as good as the commercial game *Civilization*, a game which lets players experience and use principles about historical change and the clash of civilizations, but does not set out primarily to teach history, civics, or environmental science. If one really wants to test claims about games being good for learning, we need civics and environmental games as good as *Civilization*.

So while we most certainly need to collect evidence as to whether games are good for learning—and this is hard, complex work—we also need to keep working on our theories of games and learning and on our game design. And we need to closely integrate all three.

Conclusion

My point in this paper is simply to point out some work that needs to be done to make empirical claims about games and learning clear and relevant. Since we do not yet have entirely good theories of game genres, what makes a game good, the nature of good (big 'G') Games, the categories of learners and gamers and their interactions, and the predicted interacting, integrated outcomes of combinations of theories of learning and game design, there is a lot of work that remains to be done. It can only be done, in my view, if empirical and theoretical work stay in close contact. Simplistic empirical research, just like simplistic theories, does not buy us much.

REFERENCES

]Barab, S., Zuiker, S., Warren, S., Hickey, D., Ingram-Goble, A., Kwon, E-J., Kouper, I., & Herring, S. C. (2007). Situationally embodied curriculum: Relating formalisms and contexts. *Science Education, 91*(5), 750–782.

Barsalou, L. W. (1999). Language comprehension: archival memory or preparation for situated action. *Discourse Processes*, *28*(1), 61–80.

Bartle, R. (2004). *Designing virtual worlds*. Indianapolis, IN: New Riders.

Brown, A. L., Collins, A., & Dugid, P. (1989). Situated cognition and the culture of learning. *Educational Researcher*, *18*(1), 32-42.

Clark, A. (2008). *Supersizing the mind: Embodiment, action, and cognitive extension*. Oxford: Oxford University Press.

Gee, J. P. (1990). *Sociolinguistics and literacies: Ideologies in Discourses*. London: Taylor & Francis, Third Edition, 2007.

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 schooling*. London: Routledge.

Gee, J. P. (2005). It's theories all the way down: A response to *Scientific Research in Education*. *Teachers College Record, 107*(1), 10-18.

Gee, J. P. (2007). *Good video games and good learning: Collected essays on video games, learning, and literacy*. New York: Peter Lang.

Gee, J. P. & Hayes, E. R. (2010). *Women and gaming: The Sims and 21st century learning*. New York: Palgrave/Macmillan.

Glenberg, A. M. (1997). What memory is for. *Behavioral and Brain Sciences, 20*(1), 1–55.

Juul, J. (2005). *Half-Real: Video games between real rules and fictional worlds*. Cambridge, MA: The MIT Press.

Salen, K. (Ed.). (2007). *The ecology of games: Connecting youth, games, and learning* Cambridge, MA: MIT Press.

Shaffer, D. W. (2007). *How computer games help children learn*. New York: Palgrave/Macmillan.

Squire, K. D. (2006). From content to context: Video games as designed experiences. *Educational Researcher*, 35(8), 19-29.

Taylor, T. L. (2006). *Play between worlds: Exploring online game culture*. Cambridge. MA: The MIT Press.

Quine, W. V. O. (1951). Two dogmas of empiricism. *The Philosophical Review*, *60*(1): 20-43.

Wittgenstein, L. (1953). *Philosophical investigations*. Oxford: Blackwell Publishing.