Researching and Developing Serious Games as Interactive Learning Instructions

Christian Sebastian Loh, Southern Illinois University-Carbondale, USA

ABSTRACT

As serious games gain momentum in the academic arena, no doubt more educators and instructional technologists will begin considering the possibility of making their own games for instruction. As developers of instructional resources, instructional technologists need to steer clear of producing more 'video' games, and instead, developing more 'serious' games that incorporate both learning and assessment. The research community needs to learn from tested processes and best practices to avoid repeating old mistakes. The model for serious game making presented in this article has been used successfully for the creation of an award winning project, and will now be shared for the benefits of fellow researchers, educators, and instructional technologists.

Keywords: *Instructional Development, Research Methods, Modding, Interactive Learning Instruction, Serious Games*

INTRODUCTION

Games and education have had a long-standing partnership for a large part of the known human history. Botturi and Loh (2009) showed that the ancient Greek used only one word, *ludus*, to mean both *school* and *game*, as learning and playing games were once considered to be the same. School teachers of that time were referred to as *magister ludi* (literally, game masters) because they were experts who drew upon the principles of game playing for the training and instruction of their pupils. Based on this *game-is-education* perspective, the use of digital video games for serious learning can hardly be called revolutionary. Hence, when nearly all (99% of boys and 97% of girls) teenagers report playing video games regularly as a preferred pastime (Lenhart, Kahne, Middaugh, Macgill, Evans, & Vitak, 2008), many educators acknowledged this to be the key to the hearts and minds of the digital native generation (Miller, 2008).

The video game industry had always stayed on the cutting edge by pushing for advancement in digital (graphic) technology. When coupled with the passion among game developers to out-do one another, this has given rise to an industry that is relentless in its pursuit for products with ever-escalating production qualities. Compared to just a few years ago, not only are players able to perform a lot more actions

DOI: 10.4018/jgcms.2009091501

within a game environment; the shelf-lives of commercial, off-the-shelf (COTS) games are constantly diminishing, being given over to newer games to fuel tomorrow's technology. This means that many well-known computer games (such as *The Oregon Trail*, *Math Blaster*, and *Reader Rabbit*) were not only outdated, but would cease to work on the newest computers. Even the abundant 2D-animation (Flash) games found on educational websites would pale in comparison to what the industry offers today.

Since the debut of 3rd generation game consoles (such as PS3 and Xbox 360), today's game engines can easily simulate real-world physical laws (such as gravity and inertia), and produce realistic lighting and water effects in games. As faster computer processors and online streaming technology continue to provide support for better game effects, immersive online play, and massive multiplayer virtual worlds; the knowledge gap between the gaming industry and outsiders to the industry (such as educators and researchers) will continue to widen. There was little reason to conclude non-professional game developers could ever create games at the industry production quality. Why then, should educators and instructional technologists care about making video games?

HISTORY OF GAME MODIFICATION (MODDING)

The watershed came in the form of a military training game, called *Marine Doom* (1998), created for the purpose of training soldiers in teamwork and decision making skills when live training time and opportunities were limited during peaceful times. Instead of creating the video game from scratch, the U.S. Marine (in-house) development team decided to modify (or, *mod*) a COTS game, *Doom* (1992), to take advantage of the game mechanics and resources already present in the game engine, as well as to reduce production cost and time. This game *mod*ification process—whereby a COTS game's own engine is re-used to create a "home-brew" (and very much playable) game—has come to be known as *modding* among the gamers. Since then, the U.S. Marine Corps have gone on to create other military game modules (or *mods*), including the highly successful *America's Army*, with over 26 versions released since its debut in 2002. Gamers have easy access to thousands of game *mods* (made from a plethora of COTS games) that were distributed through repositories and websites created just for mod enthusiasts—for example, the Vault Network (http://vault.ign.com), and the Game Mod Database (http://www.moddb.com).

Instead of producing a full-fledged video game, it would be far more likely for educators, researchers, and trainers to develop prototypic games for the demonstration of educational concepts, research frameworks/methodologies, or training procedures. As such, game modding would prove to be most appropriate and invaluable in reducing development cost and time while attaining industry production quality in the artifact produced. Because a game mod would be of a similar feel and quality to the original COTS game used to create it, learners would be motivated by the medium and willing to learn the new training tool. Similarly, research projects using mod of well-known COTS games could benefit from easier recruitment of human subject participants and reduce the need for (re-)training.

Increasingly more game developers have chosen to give away game development kits (GDKs) along with the sale of their games; a move which is sure to encourage more modding projects. Priced at US$20–50 per game, these GDK/game bundles are a small price to pay, when compared with commercial-grade game engines (e.g., *Unreal*), which could cost up to several hundreds of thousand dollars per seat of license. Examples of GDK/game bundles include the *Crysis Mod SDK* for *Crysis*, the *Electron Toolkit* for *Neverwinter Nights 2*, and *Hammer Editor* for *Half-Life 2*. Several of these tools have also been uses in the creation of educational projects, research test-beds, and workplace training. Notable projects in this arena include *HistoriCanada: The New World* (a *Civilization III* mod) by Bitcasters (completed in 2008), and *Revolution* (a *Neverwinter Night* mod) by Education Arcade (completed in 2004). Some educators have come to regard game mods as computational literacy artifacts (Steinkuehler & Johnson, 2009); hence, students who engaged in modding must possess high standards in computational and literacy skills.

THE RISE OF SERIOUS GAMES

The success of video games as an educational and recruitment tool (e.g., *American's Army*) (Bounds, 2007) has prompted renewed interest among educators and researcher to re-examine digital games for raising literacy (Gee, 2003), as well as for supporting research and instruction (Ferdig, 2008; Miller, 2008). The mounting interest eventually translated into the Serious Games Summit 2004, marking the beginning of a new sub-industry known as *serious games* (Sawyer, 2005). The term was chosen to encompass any type of digital game and game-like application (including simulations and virtual worlds) that had been specifically designed for serious learning or training purposes. Even though the serious games industry began as a niche market targeting human performance improvement and personnel training, it has attracted the attention of several industries, including the military, healthcare, business, government, and educational sectors. As one study (Hewitt, 2008) revealed, out of 70% of major U.S. corporations that were already using interactive software for human performance improvement, many had expressed interest in digital game-based training.

SERIOUS GAMES AS INTERACTIVE LEARNING INSTRUCTION

Unfortunately, the last joint effort between game developers and instructional designers to create (revenue generating) educational video games was less than successful. As the fateful name suggested, those early *edutainment* titles were half-baked attempts at doing *edu*-cation and enter-*tainment* concurrently; thus, resulting in a large collection of *boring* games (Hopson, 2006; Prensky, 2005; van Eck, 2007). Since history has revealed that the notion of creating educational games simply by mixing learning materials with "gaming activities" was wishful thinking, serious games developers have exercised much caution in what they produced. They chose to focus on discovery (or exploratory) learning, which dealt more with critical thinking and problem solving skills, than trying to sensationalize fanciful gaming activities, such as shooting objects, navigating pathways, or *twitching* (i.e., pressing quick combination of button in set order).

Unlike the edutainment titles which were available for sale through nation supply chains, such as WalMart and BestBuy, most serious games chose *not* to compete with COTS game titles. Instead, these serious games were either distributed at no cost through special educational downloads (Amarelo, 2008; Federation of American Scientists, 2006), or through direct marketing to specific target industries (e.g., fire fighting, police enforcement, and disaster response). Examples of serious games include: *Immune Attack* (high school and college immunology, by the Federation of American Scientists), *Flame-Sim* (fire fighting, by Flame-Sim, LLC.), and *Tactical Iraqi* (foreign language acquisition, by Alelo, Inc.).

Assessment of Learning

Readers may question what advantages serious games have over edutainment, if it is to become successful. Having interviewed experts from the industry, Michael and Chen (2006) suggested that the *assessment of learning* is the key element in making serious games better than edutainment. Because "data-driven assessments of learning" have become the foremost issue on the minds of many trainers and educators today (Mandinach & Honey, 2008), the inclusion of assessment components in serious games would allow instructors and trainers access to game-related statistics (that were otherwise unavailable) for the creation of performance improvement and Return of Investment (ROI) indices.

Depending on the training industry, these game-related data may be tailored to include any pertinent information; for example, amount of time taken to complete a task, total number of cases solved and percentage of tasks/missions accomplished. In most cases, a *screen-dump* (where mission-related achievement scores are shown just *once*, on screen, at the end of a game mission) may be sufficient for self-evaluation by the players. In other cases, a game *log* file may be necessary to provide more permanent data for evaluation by instructors. Figure 1 shows an example of a (primitive) game log in plain text format; Figure 2 shows a more comprehensive game log in tag-enclosed XML format.

Besides game logs, military training games and simulators must further support the feature of After-Action Reports (AAR)—essentially a form of graphical game log, detailing a variety of statistics indicating how much rank and experience a unit gained during a simulated operation. Although AARs could also be found in COTS games like *Call of Duty IV* and *Tom Clancy's EndWar*, they were included mostly for the sake of military gameplay authenticity and served no training purpose beyond the game. Unlike their game counterpart, real AAR from live-training sessions (e.g., a flight simulator with 3-axis tilt) could include biofeedback data collected during *test flights*. Such biofeedback data has been useful to predict soldiers' performance and the likelihood of adverse physiological reactions (e.g., gravity-induced loss of consciousness during a flight) in future, real life operations. Since the assessment of learning (and performance), as well as data visualizations, are salient features that set serious games well above edutainment. It is important to design these features into future serious games, and not to regard them as *optional* features—to be added only as an afterthought.

Media Comparison Research

In many educational sectors where scientific methods of research were upheld—including Medical/Pharmaceutical studies, Learning Science, Agricultural Education, Science, and Computer Science Education—the research design would typically call for a comparison between treatment and control groups. Researchers would test for the effects of an intervention by comparing the recipient group with a control group that did not receive the treatment. The rationale for this design being: should the intervention (e.g., a new drug in a clinical research) prove effective for the condition (e.g., cancer), a measureable change (positive or negative) would be detectable in the treatment group, and would not be present in the control group (i.e., use of placebo, or no intervention). If no statistically significant difference between the intervention and control groups could be found, then the intervention was regarded as having no effect (ineffective) for the recipient group under the circumstances.

For this type of research design to be applied in digital game-based instruction, researchers must make the assumption that the educational interventions (or technology used) were similar in effects to chemical/biochemical/pharmaceutical used in agriculture/medical/clinical interventions, respectively. If technology or technology-based instruction was the *magic pill*, then the correct control condition (to counter the treatment condition) would be no pills, or

Figure 1. A game log in plain-text format

```
Alice (Fordlindon, 7 armies, 3 lost) attacked Bob (Harlindon, 1 armies, 1 lost),
conquering it. 4 armies advanced.
Alice (Harlindon, 4 armies, 3 lost) attacked Bob (The Shire, 13 armies, 0 lost),
failing to conquer...
Bob.....
```

Figure 2. A game log in XML format


```
<?xml version="1.0"?>
<game>
<maptype>Africa</maptype>
<timeleft>12hrs 16min 20sec</timeleft>
<round>3</round>
<players>
<starting>8</starting>
<surviving>4</surviving>
<teams>4</teams>
<player>
<name>trey</name>
<rank>New Recruit</rank>
<team>3</team>
<cards>4</cards>
<armies>20</armies>
...
<event>
<type>deploy</type>
<from>NA</from>
<to>Sparta</to>
<armies>3</armies>
<defender>NA</defender>
</event>
<event>
<type>attack</type>
<from>Sparta</from>
<to>Yarmen</to>
...
```


placebos. In the case of serious game learning, a treatment group would naturally be playing serious games, and a control group would be receiving teacher-only instruction.

A Flawed Assumption

However, since the 1980s, these types of treatment/control studies—more commonly known as *media comparison studies* in educational and instructional technology circles—have long been criticized as flawed based on many inherent theoretical and design problems (Locke, Moore, & Burton, 2001; Thompson, Simonson, & Hargrave, 1992). Although the comparison design was a proven method in many scientific fields, they were deemed to be an "inappropriate research design for measuring the effectiveness of instructional technology" (Lockee, Burton, & Cross, 1999, p. 33). Researchers who conduct media comparison studies have overlooked inherent factors found only in human learning, including learner characteristics, media attributes, instructional strategies, and influence of teachers, that were not present in plant growth, biochemical pathways, or human physiology. Technology (such as serious games) in and of itself could not affect learning, it was the technology-mediated instruction that would affect learning in the learners (Haertel & Means, 2003).

Since the research question of comparing instructional media/method to 'direct instruction' (as control condition) was inherently confounded, any discussion of the invalidated research findings would become meaningless. This explains why media comparison studies (such as the case study above) would commonly yield *no (statistically) significant effect* findings, which could be doubly damaging. Firstly, policy-makers might interpret the findings to mean the intervention was ineffective for learning since there was no measureable effect (and such were the criticisms put forward by pundits). Secondly, researchers were led about on a wild-goose chase, as they tried to improve upon the (poor) experiment believing in the presence of a Type II (false-negative) error—when it was the research design that was flawed.

To illustrate the problems and flawed arguments commonly found in this type of research, a media comparison study is presented in Table 1 to serve as a case study. The case was taken from a true research study conducted in 2009 in a rural high school in the United States. Details were withheld/ altered to protect the identity of the researcher and the institution.

Just As Effective?

History revealed that media comparison studies tended to be conducted when researchers tried to justify (or prove) the effectiveness of a new instructional technology to stakeholders, by comparing it against traditional classroom instruction (or, direct instruction). Some researchers have argued for the *no significant difference* findings to mean the two media under comparison were *just as effective*, in

Table 1. Example of a media comparison study

Rationale: Because of the infancy of research on digital game-based learning, little is known about how to effectively design effective situated learning environments. This study adopts a design-based research approach to investigate the effects of digital game based learning based on gender, ethnicity, and socio-economic status of students. **Research Method:** A total of 250 students (10^{th} grade) from a rural high school participated in this study. Students were randomly assigned to one of the twenty classes. Out of the twenty classes, ten (with 150 students) were randomly assigned to the **Treatment condition** (playing a digital online game: name withheld) and ten classes (with 100 students) were assigned to receive the **Control condition** (direct instruction). There were 125 males and 125 females (comprised of 20 African American, 4 Asian American, 170 Caucasian, 1 Hispanic, and 55 Native American students). A total of 134 students received free or reduced lunch, 1 English language learner (ELL), and 15 special education students. **Procedures:** A set of pretest and post-test with parallel test items were developed by the researchers and a mathematics teacher to determine the effects of digital game play with MMOG on mathematics achievement. The pretests and post-tests were administered to students from both the treatment and control groups, prior and after the 5-month long implementation period. The first part of the instruments included a background survey to determine student involvement in digital game play, the types of games played, the amount of time spent playing games daily, and what a game must have to keep one engaged. The second part of the instruments included 20 multiple choice test items constructed from released 10^{th} grade math and Algebra II state test items and from sample test items from the state department of education's website. The structure of the math test was similar to the state's high-stakes tests, the 10^{th} grade criterion-referenced test (CRT) and the Algebra II end-of-instruction (EOI) test. **Findings:** A 2 (treatment, control) X 2 (male, female) X 3 (Caucasian, Native American, Other) X 2 (F/R, non-F/R) ANCOVA was conducted to examine the interactions of treatment/control groups, gender, ethnicity, and socio-economic status of students. The dependent variable was posttest results and the covariant was pretest results. While ***no significant main effects*** in mathematics achievement were found within group, gender, and socio-economic status. There was a significant interaction among group and ethnicity. $F(2, 191) = 3.14$, $p = 0.045$, $\eta^2 = 0.03$. **Interpretation:** Results suggested that while students' gender or social class did not have hypothesized impact, students' ethnicity may impact their learning outcomes with educational games. Our findings suggest that it may be very difficult to reach certain ethnic groups, such as the Native American population in our study, with digital game-based learning. Therefore, we recommend that in designing and implementing digital game-based learning environments educators have to consider cultural issues. We also recommend further studies in this area to identify why, for some ethnic groups, digital game-based learning may not be a viable educational alternative in the classroom.

order to circumvent the problem of having no strong reason to recommend a change in the instructional media (Locke et al., 2001). They have, in effect, failed to recognize that these findings were indicative of a poorly designed and confounded study.

As history tends to repeat itself, another wave of media comparison studies is currently underway; no doubt being conducted by serious games researchers who were overly eager to prove the superiority of serious games over traditional teaching. As evidenced in the above case study, having made the first mistake of "media comparison," the researcher had gone on to comment about the discovery of a *significant* interaction effect among group and ethnicity even when such findings make little sense: Why would race affect game playing? Frequently, unexpected interaction effects were detected in media comparison studies due to the presence of the confounding factors. However, instead of recognizing the unexpected findings to point to potential flaw in the research design, many researchers went on to commit a second mistake by trying to explain away the unexpected effect as a valid finding and even to recommend "further studies" in that direction. Researchers conducting further studies based on such poor advice would eventually arrived at even more bizarre results with no hope of finding any evidence pointing to the good of the technological intervention. In order to advance the field, serious game researchers must abandon media comparison research immediately, and recognize digital games as a new kind of instructional media, which must be evaluated through new research methodologies (see Haertel & Means, 2003). In the words of Hastings and Tracey (2004, p. 28), "After 22 years, it is time to reframe the original debate to ask, not *if*, but *how* media affects learning. We agree that media comparison studies are inherently flawed and support the argument that we must identify research designs that will provide answers to this question in significantly less time."

Better Serious Games Research

The many problems associated with "media comparison" studies should not preclude researchers from using the empirical research experiment involving intervention/control conditions as a research design. Those who must use a control condition in their experiments only need to be careful so as to not fall into the same trap. For example, the following four designs *all* made use of empirical research methodology. Three of these methods (A-C) involved comparing a learning method (i.e., game-based learning) against a controlled condition. The last method (D) was a whole new way of looking at designing serious game research: by turning the entire game into one big assessment. (Unlike typical media comparison studies, the research measurement employed by Methods A-C did not directly compare technology against traditional instruction by teachers (i.e., control), and hence, they were considered to be *meaningful* comparisons.)

A. **Compatible comparisons:** By designing two similar games (A and B) with different instructional strategies (say, individual learning vs. social learning) and having all 250 students randomly assigned to play either game A, or game B. In this manner, the two games would be compatible with each other and the differences in achievement could be safely attributed to the difference in instructional strategies.
B. **Repeated-Measure Studies:** Allow all 250 students to participate in the intervention, and repeatedly test them (say, monthly) throughout the intervention period in regular intervals. Because participants in a repeated-measure study became their own control, there is no need for an isolated control group. Moreover, this approach also eradicated the possible ethical/fairness concern as to why only some students were being exposed to a beneficial (or harmful) intervention.

C. **Improved Repeated-Measure Study:** By temporarily removing the intervention in the middle of the (5-month long) treatment period. The rationale for this removal was to verify if the achievement of the participants was truly attributed to the intervention. The expected outcome would be for the achievement to plateau or declined during the absence of intervention. If student achievement score rose despite the removal of the intervention, then the "learning" must come from another source.

D. **Designing the Game as Assessment:** One of the criticisms of the video game designers from the industry was that teachers made games boring by testing the students after they played the games. The students might also come to resent the game sessions as nothing more than disguised lessons. Hence, a fourth (but much better) approach would be to design the entire game as an assessment, in which the students' actions and behaviors would be used as indicators of their understanding in the topics or subjects being studies. (Further explanations are given in point *No. 6: Interactive Learning Instruction Design* in the section below.)

ALL GAMES ARE "NOT" EQUAL

Besides the pitfalls of media comparison studies, critics should avoid applying one label on all game studies as if all games are created equal. Instead of examining the instructional contents and contexts found in various genres of games, and how the design and presentation of instruction and information might affect learning outcomes; many pundits chose to intermix findings of research using pen-and-paper games, board games, text-based computer games—such as Multi-User Dungeons (MUD) and MUD Object Oriented (MOO) from the 1990s, and edutainment from before 2005—essentially treating all games as equal. Using a counter argument similar to the approach by the proponents of media comparison studies, they concluded that serious games were *just as ineffective* as all other forms of games for instruction (cf., Clark, 2007; Kirschner, Sweller, & Clark, 2006). Readers should recognize that this was no different to the error committed by the other group of researchers who claimed that instructions using technology were *just as effective* as traditional teaching when findings showed no statistically significant differences between the two.

First, many of these earlier studies were media comparison studies that should have been invalidated, and not be included in any meta-analysis study. Secondly, critics who interpreted no significant differences finding to mean *just as ineffective* were equally guilty as their counterparts who claimed the instructional media to be just as effective. Thirdly, intermixing findings from research studies using board games, video games, and MMOGs, are like comparing apples, lettuce, and beef. The exercise is meaningless because it is simply another form of media comparison.

Unless researchers can look beyond the quagmire of media comparison studies, they will be dragged into this meaningless debate. It is far more useful to research appropriate methodologies for measuring the assessment of learning, and just as important to understand what kind of learning is possible with serious games. It is only to be expected that critics who do not grow up with video games (Prensky, 2007) will continue to distrust the technology and arrive at a conclusion contrary to that of the digital-natives (e.g., Barab, Thomas, Dodge, Carteaux, & Tuzun, 2005; Parker, Becker, & Sawyer, 2008). Last but not least, game developers who are trying to make a quick profit with serious games need to realize any attempt to pass off poorly designed video games—albeit with high entertainment values but lacking the means to instruct and assess learning—as serious games are likely to hurt the industry during this critical growth period. Instead, serious game publishers should seek to work with selected experts who were interested in creating exemplary serious games that will take the field to the next level.

Developing Serious Games as Interactive Learning Instructions

In the near future, instructional technologists might assume the role of designers of interactive learning instructions (Squire, 2003) and be tasked with designing/developing prototypes of learning games/mods using newly developed game design toolkits. Because such mods would be just as playable as COTS games (but much smaller in scope), they can be used as research platforms for data collection, for the testing of new instructional paradigms, or for soliciting '*buy-in*' (Castillo & Novak, 2008) from clients, stakeholders, game publishers, and funding agencies for the support needed to fully produce the *interactive learning instruction*, or serious game.

Instead of delegating the design of serious games to game developers, instructional technologists need to learn from experts from other fields, such as the cognition and learning sciences, and the game design industry. They would also need new design processes and development models that could integrate interactive learning instructions and assessment of learning into one package. Collaboration with others might even produce new authoring tools for serious game making, as dedicated serious game authoring tools would be invaluable to instructional technologists to ease them over the initial steep learning curve. This idea is not new, as attested by Wikipedia's page on "video game making software." Instructional technologists with game modding/development experience should share their working models and the lessons learned readily with the community, to make up for the knowledge gap found in the literature at this moment.

Besides the abovementioned research and design issues that needed to be addressed, another problem faced by many instructional designers and technologists is the lack of a development model for serious games (as interactive learning instructions). Without a proper model to anchor the development work, designers must resort to trial and error and may end up with a well-designed game that is too expensive to build; or worse, run out of budget in mid-stream and end up with an unusable game. The following sections present a viable (and tested) development model based on a real serious game project (Loh & Byun, 2009). The below model was distilled from the original process and should prove useful as a launch pad for other developers of interactive learning instructions.

[***Note:*** The following 10-step development model was presented as a viable guide for instructional technologists. As it would be highly unlikely for instructional technologists to be called to produce a full-fledge game-for-profit, the following guide may be used for the development of smaller serious game and game mods, such as prototypes for product demonstrations, classroom instructions, and research studies.]

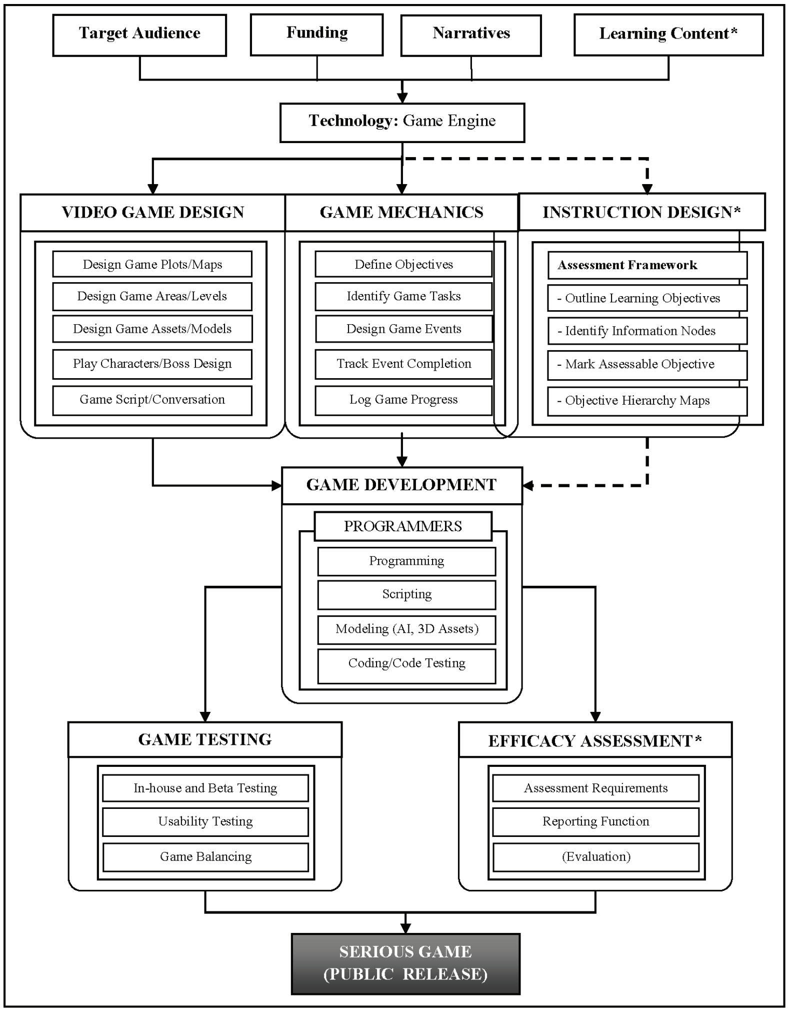
10-Steps Instructional Development Model

A total of 10 distinct steps are described in the game modding/development process (Figure 3). Serious game development components that are not found in video game development are distinguished using asterisk (*) marks. The game development cycle for the original project (i.e., *Saving Adryanee*) took less than four months for completion. Only the *Aurora Toolkit/Neverwinter Nights* bundle was used in the creation of the serious game mod. Subsequently, a screen capturing program named *Camtasia* was used (in conjunction with *Microsoft PowerPoint*) to create an endgame movie, which displayed a scrolling credit listing the names of the four-member development team and their affiliated higher institution.

1. Determining Target Audience and Learning Content *

The first step in any development project should rightly be a full analysis of the Learning Contents and the Target Audience. Instructional technologists understand this to be a very important step in the instructional design

Figure 3. A serious game development model



* *Essential Serious Game Components* (without which, this would be a Video Game Development process)

process because without a proper understanding of the audience's needs, the instructional media (learning contents) created may miss the mark completely. Special characteristics of the learner can also affect how the game or instruction is to be designed. For example, male players may prefer more *high-impact* actions, whereas female players tend to favor problem solving, emotional exchanges, and unexpected plot twists.

2. Determining the Amount of Funding and Time Available

The amount of funding and time allocated can also affect the outcome of the game project. Serious game designers must consider these two factors carefully, particularly when the project involves tax-payers' money or federal funding. Besides reporting the usual research findings, it will be well for researchers to report the amount of money and time spent, in order for the community to gain a better sense of the cost vs. benefit ratio of the project. Instructional technologists need to be more pragmatic about serious game development because they tend to be measured in months and years, not weeks. Sufficient time must be allocated (along with enough buffers) to ensure the project will proceed as planned. For example, *Immune Attack,* a full-fledged game took 3 years to complete. Since game modding projects are much smaller in scope, they may be completed within a much shorter time frame.

3. Writing Game Narratives

Finding a suitable story or writing an original game narrative can easily be the most difficult task in the game development process. When time is of the essence, it will be wise to settle on a storyline as quickly as possible. Since prototype games are not usually made to earn a profit, but are created for demonstration or for research, the narrative can easily take second place. A short and simple, but believable narrative is much more effective than an elaborated story. Some unique or memorable event should be planned within the first fifteen minutes to grab the player's attention and to draw them into the story. Game narratives have one other important purpose: they serve as acting scripts, listing all the props and characters needed to stage the game story. Should a designer be hit with a mental block, one can always hire professional game writers (Despain, 2008) to lend a helping hand.

4. Selecting the GDK/Game Bundle

The GDK/game bundle should ideally be selected only after the decisions of the target audience, learning content, amount of funding/time, and game narratives have been finalized. This will help ensure an unbiased choice for the best development platform. However, because it can take much time and efforts to learn a GDK and to become familiar with its use, some professional development houses have chosen to place Step 4 above Steps 1–3. This flip-flop in the sequence of decision may, at times, result in a phenomenon known within the game industry as *sequelitis*—games produced by one production house becoming less and less original (i.e., as if they are mere sequels of one game). On the contrary, some game publishers argue that sequelitis can, in fact, be a benefit. As the development team becomes more familiar with a particular GDK through frequent use, the production time will speed up and the time saved can be channeled into other projects. Moreover, there are many examples of game sequels that are just as successful as the original title, because they have remained true to the winning formula, particularly, evergreen series such as *Castlevania*, *The Legend of Zelda*, and *Final Fantasy*.

Instructional technologists need to bear in mind that the choice of the GDK/game bundle will frequently impact the look and feel, as well as the plot of a game. Because GDKs have been custom-made to produce a particular genre of game, such as First Person Shooter (FPS) or Role Playing Game (RPG), it will be difficult to use an FPS-oriented GDK to create an RPG, and vice versa. This explains why the U.S. Marines chose *Doom* (an FPS game) when they wanted to create a training game (another FPS) for the soldiers.

If budget is not a concern, project managers may choose from any of the following to maintain full control of the development environment: (1) licensing a commercial game engine, (2) farming out the game development project, or (3) commissioning the creation of a new proprietary GDK. These will naturally be very expensive approaches: Option 1 can cost several hundred thousand dollars per license; Option 2, a couple million dollars; and Option 3, several million dollars.

In a game modification project, instructional technologists must learn to work within the monetary and time constraints, and be contented with the finite resources provided by the GDK/game bundle. Hence, game modules may only cost the company a few hundred dollars (minus salaries of developers). Depending on the approaches chosen and the amount of a priori planning and design efforts put in, it is possible to create a serious game without the need for a big budget.

5. Video Game Design and Game Mechanics

Once all preparations (Steps 1–4) have been completed, it is time to create an early prototype to test out the look (design) and the feel (mechanics) of the proposed game. This tends to be a very long drawn-out process for those working in the COTS game industry. The process will usually begin with the production of concept arts by a graphic designer/artist to create the design for the proposed game world. It may include level (map) design, bosses, all the props needed for the game, and both player and non-player characters (PCs/NPCs). Once the artist's impressions have been approved, a modeler or texturer will then create all the models as conceptualized by the graphic designer. A mock-up game (ranging from pen-and-paper to computerized stick figures version) will be created to test if the proposed game mechanics will work smoothly; answering questions such as: How will a city wall that takes three game cycles to complete affect the game play? What will happen if the completion of city wall takes just two game cycles, or four?

Because the game mechanics can ultimately affect the feel of the game, it is important to take time to balance all the game objects (i.e., weapons against vulnerability) to prevent tipping the scale unfairly. This turns out to be a very easy step for game modification; since most of the game resources and game mechanics are already provided for by the GDK, there is nothing much to do, except for minor tweaking of object properties, such as textures and colors.

6. Interactive Learning Instruction Design*

[***Note:*** This is a unique step for serious game development.] Because an assessment component is the defining factor for serious games (Chen & Michael, 2005), instructional technologists must learn to identify "game-appropriate learning/training objectives" from game narratives. They will need help from the subject matter experts as to the metrics to be used as evidence of learning in relation to the player's performance within the game. The planning must be done a priori for proper assessment of learning, and will need to be integrated at this point (before Step 7).

If this step is omitted, the resulting serious game may still be useful for instruction, but only as an instructional tool. However, its value as a research tool will be diminished, because it will be very difficult to differentiate what a player really *learned* from the game. Did players learn the intended learning contents, or simply how to beat the game? No doubt many researchers will be tempted to employ multivariate testing methods (such as pretest/post-test) to ascertain the effects of serious game on learner performance. Although a pretest/post-test method can certainly measure the effects of *game playing* as a whole, it will be impossible to determine if the changes in performance are affected by the *game playing*, or the *learning contents* within the game. A better research design would be repeated measure of the effects of the learning contents using a series of pretest/post-test over a period of time (1 to 2 months) to rule out any signal interference (such as learning from another external source).

Moreover, the above method will not be able to measure for the effects of *individual learning tasks* within a game. Since a serious game can contain any number of learning objectives, it will be important for researchers to breakdown the game-based learning by tasks and by objectives. Hence, an *integrated assessment framework* (such as the one described in the next section) will be necessary to magnify the granularity of the research method, and allow for the measurement of the effects of *individual learning objectives* within serious games.

An Integrated Assessment Framework *

In order to facilitate data-driven assessment—what Mandinach and Honey (2008) referred to as "linking data to learning"—with serious games, some kind of data collection process must occur during a game play session in order to allow for a player's performance data to be collected. An integrated assessment framework will be invaluable to serious games used for research purposes. However, the literature has very little to say about an established assessment framework or software infrastructure that is targeted at data collection in serious games.

Loh and colleagues (Loh, 2006; Loh, Anantachai, Byun, & Lenox, 2007) have argued for the need of a software framework to facilitate automated data collection within virtual environments, and has subsequently presented a conceptual framework known as *Information Trails* (Loh, 2008). Conceptually, the Information Trails is a series of agent-detectable markings left by moving agents within an information ecology. Operationally, the assessment framework would facilitate virtual tracking of objects within information ecology, including that of serious games and multiuser virtual environments. Once a virtual object (such as player avatar) became traceable, the decision-making processes of its agent (the person behind the avatar)—reflected in the object's actions and behaviors—could then be used as evidence for assessment and analysis of the learners' performance.

The following steps are necessary for the creation of a serious game with Information Trails:

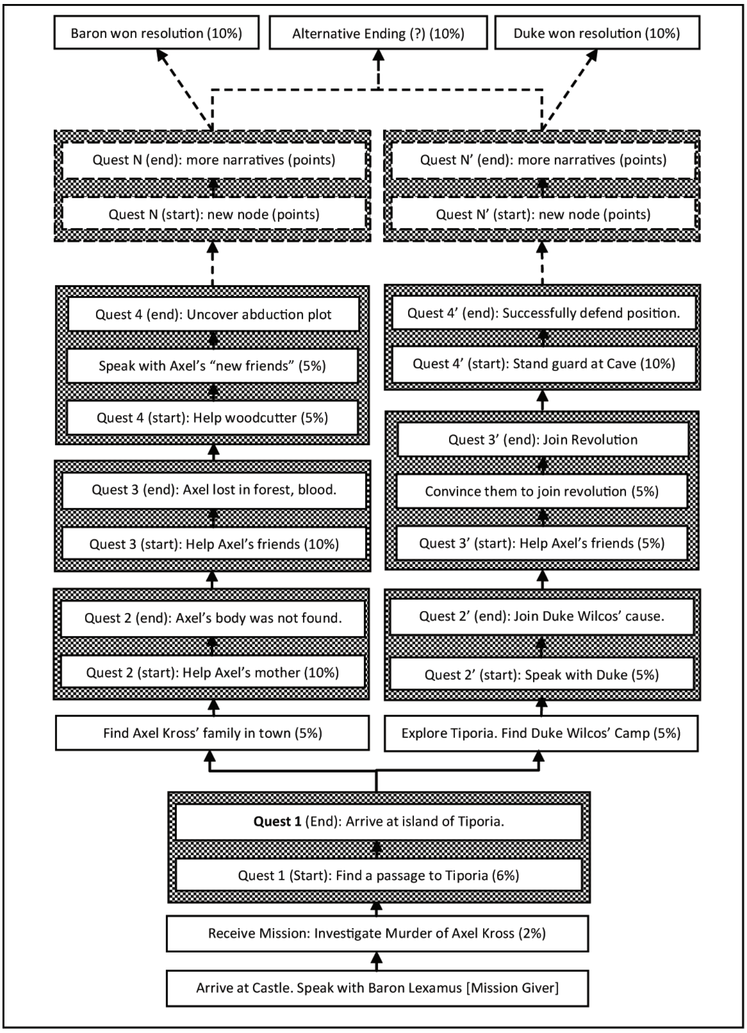
1. Starting with desire learning outcomes, create a game narrative that would incorporate as many learning objectives as possible.
2. Segregate game events from the narratives into player-dependent and player independent events—in this case, player-independent events would be plot-related events that will occur regardless of what players choose to do in the game.
3. List and match desired learning objectives to player-dependent events.
4. Breakdown *all* game events by hierarchy (into main and sub-objectives), and create an *Objective Hierarchy Map* (OHM) (Berg et al., 1999) for the game. (An example of an OHM is shown in Figure 4.)
5. Identify *Information Nodes* within the game narrative where:
    a. Player-independent events occurred (from Step 2), and
    b. Player-dependent events occurred (i.e., identified the Mission-start and -end points)
6. *Hook* event-tracers into place (to record actions and behaviors of players) at *all* Information Nodes, and send the data collected to a data store.
7. Access the data store and re-assemble events traced into Information Tails.
8. Perform assessment of learning analysis of players based on their performance and behaviors in the game, using appropriate data mining techniques and/or visualization tools.

Additional programming help may be needed to help create an appropriate Application Programming Interface (API) to facilitate cross-talk between external data stores and function calls within the game environment.

7. The Game Development Cycle

Should this be a COTS game development project, the GDK will be assembled at this point. Programmers will be working on software functions such as Artificial Intelligence (AI), collision detection and network balancing to be included in the GDK, while game developers will have first-hand use of the GDK to materialize the game world. The development process used to put together both the GDK and the game world will need to happen in tandem, a process known as *co-evolution* among electronic engineers.

Figure 4. A sample objective hierarchy map depicting the Information Trails of game events within a role-playing serious game



A Level Designer may assume the role of Team Manager and to ensure the game development cycle will move forward. All game assets will be placed in accordance with the preapproved game design document. Voice artists and musicians may be hired to record voiceover for NPCs and background music for the game, if necessary. Depending on how much time is still available, the game design cycle may be repeated up to a couple of times, or until both the design team and the game publisher can come to an agreement for sign-off. As the months roll by, the onus will be on the Level Designer to conduct game testing, draw the development cycle to a close, and adhere to the public release date.

Besides leading a serious game development team, instructional technologists must also be prepared to take on a serious game (prototype) all by themselves, especially during an economy downturn when additional personnel are hard to come by. Even though it may sound preposterous, many graduate students who need to create a game towards completing their Master or PhD degree have demonstrated that it is possible to create a game by oneself, when given enough motivation.

Putting all the pieces together in a game mod project is really quite enjoyable because it is the heart of the game modding process. Since the GDK already contains all (or most) of the necessary resources for making a game mod, instructional technologists need only to learn the toolkit and then piece together land mass, buildings, creature spawn points, background music, and conversations between PCs and NPCs, before they proceed to test-run the whole assembly.

8. Beta Testing and Usability Testing

Like many other software development processes, game development also includes a testing phase (about 4–6 weeks) to ensure production quality. During this period, groups of independent players may be recruited to test-play the pre-released game with the intention to find and eliminate software bugs before public release. Some major game publishers may undertake additional usability testing to ascertain if any segment of the game needs further tweaking or balancing. It is also common to conduct focus group interviews to solicit direct feedback from test-players.

Since testing can further delay the timeline for game release, there will come a point in time when the public release date must be adhered to. It is then up to the Level Designer to determine if last minute adjustment must be done before giving the final approval for the game to make its public debut. (This final step is akin to the final director cut in film making.) In the case of game modification, beta and usability testing are usually less of an issue because the game publishers would have already conducted these tests prior to the release if the GDK/ game bundle.

9. Public Release

In the game industry, the end of the game development project is usually marked by the pressing of the master/gold CD of the finished game. Once the gold CD is released for mass production, the development team will generally be dismissed or reassigned to work on other game projects. If severe problems are discovered after the point of sale, the problems will usually be taken care of via patch releases.

Educators may be surprised to learn that few game publishers are interested in taking steps to improve a game after point of sales. In an instructional development environment, a created artifact (instructional resource) is usually subjected to several rounds of evaluation and improvement, and may be re-used year after year to maximize investment. There is hardly any incentive for game publishers to revise a published game to make it run better.

From a game publisher's point of view, the game has already been evaluated (through beta and usability testing). Moreover, game buyers understand that the game is sold *as is*. Publishers who care about their reputations and their customers may offer to patch a broken game after release, but not much more. In very rare cases, a couple of the original programmers may be kept on a part-time payroll to provide after-sale support for a game. To date, Bioware, Inc. is the only game development house that has chosen to support *Neverwinter Nights* for an unprecedented 6 years. It is premature to speculate whether the serious games industry will choose to provide sustained support for their games, or go the route of commercial game publishers.

10. Efficacy Assessment *

Efficacy is a term used in the medical and pharmaceutical fields to measure if a particular medical intervention is able to produce a clinically measurable effect. *Efficacy assessment* is, therefore, an evaluation to judge if an intervention is "effective for the intended use" (Albrecht, 1997). Many educators believe that serious games have the potentials to turn the education process around by motivating students to learn as they play. However, the lack of efficacy assessment research in serious games has prompted critics to question its worth in the supposed education reform (Clark, 2007). Lacking clear empirical data, it will be impossible to calculate the cost-benefit ratio of serious games, meaning its effectiveness will always remain suspect.

The persistent use of media comparison studies by researchers to measure the effectiveness of serious games will only yield more no significant difference findings. Researchers and game developers must push for an *integrated assessment framework* that will allow for in situ data collection. Once individual learning objectives can be accurately measured, researchers must go on to improved the research methodologies to better analyze the data collected, and go on to model learner behavior and measure the efficacy of serious games.

CONCLUSION

During his keynote speech during the 2008 Annual Conference for the Association for Educational and Communication Technology (AECT), James Gee gave several examples to support his claim that *modding* is fast becoming the method of thinking and learning for the new generation of students. Furthermore, modding has also been proven useful for building teamwork (Antti, Tuula, & Marja, 2007) and collaboration (Hämäläinen, Manninen, Järvelä, & Häkkinen, 2006) among both young adults (e.g., Berger, 2006; Szafron et al., 2005) and school children (BBC, 2004; Wyeld, Leavy, Carroll, Gibbons, Ledwich, & Hills, 2007), in the learning of language and story-writing skills (Robertson & Good, 2005), and other social skills, such as logical thinking, communication, negotiation, public speaking (Loh & Byun, 2008), and even computer programming (Becker & Parker, 2005). Game making activities have also been successfully implemented to help at-risk children to read and write better (Peppler & Kafai, 2007), for after-school programs, student computer clubs, and by the public library to raise literacy (Gilbert, 2009). Although there are many reasons to use serious games, virtually nothing has been published about the actual cost involved in serious game development. This is an unfortunate oversight on the part of academia to focus their discussions on research findings only, when the cost of developing instructional materials (in this case, serious games) often become the sole determining factor for its adoption.

In summary, researchers need to apply appropriate learning theories when designing serious games, steer clear of media comparison studies, avoid intermixing older problematic game research findings with the new, and be on the constant look out for new methodologies that will yield conclusive empirical findings about the efficacy of serious games—possibly through new data mining and data visualization techniques. Sharing and learning from one another is not just a noble academic idea, but an essential 21st century skill to advance the increasingly global economy and society. To advance serious games as a viable instructional option in the arena of interactive learning, much work and collaboration need to occur. Game designers and developers (including game modders) need to share insights, experience, methodologies, metrics, practical lessons learned, and visionary perspectives to help craft the new paradigm for effective serious game-based learning. Researchers and educators need to achieve rigors and standards in both research and development—by avoiding media comparison research and improving instructional developing models, before moving on to assess and evaluate the efficacy of learning that results from these new instructional media.

REFERENCES

Albrecht, R. (1997, May 28). *General considerations for clinical studies... from protocol to results*. Retrieved Aug 20, 2008, from http://www.fda.gov/cder/present/genconra/sld009.htm

Amarelo, M. (2008, May 22). *Fixing the education digital disconnect one video game at a time–FAS launches immune attack*. Retrieved May 28, 2008, from http://www.fas.org/press/news/2008/may_ialaunch.html

Antti, K., Tuula, N., & Marja, K. (2007, September 24-28). *Team structure in the development of game-based learning Environments*. Paper presented at DiGRA 2007, Tokyo, Japan. Retrieved May 20, 2008, from http://www.digra.org/dl/db/07344.35576.pdf

Barab, S., Thomas, M., Dodge, T., Carteaux, R., & Tuzun, H. (2005). Making learning fun: Quest Atlantis, a game without guns. *Educational Technology Research and Development*, *53*(1), 86–107. doi:10.1007/BF02504859

BBC. (2004, August 11). Video games "good for children". *BBC News*. Retrieved April 30, 2008, from http://news.bbc.co.uk/go/pr/fr/-/1/hi/scotland/3553352.stm

Becker, K., & Parker, J. R. (2005, Oct 13-15). *All I ever needed to know about programming, I learned from re-writing classic arcade games*. Paper presented at the Future Play: The International Conference on the Future of Game Design and Technology, East Lansing, MI.

Berg, J., Bradshaw, B., Carbone, J., Chojnacky, C., Conroy, S., Cleaves, D., et al. (1999). *Decision protocol 2.0*. Washington, DC: USDA Forest Service.

Berger, A. (2006, January 31). Neverwinter nights in the classroom. *University of Minnesota News*. Retrieved April 30, 2008, from http://www1.umn.edu/umnnews/Feature_Stories/22Neverwinter_Nights22_in_the_classroom.html

Botturi, L., & Loh, C. S. (2009). Once upon a game: Rediscovering the roots of games in education. In C. Miller (Ed.), *Games: Their purpose and potential in education* (pp. 1-22). New York: Springer Publishing Company.

Bounds, J. (2007, April 6). Game on! *Dallas Business Journal*. Retrieved February 2, 2009, from http://dallas.bizjournals.com/dallas/stories/2007/04/09/focus1.html

Castillo, T., & Novak, J. (2008). *Game level design*. Clifton Park, NY: Delmar Cengage Learning.

Chen, S., & Michael, D. (2005). Proof of learning: Assessment in serious games. *Gamasutra*. Retrieved from http://www.gamasutra.com/features/20051019/chen_01.shtml

Clark, R. E. (2007, May-June). Point of view: Learning from serious games? *Educational Technology*, *47*, 56–59.

Despain, W. (Ed.). (2008). *Professional techniques for video game writing*. Wellesley, MA: A. K. Peters, Ltd.

Federation of American Scientists. (2006). *Summit on Educational Games 2006: Harnessing the power of video games for learning*. Washington, DC: Author.

Ferdig, R. E. (Ed.). (2008). *Handbook of research on effective electronic gaming in education*. Hershey, PA: Information Science Reference.

Gee, J. P. (2003). *What video games have to teach us about learning and literacy* (2nd ed.). New York: Palgrave Macmillan.

Gilbert, B. (2009, March 5). *American library association goes gaming*. Retrieved March 9, 2009, from http://www.joystiq.com/2009/03/05/american-library-association-goes-gaming

Haertel, G. D., & Means, B. (Eds.). (2003). *Evaluating education technology: Effective research designs for improving learning*. New York: Teachers College Press.

Hämäläinen, R., Manninen, T., Järvelä, S., & Häkkinen, P. (2006). Learning to collaborate: Designing collaboration in a 3-D game environment . *The Internet and Higher Education*, *9*(1), 47–61. doi:10.1016/j.iheduc.2005.12.004

Hastings, N. B., & Tracey, M. W. (2004). Does media affect learning: Where are we now? *TechTrends*, *49*(2), 28–30. doi:10.1007/BF02773968

Hewitt, D. (2008, June 23). *Use of video game technology in the workplace increasing*. Retrieved Oct 20, 2008, from http://www.theesa.com/newsroom/release_detail.asp?releaseID=24

Hopson, J. (2006). We're not listening: An open letter to academic game researchers. *Gamasutra*. Retrieved from http://www.gamasutra.com/features/20061110/hopson_01.shtml

Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational Psychologist*, *41*(2), 75–86. doi:10.1207/s15326985ep4102_1

Lenhart, A., Kahne, J., Middaugh, E., Macgill, A. R., Evans, C., & Vitak, J. (2008). *Teens, video games and civics: Teens' gaming experiences are diverse and include significant social interaction and civic engagement*. Washington, DC: Pew Internet & American Life Project.

Locke, B., Moore, M., & Burton, J. (2001). Old concerns with new distance education research. *EDUCAUSE Quarterly*, *24*(2), 60–62.

Lockee, B. B., Burton, J. K., & Cross, L. H. (1999). No comparison: Distance education finds a new use for 'No significant difference.' . *Educational Technology Research and Development*, *47*(3), 33–42. doi:10.1007/BF02299632

Loh, C. S. (2006). *Tracking an avatar: Designing data collection into online games.* Paper presented at the annual conference of the Association for Educational Communications and Technology (AECT 2006), Dallas, TX.

Loh, C. S. (2008). Designing online games assessment as "Information Trails." In V. Sugumaran (Ed.), *Intelligent information technologies: Concepts, methodologies, tools, and applications* (Vol. 1, pp. 553-574). Hershey, PA: Information Science Reference.

Loh, C. S., Anantachai, A., Byun, J., & Lenox, J. (2007, November 21-23). *Assessing what players learned in serious games: In situ data collection, information trails, and quantitative analysis.* Paper presented at the Computer Games: AI, Animation, Mobile, Educational & Serious Games (CGAMES 2007), Louisville, KY.

Loh, C. S., & Byun, J. H. (2009). Modding neverwinter nights into serious games. In D. Gibson & Y. K. Baek (Eds.), *Digital simulations for improving education: Learning through artificial teaching environments* (pp. 408-426). Hershey, PA: Information Science Reference.

Mandinach, E. B., & Honey, M. (2008). *Data-driven school improvement: Linking data and learning*. New York: Teachers College Press.

Michael, D., & Chen, S. (2006). *Serious games: Games that educate, train, and inform*. Boston: Thomson Course technology PTR.

Miller, C. T. (Ed.). (2008). *Games: Purpose and potential in education* (1st ed.). New York: Springer Science+Business Media, LLC.

Parker, J. R., Becker, K., & Sawyer, B. (2008, January/February). Re-reconsidering research on learning from media: Comments on Richard E. Clark's point of view column on serious games. *Educational Technology*, *48*, 39–43.

Peppler, K. A., & Kafai, Y. B. (2007, September 24-28). *What videogame making can teach us about literacy and learning: Alternative pathways into participatory culture*. Paper presented at DiGRA 2007, Tokyo, Japan. Retrieved May 20, 2008, from http://www.digra.org/dl/db/07311.33576.pdf

Prensky, M. (2005). *Engage me or enrage me: What today's learners demand*. Retrieved July 25, 2005, from http://www.marcprensky.com/writing/Prensky-Engage_Me_or_Enrage_Me.pdf

Prensky, M. (2007). *Digital game-based learning*. St. Paul, MN: Paragon House Publishers.

Robertson, J., & Good, J. (2005). Story creation in virtual game worlds. *Communications of the ACM*, *48*(1), 61–65. doi:10.1145/1039539.1039571

Sawyer, B. (2005). The state of serious games. *Gamasutra*. Retrieved from http://www.gamasutra.com/features/20051024/sawyer_01.shtml

Squire, K. (2003). Video games in education. *International Journal of Intelligent Games & Simulation*, *2*(1).

Steinkuehler, C., & Johnson, B. Z. (2009). Computational literacy in online games: The social life of mods. *International Journal of Gaming and Computer-Mediated Simulations*, *1*(1), 53–65.

Szafron, D., Carbonaro, M., Cutumisu, M., Gillis, S., McNaughton, M., Onuczko, C., et al. (2005). Writing interactive stories in the classroom. *Interactive Multimedia Electronic Journal of Computer-Enhanced Learning, 7*(1).

Thompson, A. D., Simonson, M. R., & Hargrave, C. P. (1992). *Educational technology: A review of the research*. Washington, DC: Association for Educational Communications and Technology.

van Eck, R. (2007). Building intelligent learning games. In D. Gibson, C. Aldrich, & M. Prensky (Eds.), *Games and simulation in online learning: Research and development frameworks*. Hershey, PA: Idea Group, Inc.

Wyeld, T. G., Leavy, B., Carroll, J., Gibbons, C., Ledwich, B., & Hills, J. (2007, September 24-28). *The ethics of indigenous storytelling: Using the torque game engine to support Australian Aboriginal cultural heritage*. Paper presented at DiGRA 2007, Tokyo, Japan. Retrieved April 30, 2007, from http://

Christian Sebastian Loh wears many hats in real life. He is a researcher, an educator, a gamer, and a father. Professionally, he is an assistant professor of instructional design technology and one of the coordinators of the Collaboratory for Interactive Learning Research (C.I.L.R.) at the Southern Illinois University Carbondale. He also serves as president (2008-2009) for the Multimedia Production Division (MPD) of the Association for Educational Communications and Technology (AECT), reviews articles for the IJGCMS *as associate editor, and judges professional serious game competitions. His publications include many journal articles and book chapters on and about serious games, assessment, and emerging technology. Currently, he is pioneering the research on information trails, a user behavior analysis framework for in situ assessment within multi-user virtual environments, including serious games and virtual worlds.*