



Digital Culture & Education (DCE)

Publication details, including instructions for authors
<http://www.digitalcultureandeducation.com/>

Deep assessment: an exploratory study
of game-based, multimodal learning in
Epidemic

Jennifer Jenson^a
Suzzane de Castell^b
Kurt Thumlert^a
Rachel Muehrer^a

^a*York University*

^b*Simon Fraser University*

Online Publication Date: 10th March 2016

To cite this Article: Jenson, J., de Castell, S., Thumlert, K. & Muehrer, R. (2016). Deep assessment: an exploratory study of game-based, multimodal learning in *Epidemic*. *Digital Culture & Education*, 8(1), 20-40.

URL: <http://www.digitalcultureandeducation.com/cms/wp-content/uploads/2016/03/jenson.pdf>

PLEASE SCROLL DOWN FOR ARTICLE

DEEP ASSESSMENT: AN EXPLORATORY STUDY OF GAME-BASED, MULTIMODAL LEARNING IN *EPIDEMIC*

Jennifer Jenson, Suzanne de Castell, Kurt Thumlert, Rachel Muehrer

Abstract: *In this study, we examine what and how intermediate age students learned from playing in a health-focused game-based digital learning environment, Epidemic. Epidemic is a playful interactive environment designed to deliver factual knowledge, invite critical understanding, and encourage effective self-care practices in dealing with viral contagious diseases, using a social networking interface to integrate both serious games and game-like multimodal design projects. Epidemic invites a playful approach to its deadly serious core concern - communicable disease - in order to see what happens when students are encouraged to critically approach information from multiple or contradictory perspectives. To identify what participants learned while interacting within Epidemic, we deployed two instructional and assessment models, noting the differences each instructional approach could potentially make, and what approach to assessment might help us evaluate game-based learning. We found that each approach provided importantly different perspectives on what and how students learned, and on the very meaning of student success. Recognizing that traditional assessment tools based in print-cultural literacy may prove increasingly ill-suited for assessing emergent multimodal literacies in game-based learning environments, this study seeks to contribute to a growing body of work on the development of novel assessments for learning.*

Keywords: educational assessment, educational media, interactive learning environments, game-based learning, multimodal literacies, serious play.

Introduction

In an era when proponents of 21st Century Learning are promoting immersive, multimodal, digitally-mediated learning environments supportive of 'deep learning' (Dede, 2014), near-universal mandates promulgating standardized assessment models continue to work, antagonistically, to undercut the very potential of these novel educational models. If assessment systems largely prefigure what significant learning looks like, what is measurable, and therefore what is pedagogically possible, then transformations in learning environments must arrive with equally innovative assessment tools. While Dede (2014) signals an urgent need to transform traditional assessment tools and systems, drawing our attention to the eventuality of 'real-time diagnostic assessments...woven into immersive simulation[s]' and learning sites (p. 19), his work does not explicitly address how we might, more presently, develop practical instruments that enable us to rethink and reassess the outcomes – and the creative expressions – of student learning that emerge through deep engagement in dynamic game-based, socially-networked learning environments.

To address the immediate, sociotechnical concern of designing assessment instruments adequate to transformed – and educationally transformative – learning environments, the study reported in this article examines what and how students learned from playing in a

health-focused digital learning platform, *Epidemic*, and how their learning was expressed - and assessed - using novel tools and instruments. *Epidemic* is an interactive game-based platform designed to deliver contextualized knowledge, invite critical reflection on that knowledge, and encourage effective self-care practices in dealing with contagious diseases – a particularly timely intervention given the recent global Ebola and Zika pandemics. Using a social networking interface to bring together serious games and multimodal production projects (through which students creatively demonstrate understanding), we identify what and how participants learned while interacting within, and creating knowledge through, the *Epidemic* platform. For this study, we deployed two instructional and assessment models (standard and experimental), coding and analyzing the differences each instructional and assessment approach made and, further, evaluating which approaches might help us better understand and assess significant learning in multimodal and game-based environments.

The stakes of this particular research study transcend matters of ‘educational enhancement’ to make visible novel pedagogies, new modes of student engagement and creative action, and new assessment forms that might challenge the global drive toward the educational shallows of standardization and accountability. In our conclusion, we contend that it is only through transformed forms of assessment that formal education might connect with the changing worlds of knowledge, creative practice and critical agency outside of schools.

Background to study

Claims about the educational value of digital gameplay and immersive, playful virtual worlds are by now widely rehearsed, with proponents arguing for games as designed learning environments that can offer their players experiences different from, contextually richer than, and more engaging than those available in traditional schooling models (de Castell & Jenson, 2003; Gee, 2003; Prensky, 2006; Rieber, 1996; Squire, 2011, Wouters, van der Spek & van Oostendorp, 2009; Apperley & Beavis, 2011). However, hard proof of the learning potentials of games has been elusive (Linderoth, 2012; Mayer, 2014), with some education and gameplay studies showing no significant gains using standard testing measures (Sward, Richardson, Kendrick, & Maloney, 2008; Young et al., 2012), and others showing significant gains (Barab et al., 2008; Ke, 2008; Sitzmann, 2011; Wouters, et al., 2013) or indicating that informal learning is occurring, incidentally, by virtue of players engaging diverse challenges through games, situated play, and multimodal production activities (Steinkuehler, 2006; Salen, 2007; Alexander, Eaton & Egan, 2010). Beyond the classical “learning outcomes” question in its direct form, the position most often argued is that the real potential in games lies in their capacity to attract, capture, engage, and sustain student attention (Boyle et al., 2012; Connolly et al., 2012), in order to, indirectly, support other, more conventional, educational tasks and activities.

Research on game-based health education, specifically, has largely focused on: 1) increasing awareness of health-related issues, like food choice, obesity and exercise (Papastergiou, 2009; Partridge et al, 2007); 2) using games to promote changes in health-related behaviours, like appropriate post-operative care (Arnab et al., 2013; Baranowski et al., 2008; Thompson et al., 2010); and 3) developing games that can develop and/or support specialized populations struggling with particular diseases like diabetes or cancer (Beale et al., 2007; DeShazo, Harris, & Pratt, 2010; Knight et al., 2010; Kato et al., 2008). Generally speaking, it is probably fair to say that health games, and health education more broadly, have focused primarily on compliance and, secondarily, on factual knowledge, and research on health games, accordingly, reports outcomes, impact and effectiveness in these terms.

Studies by Lieberman (2001, 2012) in the area of diabetes and self-management, for example, asked diabetic youth to play a video game related to the disease and its management for six months. He found that those who played the game knew more about their disease and its management than those who did not. More recently, DeShazo et al. (2010) reviewed video games used in diabetes education and concluded that “video games hold great potential as an alternative modality for diabetes education” (p. 819). In addition, a growing number of studies in the area of exer-gaming (fitness games) has investigated whether and how fitness games successfully support physical activities for a wide range of different users (Papastergiou, 2009), although those studies tend to be much more focused on altering player behavior than on developing player knowledge and understanding.

While acknowledging that studying ways to alter behavior is very much an important (and arguably fundamental) dimension of game-based health education (Beale et al., 2007; Unnithan et al., 2006), behavioral change was not the primary intent of our own study. Rather, we sought to discover whether and how a designed playful environment might effectively support participants’ development of critical knowledge and understanding about contagious disease processes.

Few studies of games for health have examined how playing health games might help develop the kinds of knowledge and understanding that can cultivate a more critical, self-reflective and less compliant relationship to personal, community, and global health challenges, conditions and crises. In one small scale study, more exploratory than definitive, of learning about contagious disease and its management, Lennon and Coombs (2006) report on a single case of a child (aged 8) creating a dengue fever related board game, and they detail the kinds of learning demonstrated, including a “diagnostic of a child’s understanding of a topical knowledge (in this case dengue)” (p. 96). Lennon (2010) also studied a single player of immunity-based games and malaria games (Lennon 2006) designed as part of the Nobel Prize suite of web-based games. In both cases, the studies emphasized the player’s feedback about the games and the debriefing strategies used post-game, as well as providing some accounting of the content or topical knowledge players demonstrated.

Another important, and again small scale, study (Amory, 2010) of twelve Soweto, South African teenagers playing a health-related game shifts the focus of study from a game as a “stand in” instructor to the game as one tool (among others) to support learning. Amory argues that within the context of learning, games should be viewed not as “instructional media”, but instead as a “tool to mediate learning” (p. 825). The study demonstrates how these young players were better able to understand key concepts related to disease, including HIV/AIDS, than were a group of first year university students studying biology.

Illustrating the challenge of identifying learning gains from educational games is a large-scale study of users of a Multi-User Virtual Environment (MUVE), *River City*, that was designed to support science learning through problem-based inquiry. The project has been documented in many different forms over a number of years (see for example: Dede, 2009; Ketelhut, 2007; Nelson & Ketelhut, 2008), and in general demonstrated that “a broader range of students gain substantial knowledge and skills in scientific inquiry through immersive simulation than through conventional instruction” (Dede, 2009). In one large scale study of 2000 middle school students using *River City*, Ketelhut, Dede & Clarke (2005) used both standard measures (pre- and post-test) as well as a more inquiry-driven measure in which students wrote a “Letter to the Mayor” of River City. Analysis of student letters revealed that the pre- and post-test did not necessarily adequately capture what students learned. They write: “... students who scored low on the science inquiry post-test wrote letters that were of similar quality to those written by students who scored higher on the post-test” (p. 8). In short, after applying non-standard evaluation and

assessment models, the study demonstrated more gains in students' understanding than the standardized test results had indicated (Ketelhut, 2007; Nelson & Ketelhut, 2008).

These findings resonate with the arguments of Dede (2014), Merchant (2010), Curwood (2012), and Bezemer and Kress (2008), who suggest that conventional, text-based assessment tools used to measure student learning of well-specified curricular knowledge are unable to measure, or even countenance, the forms of learning and the meaning-making performances being enabled and enacted within virtual worlds, digital games, and digitally-mediated multimodal learning environments, which draw upon equally important, if less familiar, semiotic resources and affordances. Bezemer and Kress (2008) conclude their discussion of emerging multimodal learning environments by signaling the “pressing issue” of developing “apt forms of assessment for representations in different modes, treated as signs of learning” (p. 193).

Keeping in mind these and similar arguments (Young et al., 2012; Ketelhut, 2007; Klopfer, 2011) about the limitations of standardized assessment models for measuring learning through games and digitally-mediated multimodal environments, we wanted to examine whether, what and how students (ages 11 to 14) learned about infectious diseases and their transmission after having played in a ludic online learning environment, *Epidemic*, and what kind of assessment was best able to make that learning evident.

Materials

Epidemic: A playful learning environment

Intended to teach adolescents (11-14) about contagious disease infection, self-care and prevention, *Epidemic* is a modular, Flash-based online environment that allows players to access text-based material on over 30 contagious diseases (“Virus Profiles”), create their own or edit other users’ disease-related “public health” posters and illustrated comics, and create and customize a fictitious disease avatar that gets stronger (more viral) as users complete more activities in the environment, playing – literally – with disease related images and information.

The initial impetus for developing *Epidemic: Self-care for Crisis* was the SARS outbreak in 2003, and the recognition that elementary and secondary school curricula in Ontario and elsewhere did not explicitly address issues around prevention and self-care in the face of new contagious disease strains, some of which have in recent years threatened to become global pandemics (SARS, avian influenza, H1N1 and, most recently, Ebola and Zika). As private and community health becomes, increasingly, not just a local public, but also global matter over which individuals have limited control, our focus on *Epidemic* began with helping users learn how to protect themselves and their communities from particularly prevalent viruses, from HIV to chicken pox to common influenza strains, and to enlarge the context for players’ understanding to include other historical and/or rare diseases such as Ebola, polio, and the hantavirus. For each of the 30 viruses featured in the environment, we include not only practical information for self-care (identifying and treating symptoms, managing contagion and prevention), but also scientific facts and discourse practices (epidemiology and virus morphology) and ethical and social-scientific understandings (i.e., relating to social, economic and material conditions, and/or ongoing misconceptions, as conveyed through public media channels).

The development of *Epidemic* is more fully described elsewhere (Authors redacted for review), so we touch here just on the aspects of its interactive environment most relevant to this study. The overall interface and functionality of *Epidemic’s* user home page is a Facebook-style social networking tool that allows users to “friend” each other and view one another’s disease-related stories and posters (see Figure 1). Unlike Facebook, however, users create alternate identities¹ – specifically, custom-designed viruses that become players’ avatars for the site. The avatar creation tool (see Figure 2) allows users to

select their body, each physical structure representing a different family of viruses; this decision as to what kind of virus the player chooses subsequently informs what kinds of symptoms, transmission vectors, and weaknesses users can attribute to their virus – and herein resides much of the curricular content, as all of these avatar design options are based on epidemiological fact. Thus, in developing their own custom avatar for use in the *Epidemic* environment, users must apply and extend authentic epidemiological knowledge.

A significant and playful part of this interface is its procedural game dynamic: as users friend one another, post new content (e.g., stories made with the “FluTube” story and poster generator) or post high scores from a mini-game (that has players avoiding certain airborne and blood-borne viruses), users receive immediate visual feedback that shows an increase in the potency of their virus (avatar). As a form of a game-play, the more potent the virus, the more potent a user’s social network (much like Facebook), where the communicability of a virus is itself reflexively modeled by, and enacted through, the “going viral” of social media communication—a contagious digital-era trope originally contracted from the field epidemiology.

EPIDEMIC *spread. multiply. dominate.*

CONTAGION METER

STATUS
Outbreak

FACT: Those exposed to Hep A develop a life-long immunity upon recovery.

CHARACTER PROFILE | [Update](#)

FRIENDS | [Add Friends](#) | [View Invites](#) | [Accept Invites](#)

Glitchovirus

Symptoms: genital warts, can lead to cervical cancer

Transmission: unprotected oral, anal or vaginal sex with infected individual

Treatment: warts can be treated/removed

GAMES

PROPAGANDA
SPREAD LIES & MISINFORMATION.

HALL OF INFAMY

Rabies
The Rabies virus can infect humans and animals, and is almost always fatal unless treated.

Hand, Foot & Mouth
The Coxsackie Virus (Hand, Foot and Mouth Disease) was first isolated in fecal matter from polio patients, and was

VIRAL NETWORK

Select an existing chatroom

-or-

Create a new one

RSS FEEDS

True: women's weight tied to Healthy Aging
Category: Health News
Created: 9/30/2009 9:01:00 AM
Last Editorial Review: 9/30/2009 9:01:54 AM

[Few Eating Enough Fruits, Veggies](#)
Title: Few Eating Enough Fruits, Veggies
Category: Health News
Created: 9/30/2009 8:55:00 AM

LINKS

- [CDC \(Center for Disease Control and Prevention\)](#)
- [Center for the Advancement of Health](#)
- [Familydoctor.org](#)
- [Family Health International](#)
- [HealthInSite](#)
- [Infectious Disease Society of America](#)
- [Institute of Medicine](#)

Figure 1: Epidemic home page

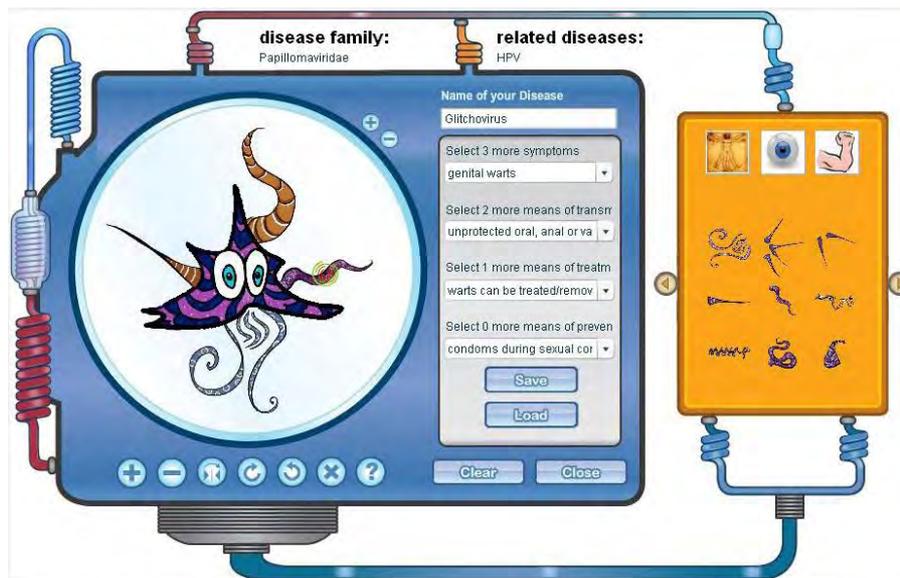


Figure 2: The avatar creation tool

The second aspect of *Epidemic* that participants in this study encountered was FluTube, the story-creation tool that drives *Epidemic*'s central multimodal activity (see Figure 3). Like dynamic adventure, non-linear narrative, and role-playing games, FluTube allows for ludic, hands-on experimentation with—and manipulation of—objects, characters, and narrative scenarios. This multimodal story-building tool is intentionally playful, even as it engages users with imagery and topical information related to the very serious matter of contagious diseases. When users compose and complete a multimodal document or *Epidemic*-related story, it is posted to their “wall”, and to the “walls” of their friends; posting new content increases viral potency and, thereby, social status within the game.



Figure 3: FluTube

The third module of *Epidemic* that students in this exploratory study interacted with was the “Propaganda Maker” (see Figure 4). Its function was to enable users to create informative (to protect from an enemy disease) and dis-informative (to trick people in ways that would make them more vulnerable to one’s own disease-avatar) posters that might describe disease side effects, transmission vectors, and methods of treatment, as a

way to actively engage students in critically understanding how health information can be accurately—or misleadingly—represented. The “Propaganda Maker” is essentially a digital remix tool (Knobel & Lankshear, 2008) that enables students to create propaganda posters by juxtaposing images and original text to informative and/or critical/parodic ends (see Figure 4). After composing a poster, students can directly save and post their document to their *Epidemic* home page, where it becomes visible to others in the network, inviting communicable interaction (“Likes”) which will in turn boost the poster designer’s viral potency.

The tool is designed to position students as producers and participants in the process of public health communication rather than simply asking them to reiterate and comply with purportedly value-neutral facts about diseases. This opportunity to rehearse, play with, and critically remediate PSA-style health communication was particularly relevant given that both schools we worked at featured public health posters prominently in hallways, bathrooms, and other high-traffic areas. As with the FluTube story-building module, completing a multimodal poster and updating it on a user’s wall increased viral potency and thereby social status. By “friending” other users, they could also view the propaganda posters other classmates had published to their walls.



Figure 4: Propaganda Maker

Two features of the game were not used in the study, though they were available to the participants. The first was a mini-game called *Host-Hop*, a Frogger-like game that challenges the player to jump on droplets of fluid travelling between two hosts. If the virus can stay on the droplets long enough to cross the screen, it can reach the next host. The second feature was an in-game chat window where players could message their “friends” who were currently online.

Methods

Our exploratory mixed-methods study had four main components: 1) We used observations, audio-visual recordings of the play sessions, and daily fieldnotes to

document incidental learning that occurred as participants interacted with *Epidemic*, its activities, the researchers, and the contagious-disease related content; 2) We used a set of pre- and post-test content-questions to see what fact-driven *standardized* assessment measures could capture of students' learning from interacting with *Epidemic*; 3) We used a *non-standard* assessment framework (Beavis, 2004) designed to evaluate multimodal student work – in our case, posters and serious comics (in both traditional paper-and-pencil and digital forms). This multimodal assessment tool (described below) was designed to look for different kinds of evidence of *what* and *how* students learned; and 4) we used a questionnaire about participant's demographic information (age, grade, sex, ethnicity), as well as media and videogame habits to contextualize this exploratory study's participants and its preliminary findings.

Assisted by the principal and school staff at 2 comparable, mid-high SES large suburban schools in Ontario, we recruited a group of teachers to run the study over one week during their regularly scheduled class times, and students in these classes were invited to participate. In total, 178 students aged 11 to 14 participated in the study.

In total, across the two sites, both grade 6 and grade 8 students were assigned to a standard group, who were directly instructed in a pedagogically traditional class: a didactic, lecture-based knowledge-presentation of the same content knowledge that was situationally and multimodally embedded in *Epidemic*'s play-based environment (n = 66). Two classes, one grade 6 and one grade 8, were assigned to an experimental group who played *Epidemic* (n = 89), and 1 of the classes was assigned to the baseline group (n = 23) who took the pre- and post-tests without engaging in any of the health related activities until they had completed the post-test. All groups spent five 40-minute sessions with the researchers. In the first session, all were given the demographic questionnaire and a multiple choice pre-test to determine students' prior content knowledge.ⁱⁱ The test questions were composed of images and text drawn directly from the "virus fact sheets" in *Epidemic*.

Standard Group: After completing the questionnaire and the pre-test in the first session, the participants in the standard group were given a conventional lecture (with PowerPoint) on five contagious diseases (HIV/AIDS, SARS, West Nile, smallpox, and influenza). The slide deck was developed and presented by the researchers, and composed of information and images drawn directly from the *Epidemic* website to ensure that the students in the standard group were presented with the same information that the students in the experimental group would encounter through play within *Epidemic*. In the second and third sessions the standard group was provided with the 5 virus fact sheets printed from *Epidemic*, and asked to create either two public service announcements in the form of posters (one accurate and one dis-informational) or a comic strip, using card stock, colored pencils, markers, glitter-glue, and printouts of art assets from *Epidemic*. At the beginning of the fourth session, the standard group was given a few minutes to finish that task, and then took the post-test. The remainder of the week, they played *Epidemic*. Ensuring all participants had an opportunity to play *Epidemic* was important because it was this opportunity that had motivated participants' involvement in the study.

Baseline Group: After completing the questionnaire and pre-test, the baseline group (23) did not have instruction and did not play *Epidemic*, but was asked to play, instead, their favorite online game for that day. They took the post-test (which was identical to the pre-test) at the end of that play session. In the second and third sessions this group, too, was asked to produce either two public service announcements in the form of posters (one accurate and one dis-informational) or a comic using the same (non digital) materials as the standard group. At the beginning of the fourth session, they were also asked to finish

their work and then invited to play *Epidemic* for the remainder of the week (the end of the 4th session and all of the 5th session).

Experimental Group: After completing their questionnaire and pre-test, participants in the experimental group were invited to play *Epidemic*. Beginning by creating and naming a virus to act as their public avatar for *Epidemic*, they explored the environment (see Figure 5). In subsequent sessions (2, 3 and 4), this group engaged and applied topical knowledge using the digital design tools in FluTube and Propaganda Maker to develop their own creative content. Like the standard group, they were asked to produce either two public service announcements in the form of posters (one accurate and one dis-informational) or a serious comic, but in this case they were explicitly requested to use *Epidemic*'s multimodal digital-design tools. As noted above, once *Epidemic* participants design an artifact (a serious comic or poster), they upload the image to the their user page (as you would on any social media site) where the other students/players may respond to or “like” the image. By *serious comic*, we refer to multimodal artifacts that take the aesthetic and communicative possibilities of graphic texts and comic genres seriously in conveying serious ideas through sophisticated digital narratives.

After participants in the experimental group completed and posted either posters or comic strips (and sometimes both), they moved on to use other features of the environment, spending time especially on its social networking capabilities and games, as well as engaging with the posters/comics created by other participants. The chat function in the social networking aspect of *Epidemic* was what these students appeared predominantly interested in, with most of their remaining time (having completed the activities) spent “friending” each other and having informal chats. At the beginning of the fifth session, they took the post-test and were then free to further explore the *Epidemic* website.



Figure 5: Two examples of avatars created in the experimental group.

Data/Analysis

All of the data from the questionnaire, the pre/post tests, along with fieldnotes and video-data, as well as all data related to the creation of digital posters and serious comics, was compiled and stored on a server, and we took pictures of all non-digital artifacts that students created. We first explain more fully the way we analyzed students' multimodal work using an innovative assessment instrument, and then report on the analysis of quantitative data.

To evaluate the posters and comics, we used Green's 3D analytical model (1988), which has been adapted and further refined for multimodal and digital contexts (Durrant & Green, 2000; Beavis, 2004) as a framework for multiliteracy and technology curriculum assessment. “Three dimensions” frame Green's model for multimodal literacy assessment:

operational, cultural and critical. We use the framework here in keeping with Beavis' (2004) adaption in which she uses the categories – operational (technical competence and “how to” applications), cultural (demonstrating general understanding of content knowledge and the broader contexts of the activity), and critical (linking content and context while demonstrating some self-reflective or critical perspective on what was created or enacted)—to analyze multimodal artifacts that students produced.

For the present project, we codedⁱⁱⁱⁱ multimodal posters and serious comics, which required a translation of terms to apply across both textual and multimedia literacies. Applying Green's (1988) model to the multimodal work students in all three groups produced, both digital and traditional (i.e. conventional paper-and-pencil), we coded student work as: operational if the work demonstrated a literally correct and technically competent grasp of the assignment, reproducing factual disease related information in a technically capable way. We coded work as cultural if it demonstrated an understanding of the content in context, evidenced an ability to apply to the task information or understandings drawn from students' wider social/cultural context (e.g. from posters in the school, or derived from other outside-the-project sources, other media models or examples), and any grasp of the wider implications of the information they used beyond its original presentational context. Work was coded as critical if it showed analytical or deconstructive engagement with the ideas and information given, and evidence that students grasped wider significance (i.e. beyond its initial presentational context) and rhetorical purposes of their health-promotion project. And sometimes, appropriately enough for this play-based activity, students demonstrated that critical perspective through intertextual references, ironic humor and/or critical parody of health education messages and public service announcements.

Standard measures: Pre-test to post-test score comparisons.

Data from the baseline group of grade 6 students, who took the pre-test and the post-test before engaging in any of the activities related to the study, was used to detect a practice effect: how much improvement could we expect to occur simply because participants had taken the same quiz after a short period of time? We needed to know that improvements we might see in experimental and standard groups were a function of specific instructional interventions, and not artifacts of repeated test-taking.

A paired-samples T-test revealed no significant change from pre- to post-test ($p = .853$) in the baseline group. No one improved by more than 1 point between pre- and post-test and 42.1% of the group got the same score on the pre and post-tests (see Table below). The average score of the baseline students on the pretest was quite high (10.63 correct items out of 14, or 75.9%), and, far from improving, the baseline group as a whole did marginally *less* well on the post-test, so we were confident the tests were not themselves a contributor to improved scores.

Table 1. Pre- to post- test score change for baseline group.

Group Status, Baseline group	# of Students	Percent change	
Valid	-3.00	2	10.5
	-1.00	2	10.5
	.00	8	42.1
	1.00	7	36.8
	Total	19	100.0
Missing	4		
Total	23		

We compared the relative improvement between our standard and experimental groups by running a General Linear Model analysis (Mixed-ANOVA) to include both the repeated measure variable of time and the between-measure variable of group status.

There was no significant difference ($p=.072$) between pre-test scores for the two groups (8.877 vs. 9.549) groups, when the scores of the experimental groups and the standard groups in both schools were combined.

However, when the schools were looked at separately, at School 2, the standard group scored significantly higher on their pre-test than did the experimental group ($t(30)=-2.860$, $p = .008$) – in other words, the standard group students at school 2 *started off* with a major and statistically significant advantage.

Table 2: School 2 Standard group's initial higher scores than experimental group

Group Statistics				
	Group Status	N	Mean	Std. Deviation
Pretest total answers correct	Experimental Group	18	8.2778	2.16403
	Standard Group	14	10.2857	1.68379
Posttest total answers correct	Experimental Group	18	9.1111	1.96705
	Standard Group	14	10.4286	2.20887

It is of interest to note that for participants in school 2, the *learning gains* were greater for the experimental group than for the standard group, even though the standard group ended up with higher post-test scores than the experimental group. This indicates that, given the greater post-test gains of the experimental group, the game worked better as a learning tool for students in school 2 than the standard/traditional lecture-based pedagogy did. In any such small study (and classroom-based studies are mostly of this kind), we risk overlooking suggestive outcome differences if we dismiss all findings that are not statistically significant – a finer-grained analysis is often required to make educationally sound inferences from data of this kind. Schools and classrooms and student learning are, after all, very diverse and individual matters, and can be as much obscured as illuminated by basing educational decisions on large participant samples averaged across several potentially very different schools, teachers, and learners.

From a statistical standpoint, for example, experimental and standard groups, when both schools were taken together, did *not* show significantly differing average scores ($t(90)=-.445$, $p = .658$), and while both groups made positive gains in post-test scores, the *Epidemic* group showed a smaller increase than the Baseline group (0.79 points versus 1.549 points respectively) using the standard assessment tools (conventional tests).

Because the average post-test scores were significantly higher for the standard group when compared to the Experimental group ($t(90)=-3.194$, $p = .002$), it would be easy to conclude that the traditional pedagogy was more effective than the game-based approach, and thereby miss what a finer-grained analysis suggests: that as a learning tool, the game might actually help students learn better than traditional methods, *even if traditional methods produce better results on standardized tests*. In fact, when we look back at the measured (but non statistically significant) differences between experimental and standard group starting points (as indicated by the pretest), we see that, *across the board*, the standard group started off with higher scores in School 1 as well – it is just that this difference between the pre-test scores of both experimental and baseline groups at School 1 did not reach statistical significance ($t(90)=-.445$, $p = .658$). Interesting, too, at this school (1), standard group students made the largest gains over all groups at both schools. To repeat, in school-based research, it is often very useful to interrogate high-level and larger N-based findings at a more granular level (as it is to triangulate data types, something we have not done here) in order to better discern the differences schools and teachers might make to learning outcomes, even where (the same) pedagogical tools and activities are controlled for.

Table 3: Pre-and Post-test results by group, School 1.

Group Statistics				
	Group Status	N	Mean	Std. Deviation
Pretest total answers correct	Experimental Group	55	9.0727	2.18458
	Standard Group	37	9.2703	1.93862
Posttest total answers correct	Experimental Group	55	9.8545	2.41460
	Standard Group	37	11.3514	1.84415

Table 4: Both groups improved their overall average scores on the test; but the standard group improved more than the experimental group.

Group Status		Mean	N	Std. Deviation
Experimental Group	Pre-test total answers correct	8.88	73	2.19
	Post-test total answers correct	9.67	73	2.32
Standard Group	Pre-test total answers correct	9.55	51	1.91
	Post-test total answers correct	11.10	51	1.96

To see whether test outcome differences had to do with content knowledge or with representational medium, we grouped pre- and post-test questions into “text” and “image” categories, and then regrouped the questions in terms of the content-area knowledge they were testing. We had presumed we might see improvement in the standard group on the text-only questions because they had encountered the information in textual form through the PowerPoint lecture and then demonstrated their knowledge of contagious disease through paper and pencil activities. Similarly, we expected that the experimental group would improve their scores on the image-based questions, as they were interacting with content through an image-rich visual medium.

The standard group did significantly improve their scores on text-based questions: their average pre-test text-based score was 3.97 out of 7 in comparison to their average post-test score of 4.79 out of 7, but neither the experimental group nor the standard group showed a significant improvement on the image-based questions.

We expected that students who reported greater experience with technology might find it easier to play with and within *Epidemic* and might therefore gain more health-related knowledge than their peers, so we looked for relationships between gameplay experience and test score improvement, as well as between cell phone use and test score improvement. Participants in the experimental group who reported that they played games every day improved their scores more than non-gamers, however this difference was statistically non-significant, because of the small number of self-identified gamers. Nevertheless it is worth mentioning that there was a one-point increase between gamers and non-gamers in both the pre- (9.94 vs. 8.88) and post-test scores (10.67 vs. 9.67), an indication, even if not statistically significant, that digital game-based learning may be more effective with students who have technological (digital game, in this case) experience. Students in the experimental group who had cell phones also had higher scores on both pre and post-tests than students who did not, with their average improvement between pre- and post-test scores being 0.72, a statistically significant difference (because cellphone users were a larger group than gamers, whose similar improvement did not, however, reach statistical significance). It is also worth mentioning, though, that both gamer experience and cellphone use are correlated with higher SES, and

higher SES is invariably correlated with greater school success, so we have no way to disambiguate technology use from higher SES in this study.

Non-standard measures: Posters and serious comics

The projects created by the standard group were almost all in the form of posters for PSA's (public service announcements). This may have been a function of their work-group structure – most students chose to work in pairs with one student completing the “true” PSA while the other completed the “false”, propaganda PSA. Most spent a considerable amount of time on their PSAs cutting out the characters provided for them, drawing images of their own, and decorating the posters. The standard group posters were, however, predominantly text, heavy on factual information, and often much of the text had been copied word-for-word from the fact sheets about the viruses provided to them. Most of the poster creation work by the standard group was coded as operational (81%).

Nearly as many (16) of the experimental group chose to produce serious comics as posters (20). In the serious comic projects created by the experimental group using FluTube, text did not predominate, but was integrated with the backgrounds, characters, and props to narrate and illustrate narrative scenarios in which people might become ill or transmit illness. The experimental group's poster/public service announcement projects typically had no more than three to four phrases and, in contrast to the standard group's projects, the experimental group used more graphics, and often creatively and humorously, to ironic or critical ends. We coded as *incomplete* work that was obviously not finished.

Using Green's 3D model to analyze these multimodal projects gave us a different understanding of what students in each group had learned. It opened up a different assessment lens capable of evaluating a wider range of symbolic-semiotic action and meaning-making than had the conventional pre-post tests, enabling us to seek out and take account of student learning that was not discernible with, nor measurable by, traditional assessment instruments.

All students' multimodal work was blind-coded by the first two authors of the article, who established the criteria together, coded approximately one third of the work together, and then coded separately all remaining work. We next compared our coding decisions for work done separately, and where we found a discrepancy (in fewer than 10 cases), we discussed why, explicating our respective reasons for the way we had each coded that artifact, and were able to reach consensus about the code that best reflected the quality of work initially coded discrepantly.

In total, 80 students (45% of the participants) produced incomplete work (36% of the standard group, 17% of the baseline group, and 59% of the experimental group^{iv}). Of the 30 posters completed in the standard group, 27 were coded as operational and 3 were coded as cultural. Of the 12 standard group students who created complete comics, 7 were coded as operational and 5 as cultural. In the baseline group, 19 students completed posters (no comics) of which 16 were coded as operational and 3 as cultural. Of the experimental group's 20 posters, 15 were coded as operational and 5 as cultural, and of its 16 completed comics, 8 were coded as operational, 5 as cultural, and 3 as critical.^v

Table 5: Student artifacts grouped according to Green's 3D Analytical Model (1988)

	Standard	Baseline	Experimental
Total Students	66	23	88
Coded Work	42	19	36 ^{vi}
Posters total	30	19	20
Operational	27 (90%)	16 (84%)	15 (75%)
Cultural	3 (10%)	3 (16%)	5 (25%)
Critical	0	0	0
Comics total	12	0	16
Operational	7 (58%)	0	8 (50%)
Cultural	5 (42%)	0	5 (31%)
Critical	0	0	3 (19%)
Total Work	42	19	36

Across all 3 groups, the only work coded as critical were the digitally-created comic, whereas 81% of the complete work created by students who did not play *Epidemic* was coded as operational, 18% as cultural, and *none* as critical.

Discussion

This study was designed to engage all participants in learning activities related to contagious disease prevention for the week, in order to see what and how they learned through *Epidemic's* playful digital environment, compared with a traditional pedagogical approach to the same subject-matter. We also wanted to see what, if any, differences became evident between work done using a computer-based multimodal interface to make posters and comics, and work done using non-digital, paper and pencil media/methods. And we wanted to see what perspectives on learning outcomes two different assessment models might provide.

Incidental learning and the building of a technical register

Across all groups, participants showed a high level of sustained engagement with *Epidemic* and its topical content, and appeared to be genuinely interested in, and excited to discuss with each other and the researchers, facts and misunderstandings related to contagious disease. Two of the classes enthusiastically recited to the researchers the importance of hand-washing as there was a poster campaign around their school to discourage the spread of cold and flu viruses. Other participants recounted facts about viral diseases, checking in with the researchers to ensure that their information was correct. Interestingly, what was clear overall, and particularly in relation to HIV, was that these students had rarely encountered curriculum that was as directly related to contagious disease, none that addressed contagious disease as a category of public interest and personal importance, and none of the students could, prior to engaging in this study, articulate the difference between a viral infection and a bacterial one. By the end of the week students were fluently applying professional vocabulary and using medical discourse associated with epidemiology and contagious disease, mobilizing terms like “transmission”, “infection”, “prevention”, and cognate terms in their conversations. For the experimental group, that vocabulary was implicit in and integral to their activities, projects, and communicative exchanges, and by playing with and within *Epidemic*, these students *incidentally* (Alexander, Eaton, & Eagan, 2010; Salen, 2007; Steinkuehler, 2006) acquired and made use of the vocabulary *Epidemic* employs to model and communicate disease-related information. Students situationally acquired an epidemiological “technical

register” (Gee, 2003) to speak about, and multimodally represent, issues and facts surrounding contagious disease and health promotion, even though this vocabulary was not something that *Epidemic* or the researchers directly attempted to teach.

The picture we have from the two formal assessments we used is that whereas the standard group, who had received a lecture with slides, scored higher on the traditional post-test assessment, the experimental group who interacted with *Epidemic* produced more creative, critical digital artifacts through the course of the study and in their final projects— work grounded in a ludic production pedagogy (Thumlert, de Castell & Jenson, 2015; Toohey & Dagenais, 2015), displaying richer intertextual connections to social texts and meanings beyond prescribed lesson content, mobilizing multimodal and critical literacies, as well as relatively sophisticated design strategies when combining image and text. These results suggest that forms of evaluation better suited to apprehending the multimodality of digital media than oriented to literal correctness (as typifies conventional print-culture tests) might substantially enrich the ways educators can identify and evaluate learning outcomes, and the provision of such alternative assessment tools, models and methods might greatly assist teachers in comprehending the opportunities for critical engagement that digital technologies and ludic environments can provide in the classroom.

While no statistical significance was found when comparing score improvement between the experimental group and the baseline and standard groups, we did see statistically significant improvement in scores overall, with a higher average positive change in score in the standard group. More students in the standard group used the information found on the virus fact sheets to create true and false public service announcements that were richer in factual detail than those constructed in the *Epidemic* environment. Had we provided structured and explicit instruction to the experimental group, they may have mobilized and demonstrated greater factual knowledge and understanding, perhaps at least as much as their counterparts in the standard group, who were taught through more traditionally “school-like” health education activities. That is a question to address in future study design.

Our coding and analysis of the student work using Green’s 3D model further illuminates these findings. First, although we sought to make the production project (poster or comic) multimodal for the two non-digital media groups by providing images from *Epidemic*, encouraging the students to use them and/or draw around them, we still saw (as previously noted) the production of text-heavy posters from the standard group—and, importantly, it was those who produced these (text-heavy) posters who significantly improved their post-test scores. In other words, participants took up, used, and reused (through the poster) the literal information that was presented to them and applied it, as well, to the post-test. That participants in the standard group demonstrated the most understanding of propositional fact was no surprise given the traditional pedagogy used to prepare them for the activity (a formal lecture with PowerPoint). But they only demonstrated that understanding at *operational* and *cultural* levels: correctly reporting information and in some cases applying it beyond its original (given) context, and/or mobilizing knowledge drawn from outside their classroom activities. What this group did *not* demonstrate was an ability to challenge, question or contradict given information, deconstruct it, or adopt an ironic perspective toward it. In other words, the standard group was most successful at restating and applying literal knowledge, and adding (literal) outside knowledge relevant to the topic of their posters and comics. Using Green’s 3D model allowed us to re-evaluate success in the game-based learning achieved by the experimental group by using non-traditional criteria to assess multimodal work produced in non-traditional ways, making the quality of knowledge and critical understanding of *that* work measurably evident.

Conclusions: Leveling up

In hindsight, the quiz given to the participants was too easy, as evidenced by the high scores found on the pre-test in all groups. We found its multiple-choice test form ill suited to assess multimodal play-based activities—indeed, why should we be surprised if standardized individual assessment models rooted in print culture and propositional knowledge statements prove ill-suited for assessing learning activities that involve emerging media environments, game-based learning, social interaction and multimodal literacy practices. We echo Amory’s (2010) argument that emerging media environments supporting serious play, games, and multimodal student production (like *Epidemic*), “must be part of a socially collaborative learning experience and should act as tools, and not as tutors, to mediate learning objectives” (p. 825). Notwithstanding our participants’ rather unspectacular showing in terms of test results, it appeared through the projects (digital and otherwise) that the students completed, the questions they asked, and the conversations students had with each other throughout their activities, that there was learning taking place in both groups across both schools.

Recall that the only material coded as critical were the serious comics created by the experimental group: theirs was the most impressive work accomplished, suggesting that this kind of playful, multimodal, and immersive game-like environment can help students “level up” from the compliant (re)production, and literal (re)application, of propositional information that has come to characterize—and to limit—the depth and criticality of traditional (print-literacy driven) classroom tasks and tests. Thus we may, with evidence, begin to challenge, critically reconsider and recast the kinds of statements of fact and claims about values that have for generations been handed on to students more as fodder for memorization than as food for thought.

Our preliminary conclusions echo those of Merchant (2010) and Curwood (2012), suggesting that standard assessment models prefigure and routinize a systemic myopia with regard to the forms of learning transpiring in multimodal and ludic contexts and sites. As Merchant (2010) argues, in the context of learning in virtual worlds, “the current emphasis on standards, derived from measures of individual performance on a rather narrow range of literacy practices coupled with pervasive and powerful discourses of what constitutes literacy instruction, limits our capacity for innovation” (p. 148).

Having seen what using the 3D multimodal assessment tool can reveal, we would second Yancey’s (2004) warning against applying the assessment frameworks “of one medium [print-culture] to assign value and interpret work in a different medium, because by doing so we lose the chance to see new values emerging in the new medium” (Yancey, cited in Sorapure, 2012, p. 431). Indeed, we are only now beginning to find new ways to evaluate students’ capacity to learn, apply and make new meanings in the kinds of multimodal, digital-literacy environments that characterize the world outside of schools today. As Curwood (2012) insists, “if we recognize that learning and knowing within a virtual culture occur within and through multiple, multimodal, and multifaceted textual representations, our approaches to design and evaluation must change” (p. 242).

There is, we contend, an excitingly rich, imaginative, and critical band of experience, learning, and creative/critical action that our standard instruments are not even equipped to sound or detect. It is perhaps more evident today as technologies change, but this failure in the school’s standard forms of learning assessment goes back to Gradgrind, and is very likely implicated in the public system’s inability, from its inception to the present moment, to decouple parental income from educational outcome. Much that students, past as well as present, aspire to think, do and express has been rendered silent and invisible through the public system’s “textual preferences” for a superficial correctness, an easily assessable correctness which passes off in education’s name that kind of literal and uncritical replication of prescribed curricular knowledge most susceptible to near-

immediate forgetting, whose significance is largely contained within school buildings and classroom walls.

References

- Alexander, G. Eaton, I. & Egan, K. (2010). Cracking the code of electronic games: Some lessons for educators. *Teachers college record*, 112(7), 1830-1850.
- Amory, A. (2010). Learning to play games or playing games to learn? A health education case study with Soweto teenagers. *Australasian journal of educational technology*, 26(6), 210-829.
- Apperley, T. & Beavis, C. (2011). Literacy into action: Digital games as action and text in the English and literacy classroom. *Pedagogies: an international journal*, 6(2), 130-143.
- Arnab, S., Brown, K., Clarke, S., Dunwell, I., Lim, T., Suttie, N., Louchart, S., Hendrix, M., & de Freitas, S. (2013). The Development Approach of a Pedagogically-Driven Serious Game to support Relationship and Sex Education (RSE) within a classroom setting. *Computers & Education*, 69, 15-30.
- Barab, S., Goldstone, R., & Zuiker, S. (2009). Transformational play as a curricular scaffold: Using videogames to support science education. *Journal of Science Education Technology*, 18, 305–320.
- Baranowski, T., Buday, R., Thompson, D. I., & Baranowski, J. (2008). *American journal of preventative medicine*, 34(1), 74-82.
- Beale, I. L., Kato, P. M., Marin-Bowling, V. M., Guthrie, N., & Cole, S. W. (2007). Improvement in cancer-related knowledge following use of a psychoeducational video game for adolescents and young adults with cancer. *Journal of Adolescent Health*, 41, 263–270.
- Beavis, C. (2004). Critical perspectives on curriculum and ICTs: The 3D model, literacy and computer games. *Interactive Educational Multimedia*, 9, 77-88.
- Bezemer, J. & Kress, G. (2008). Writing in multimodal texts: A social semiotic account of designs for learning. *Written communication*, 25(2), 166-195.
- Boyle, E.A., Connolly, T. M., Hainey, T., & Boyle, J. M. (2012). Engagement in digital entertainment games: A systematic review. *Computers in Human Behavior*, 28, 771-780.
- Connolly, T. M., Boyle, E. A., MacArthur, E., Hainey, T., & Boyle, J. M. (2012). A systematic literature review of empirical evidence on computer games and serious games. *Computers & Education*, 59, 661-686.
- Curwood, J. S. (2012). Cultural shifts, multimodal representations, and assessment practices: a case study. *E-Learning and Digital Media*, 9(2), 232-244.
- De Castell, S. & Jenson, J. (2003). Serious Play: Curriculum for a Post-talk Era. *Journal of Canadian Association for Curriculum Studies*, 1(1), 47-52.
- Dede, C. 2014. *The Role of Digital Technologies in Deeper Learning*. Students at the Center: Deeper Learning Research Series. Boston, MA: Jobs for the Future.
- DeShazo, J., Harris, L., & Pratt, W. (2010). Effective intervention or child's play? A review of video games for diabetes education. *Diabetes Technology & Therapeutics*, 12(10), 815-821.
- Durrant, C. & Green, B. (2000). Literacy and the new technologies in school education: Meeting the l(IT)eracy challenge? *Australian Journal of Language and Literacy*, 23(2), 89-108.
- Gee, J. P. (2003). *What video games have to teach us about learning and literacy*. New York: Palgrave Macmillan.
- Green, B. (1988). Subject-specific literacy and school learning: A focus on writing. *Australian Journal of Education*, 32(2), 156-179.
- Jenson, J., Taylor, N., & de Castell, S. (2007). Shifting Design Values: A Playful Approach to Serious Content. *E-Learning*, 44, 497-507.

- Jenson, J. & de Castell, S. 2010. Gender, simulation, and gaming: Research review and redirections. *Simulation Gaming* 41(1), 51-71.
- Jenson, J., Fisher, S., & de Castell, S. (2011). Tactics, confidence & leadership: Gender, play & performance. *International Journal of Gender, Science and Technology* 3(1), 149-169.
- Kato, P. M., S. W. Cole, Bradlyn, A. S. And Pollock, & B. H. (2008). A Video game improves behavioural outcomes in adolescents and young adults with cancer: A randomized trial. *Pediatrics*. 122, 2, 305-317.
- Ke, F. (2008). A case study of computer gaming for math: Engaged learning from gameplay? *Computers & Education*, 51, 1609–1620.
- Ketelhut, D. J. (2007). The impact of student self-efficacy on scientific inquiry skills: an exploratory investigation in river city, a multi-user virtual environment. *Journal of Science Education and Technology*, 16(1), 99–111.
- Klopfer, E. (2011). A better model field: Ecology as a model for games and learning. *Teachers college record*, December 22, 2011 <http://www.tcrecord.org> [Date Accessed: 17 July 2013]
- Knight, J., Carly, S., Tregunna, B., Jarvis, S., Smithies, R., de Freitas, S., Dunwell, I., & Mackway-Jones, K. (2010). Serious gaming technology in major incident triage training: A pragmatic controlled trial. *Resuscitation*, 81, 9, 1174-1179.
- Knobel, M. & Lankshear, C. (2008). Remix: The art and craft of endless hybridization. *Journal of Adolescent & Adult Literacy*, 52(1). 22-33
- Lennon, J. L. (2006). Debriefings of web-based malaria games. *Simulation & Gaming*, 37(3), 350-56.
- Lennon, J. L. & Coombs, D. W. (2006). Child-invented health education games: A case study for Dengue Fever. *Simulation & Gaming*, 37(1), 88-97.
- Lennon, J. L. (2010). Debriefing a health related educational game: A case study. *Simulation & Gaming*, 41(3), 390-396.
- Lieberman, D. (2001). Management of chronic pediatric diseases with interactive health games: Theory and research findings. *Journal of Ambulatory Care Management*, 24(1), 26–38.
- Lieberman, D.A. (2012). Video games for diabetes self-management: Examples and design strategies. *Journal of Diabetes Science and Technology*, 6(4), 802-806.
- Linderoth, J. (2012). Why gamers don't learn more: An ecological approach to games as learning environments. *Journal of Gaming and Virtual Worlds* 4(1), 45-61.
- Mayer, R. E. (2104). *Computer games and learning*. Cambridge, MA: MIT Press.
- Merchant, G. (2010). 3D virtual worlds as environments for literacy learning. *Educational Research*, 52(2), 135-150.
- Nelson, B. C. & Ketelhut, D. J. (2008). Exploring embedded guidance and self-efficacy in educational multi-user virtual environments. *International Journal of Computer Supported Collaborative Learning*, 3(4), 413–427.
- Partridge, A., Blair, E., & Leidman, M. (2007). Promoting physical activity within the elementary school curriculum through simulations and games. In C. Crawford et al. (Eds.), *Proceedings of society for information technology and teacher education international conference 2007* (pp. 1230–1235). Chesapeake, VA: AACE.
- Papastergiou, M. (2009). Exploring the potential of computer and video games for health and physical education: A literature review. *Computers & Education* 53(3), 603-622.
- Prensky, M. (2006). *“Don't bother me mom – I'm learning”*. Paragon House: St. Paul, Miss.
- Rieber, L. P. (1996). Seriously considering play: Designing interactive learning environments based on the blending of microworlds, simulations, and games. *Education Technology Research & Development* 44(2), 43-58.
- Salen, K. (2007). Gaming literacies: A game design study in action. *Journal of Educational Multimedia and Hypermedia*, 16(3), 301–322.

- Sorapure, M. (2010). Between modes: Assessing student new media compositions. In I. L. Clark (Ed.), *Concepts in composition: Theory and practice in the teaching of writing* (pp. 431-442). New York & London: Routledge.
- Sitzmann, T. (2011). A meta-analytic examination of the instructional effectiveness of computer-based simulation games. *Personnel Psychology*, *64*, 489–528.
- Squire, K. (2011). *Video games and learning: Teaching and participatory culture in the digital age*. New York: Teachers College Press.
- Steinkuehler, C. (2006). Massively multiplayer online videogaming as participation in a discourse. *Mind, Culture and Activity*, *13*(1), 38-52.
- Sward, K. A., Richardson, S., Kendrick, J., & Maloney, C. (2008). Use of a web-based game to teach pediatric content to medical students. *Ambulatory Pediatrics*, *8*(6), 354–359.
- Thompson, D., Baranowski, T., Buday, R., Baranowski, J., Thompson, V., Jago, R., & Griffith, M. J. (2010). Serious video games for health: How behavioral science guided the development of a serious video game. *Simulation & Gaming*, *41*(4), 587-606.
- Toohey, K. & Dagenais, D. (2015). Videomaking as sociomaterial assemblage. *Language and Education*, *29*(4), 302-316.
- Thumlert, K., de Castell, S. & Jenson, J. (2015). Short cuts and extended techniques: Rethinking relations between technology and educational theory. *Educational Philosophy and Theory*, *47*(8), 786-803.
- Unnithan, V., Houser, W., & Fernhall, B. (2006). Evaluation of the energy cost of playing a dance simulation video game in overweight and non-overweight children and adolescents. *International Journal of Sports Medicine*, *27*, 804–809.
- Wouters, P., van der Spek, E. D., & van Oostendorp, H. (2009). Current practices in serious game research: a review from a learning outcomes perspective. In T. M. Connolly, M. Stansfield, & L. Boyle (Eds.), *Games-based learning advancements for multisensory human computer interfaces: Techniques and effective practices* (pp. 232–255). Hershey, PA: IGI Global.
- Wouters, P., van Nimwegen, C., van Oostendorp, H., & van der Spek, E. D. (2013). A meta-analysis of the cognitive and motivational effects of serious games. *Journal of Educational Psychology*, *105*(2), 249-265.
- Yancey, K. B. (2004). Looking for sources of coherence in a fragmented world: Notes toward a new assessment design. *Computers and Composition*, *21*(1), 89-102.
- Young, M. F., Slota, S., Cutter, A. B., Jalette, G., Mullin, G., Lai, B., Simeoni, Z., Tran, M., & Yukhymenko, M. (2012). Our princess is in another castle: A review of trends in serious gaming for education. *Review of Educational Research*, *82*(1), 61-89.

Biographical Information

Jennifer Jenson is Director of the Institute for Digital Learning and Professor of Pedagogy and Technology in the Faculty of Education at York University, Canada. She has published on gender and gameplay, technology and education, and games and learning.

Contact: jjenson@edu.yorku.ca

Suzanne de Castell is Professor Emeriti in the Faculty of Education Simon Fraser University. Her considerable work includes publications on digital games and education, gender and technology, literacy and school, and queer pedagogies.

Kurt Thumlert is a Lecturer in the Faculty of Education at York University, Canada. His published work includes art-based research and technology and pedagogy.

Rachel Muehrer is a Research Associate in the Play:CES Lab at York University, Canada. She has published on games and learning, and music education.

Acknowledgements

This research was supported by the GRAND NCE network and the Social Sciences and Humanities Research Council of Canada. We also gratefully acknowledge the considerable design work and research support provided by Dr. Nicholas Taylor, as well as are grateful for the students and their teachers who agreed to participate in the study.

ⁱ Learning ways to engage online *without* disclosing private information is a major concern, as we have discovered from prior work with children playing in virtual environments. In fact, in one extensive study of online gameplay, we discovered that use of one's own name was a reliable indicator of the player's being underage (Author names redacted for review).

ⁱⁱ For example, the pre-test tested knowledge about the symptoms and means of prevention, transmission, and treatment for a range of contagious diseases including HIV/AIDS, SARS, West Nile, smallpox, and influenza.

ⁱⁱⁱ In order to achieve consistency, two researchers initially coded each student artifact together using the 3D model.

^{iv} Although completion data suggests that the students in the experimental group may have done a lot more "playing around" than the Standard group students, technical glitches in the *Epidemic* environment meant that many experimental group students were simply not able to complete their work, or, if they completed it, may have been unable to save it on the server: we needed (but did not have) a way our completion data could discriminate between non-completion due to technical problems, and non-completion due to substantive difficulties with the task or non-engagement with it.

^v That only serious comics were coded as critical has by no means escaped our attention. This may of course indicate some insufficiency in our coding. It is also possible, however, that different media privilege different cognitive functions. In that case, the ways knowledge is understood, applied, and represented cannot be divorced from the specific media forms used: for example, public information posters might just be less conducive to criticality than the more graphically complex medium of the comic panel (which only the experimental group *elected* to use to demonstrate knowledge). This is a far larger question than can be tackled here, but it is interesting - and worthy of further investigation - that across all students in all 3 groups, notwithstanding other variations in performance, work in only one media form was judged to be critical. Although this was not something we anticipated prior to seeing the results of student efforts, it is something we will take into account in this study's next iteration.

^{vi} For an explanation of why the experimental group has two-thirds of the participants but only one-fourth of the completed work, refer to note 4.