Participatory Assessment: A Game Design Model for Impacting Engagement, Understanding, and (as Necessary) Achievement

Daniel T. Hickey, Indiana University, dthickey@indiana.edu

Abstract: Participatory Assessment is a game design model for obtaining diverse learning and/or social outcomes in innovative learning environments. It fosters participation in socio-technological interactions that ensures individual understanding of targeted concepts. As necessary, the model has also been capable of improving and documenting the impact of aggregated achievement. The model emerged from assessment-oriented design studies in several environments, including the Quest Atlantis 3-D virtual environment. This paper introduces the five general design principles that make up Participatory Assessment, along with the more specific design principles that emerged across given design cycles of the Taiga game in Quest Atlantis. Specific game features are summarized, along with evidence of the impact of those features in the Taiga design studies.

Designing games for impact often forces the choice between (a) more direct "expository" approaches, (b) more constructivist "inquiry-oriented" approaches, or (c) more sociocultural "situated" approaches. This paper introduces an assessment-oriented game design model that is intended to address core tensions that emerge when designing, using, and evaluating video games to attain educational and/or social impact. This model is called *Participatory Assessment.* The model consists of five general design principles that emerged in design studies in innovative learning contexts. This paper describes these principles as they emerged in studies of the *Taiga* educational video game in *Quest Atlantis*.

Reconciling Competing Approaches to Games for Impact

Early text-based games like *PLATO* and PC games like *MathBlaster* provide "drill & practice" of specific factual and procedural knowledge. These "expository" approaches expose players to numerous specific associations. A new generation of games like *Dimension M* embeds drill and practice in mathematics into complex immersive games. Critics point out that in most drill and practice games, the relationship between the game activity and the academic content is arbitrary. While this use of extrinsic rewards simplifies game design, any learning regarding the structure, rules, or story of the game itself does not reinforce or enhance the academic content (Rieber, 2005). Furthermore, extrinsic rewards have been shown in hundreds of studies to diminish subsequent interest and engagement (Lepper & Hodell, 1989).

The "cognitive revolution" of the 1970s led to video games that embraced constructivist theories of learning and intrinsically motivated learning. LOGO and *Zoombinis* emphasized intrinsically rewarding activities that called on fantasy and curiosity, while avoiding lower-level content and extrinsically rewarded activity (Lepper & Malone, 1987). More recently, *Spore* and *World of Goo* are presumed to build critical thinking skills and deep conceptual knowledge (Kafai, 2006). But constructivist and constructionist innovators have traditionally been hard-pressed to show educational impact (Egenfeldt-Nelson, 2006).

A third wave of innovation reflects newer situative theories of cognition (e.g., Greeno, 1998) that focus on learning as successful interaction with social, technological, and informational resources. An intriguing possibility of massive multiplayer games is that they can support specific types of social interaction that lead to broader social learning of academic knowledge, as well as the more salient and readily measured individual knowledge. In this way, players can confront formal concepts and abstract principles while solving real problems (Shaffer, Squire, Halverson, & Gee, 2005). This allows learners to directly interact with complex social systems that are otherwise inaccessible to them (Squire, 2003).

The options offered by these different views introduce tensions in game design. The tensions between expository and constructivist approaches are particularly problematic. They are premised on assumptions about individual learning that are ultimately antithetical (Case, 1996). Situative perspectives introduce additional tensions because they focus the designer's attention more on fostering productive social interaction. The five design principles introduced in this paper address these tensions by focusing *primarily* on social learning via interactive participation, and only *secondarily* on individual learning outcomes. Put differently, this means that designing and refining features should focus on helping players and teachers informally assess engaged participation in interactive discourse concerning the to-be-learned knowledge. This is theoretically consistent with Greeno's (1998) notion of *engaged par-*

ticipation and practically consistent with Engle and Conant's (2002) notion of *productive disciplinary engagement.*

The Five Design Principles of Participatory Assessment

Participatory Assessment is rooted in prior multi-year design studies of interactive multimedia for genetics (Hickey & Zuiker, 2013) and space science (Hickey, Taasobshirazi, & Cross, 2013). Insights from those studies were subsequently refined in several parallel strands of design research. This included five annual cycles of design research in *Taiga*, the first of many "worlds" that now make up *Quest Atlantis*. Taiga is a 15-20 hour game involving ecological science and socio-scientific inquiry for grades 4-6, in which students investigate the reasons for declining fish populations in a river. The game is organized around five *Quests* where players draft and submit *quests* (essentially field reports) for Ranger Bartle. The teacher then "plays" the role of Ranger Bartle in reviewing and accepting the reports. Practically speaking, this means Taiga has the potential to support a great deal of student writing, which distinguishes it from many educational games.

The studies were carried out with two elementary school teachers over five years. One teacher taught a single class of academically talented fourth graders from 2005-2007. The other teacher taught four classes of sixth graders from 2006-2010. The specific game design principles and specific features in Taiga will be described below in the context of five game design principles that emerged in this research.

Principle 1: *Let Contexts Give Meaning to Knowledge Tools*

The first step in Participatory Assessment is reframing targeted knowledge as *tools.* This reframed learning as practicing using tools appropriately in particular contexts. In Taiga, this meant first defining a compelling narrative game that required using knowledge of ecology and socio-scientific inquiry to play. This then meant fine-tuning that narrative to require the student to use more of that knowledge to succeed.

The initial design of Taiga transformed the concepts of elementary ecological science into knowledge tools that could be used to solve important *socio-scientific* problems. Such problems evade simplistic explanations and require balancing a host of issues in advancing plausible hypotheses and solutions. In Taiga, players serve as apprentices to Ranger Bartle. In this way, Taiga (and most subsequent QA worlds) incorporated the foundational characterization of situative instruction as "cognitive apprenticeship" (Collins, Brown, & Newman, 1989; Lave, 1977). As elaborated in Barab, Sadler, et al., 2007, p. 61-62), the initial effort to "narratize" ecological science in Taiga involved creating a narrative that was compelling to students and whose solution required using scientific inquiry to use scientific resources in the service of identifying underlying cause(s) of the core problem introduced by the narrative.

The initial version of Taiga was implemented in 2005 by the fourth grade teacher across fifteen periods. An open-ended essay on socio-scientific inquiry was used in 2005. The students in 2005 made tremendous gains in their socio-scientific essays. This made sense because the students really had no experience with these ideas or this kind of scientific inquiry before Taiga. However, the scores on the achievement test only went up slightly and the gain was not statistically significant (Figure 1). More importantly, interpretive analyses of the quest submissions showed that many students had failed to even mention the targeted scientific practice or resources in their reports (Barab, Sadler, et al., 2007).

As elaborated in Barab, Zuiker et al., (2007) the design team then set out to "scientize the narrative." They framed increasingly formal relationships between the targeted scientific formalisms (i.e., tools) and their context-of-use in Taiga. Their continued refinement of the narrative and the assessments distinguished between formalisms that are *embodied* by, *embedded* in, or *abstracted* from the social and material context of the game. The revised Taiga was implemented in 2006 by the same fourth grade teacher with a new class. The curriculum-oriented assessment was revised to include some knowledge that students might have previously encountered, and the achievement test was recreated to include more items for just four of the most relevant state science standards. The revisions were effective in that many more of the students enlisted many more of the domain formalisms in their quest submissions, and did so more meaningfully. The gains in understanding of the new curriculum-oriented assessment were smaller but were still statistically significant; the gain in achievement on the standards-oriented test doubled from the previous year and was statistically significant (Figure 1). Later in 2006, a sixth-grade teacher implemented Taiga for the first time in two of his four sixth grade science classes. As shown in Figure 2, the gains in understanding and achievement were statistically significant for both pairs of classes, but both gains were larger in the Taiga classes (Hickey, Ingram-Goble, & Jameson, 2009). Additional analysis showed many other very positive outcomes in the QA group and showed that the knowledge the students did gain lasted much longer (Barab, Gresalfi, & Ingram-Goble, 2010).

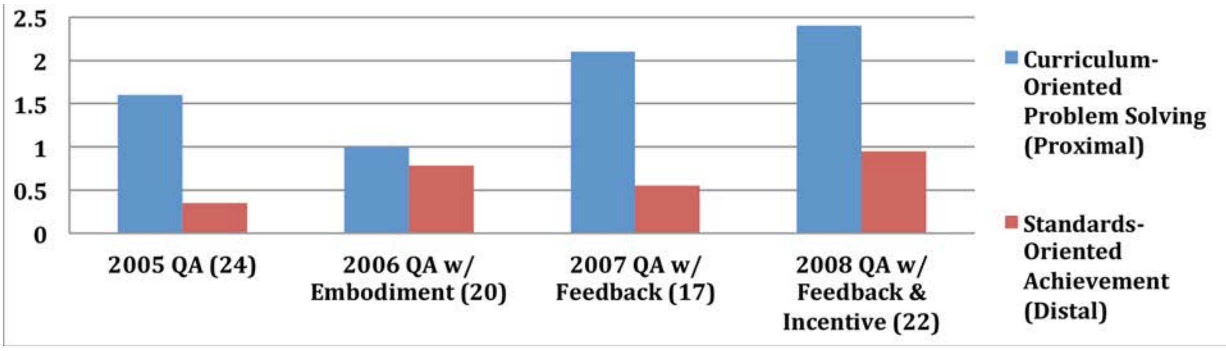

Figure 1: Learning gains in fourth grade classe across years (in SDs with number of students)

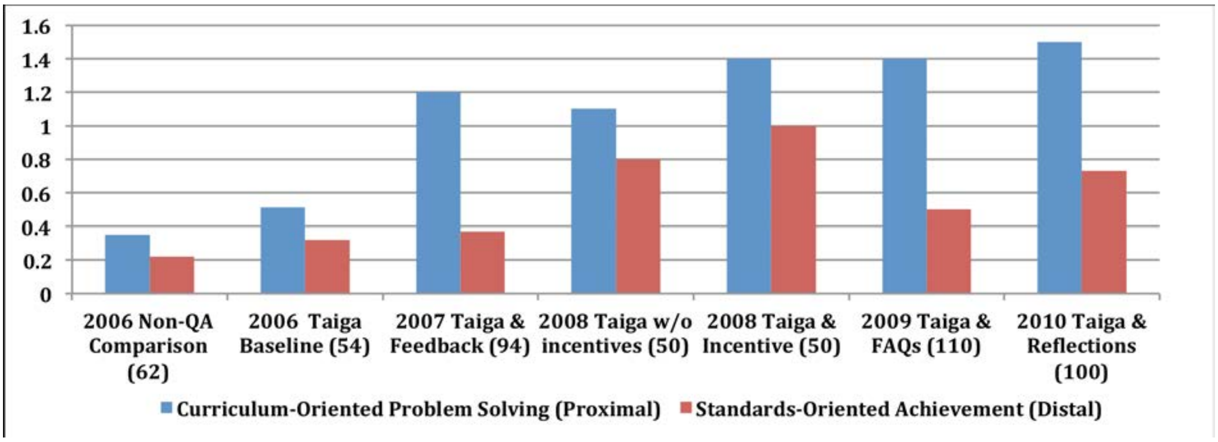

Figure 2. Learning Gains in sixth-grade classes across years (in SDs, with numbers of students)

Principle 2: *Recognize and Foster Productive Disciplinary Engagement*

This second design principle emerged from a more focused series of refinements around the process of drafting, submitting, and revising the five written quests that organized much of the gameplay. Analysis of the quests submitted in the sixth grade classrooms in 2006 confirmed that the quality of the submissions was modest, the feedback provided was hasty, and the revised submissions did not improve. An ecology graduate student was invited to join the team in 2007 and improve our performance assessment, quests, and activities leading up to the quests. She created the *Lee River* performance assessment described below and helped create two new information resources to help foster more productive disciplinary engagement around the questing process. The first new resource was a detailed scoring rubric for the crucial second quest that required the most synthesis of knowledge. The rubric provided carefully worded examples and feedback that aimed to make this structured discourse more productive and more disciplinary. Examples and descriptions of *incomplete, partial, nearly complete,* and *complete* submissions were included, along with feedback text that could be cut and pasted into the feedback window and then customized as desired for each student. The second new resource consisted of screens of information that were embedded for players to use when revising their reports.

In 2007, both teachers implemented this revision of Taiga. The submitted quests showed that these refinements led students to use many more of the domain tools *and* use them more appropriately, with the resubmitted reports containing evidence of this much more than the initial reports that students were asked to revise. As reported in Hickey, Ingram-Goble, & Jameson (2009) and shown in Figures 1 and 2, we also observed dramatically larger gains in understanding and larger gains in achievement compared to 2006. However, uneven and generally low use of feedback made us wonder about motivation to use feedback. While incentives are popular in commercial video games, they remain controversial in educational video games. In 2008, we removed the incentives in Taiga from two of the sixth grade classes and emphasized intrinsic reasons for succeeding (such as helping the park and helping Ranger Bartle do his job). In the other classes, we made the incentives more salient and added some additional ones. In addition to the backpacks and hats, accepted quests also resulted in a badge that students could place on their avatar, which corresponded to the quality of the submission as judged by the teacher/ranger (*knowledgeable, expert,* or *wise*). We also placed a physical “leader board” on the wall in those two classrooms.

That the incentives led players to enlist formalisms more appropriately suggested that they were *not* taking shortcuts in their submissions to get the incentives. While we were unable to systematically compare feedback use in the two groups, the students in the incentive classes made significantly larger gains in understanding and somewhat larger gains in achievement (see Figure 2). Additionally, they reported slightly higher levels of motivation while completing their reports and slightly larger gains in interest in solving these kinds of problems. Thus, we found no negative consequences of incentives and some positive consequences.

In 2009, we also revised the embedded feedback to make it easier to use. The previous feedback screens were slightly hard to navigate, and players had to commit to a whole series of them. We changed it so that when students went to the technician for help, he presented them with a list of about 20 questions. We also further grounded this feedback into the narrative problem context. While this certainly made the feedback more accessible, gains in understanding for the sixth graders were unchanged.

Principle 3: *Assess Reflections Rather Than Artifacts*

One of the distinguishing feature of QA and what arguably makes it so educationally useful is that players generate “artifacts” that are personally meaningful and that feature directly in the game narrative. Artifacts are things that have been made meaningful (Lave & Wenger, 1991). The artifacts that students create in project-based learning are more meaningful than worksheets because the personalization possible in a project means that the artifact takes on additional meaning. In some games, these artifacts are virtual items that players have collected or won. In Quest Atlantis, these artifacts include written communications such as commentaries. Because of their role in rewarding, acknowledging, and supporting learning, the design and function of artifacts is an important consideration and a potential source of tension in designing educational games.

In 2009, we began experimenting with reflection questions that built upon the sub-narrative that Ranger Bartle was a busy mentor. The reflections were framed as requests from Bartle to help him determine whether each report showed evidence that they were fulfilling their responsibility as an apprentice. The reflection questions for Quest Two were as follows: *Remember, you are here as an apprentice. Help me make sure you are becoming a skilled ranger. Explain what it is about your quest that shows you understand the following things about hypotheses:*

1. *The things that a hypothesis must include to be scientific;*
2. *That a testable hypothesis must have enough detail for someone else to test it if they want;*
3. *That experts always look for and include other alternative explanations for their hypothesis;*
4. *That experts always consider what they might have overlooked.*

We asked the teacher to review the reflections primarily and were pleased with the way doing so seemed to streamline the reviewing process. Sometimes, when students went to draft the reflection, they would realize that there were things missing from the submission, which they would then go back and complete. We realized that completing the reflections for one quest submission could shape the way that students approached the next submission. The next time, it was expected that students would *start* their submission by considering the reflection while drafting and revising the report; this, in turn, had the potential to shape engagement in the activities leading up to each report. Reviewing the Quest Two submissions in 2009 showed that most (but not all) students took the reflections seriously, and we concluded that reflections were a promising strategy for increasing disciplinary engagement while streamlining the review process. These ideas did not fully come together until the final 2010 study, and several factors precluded systematic study of the reflections. While gains in understanding and achievement were about the same, we ended up with a more sustainable teacher workload. Compared to 2008, the number of resubmissions declined (from an average of 3.1 to 1.9 per student).

Principles 4: *Assess Individual Understanding Prudently*

The last two principles in the model respond to one of the central tensions facing educational game designers. In our view, all learning involves assessment, which means that the clear distinction between “instruction” and “assessment” disappears. This view also does away with the sharp distinction between formative and summative assessments. Rather, particular assessment practices are understood in terms of their potential formative and summative functions, along a continuum ranging from informal to formal. Of course, assessment and testing raise complex issues about the *validity* of the inferences that can be drawn from scores. A situative perspective on assessment argues that one must specify a theory of knowing *and* a theory of learning when discussing validity.

Achievement tests can provide valid evidence of how much individuals know about broad domains of knowledge that accrues over very long time scales; this means that achievement tests are hard-pressed to provide valid evidence of how much individuals learn from specific learning activities. A situative perspective on assessment raises complex theoretical issues that are beyond the scope of this paper and are elaborated elsewhere (Hickey & Anderson, 2007; Hickey & Zuiker, 2012). The most important point for this chapter is that situative perspectives suggest careful *alignment* of less formal assessments with more formal assessments. Doing so allows the summative function of the more formal assessment to "protect" the formative function of the less formal assessment.

Administering a curriculum-oriented assessment before and after Taiga showed how much particular students were learning about the scientific concepts in Taiga and how much students overall were learning about particular concepts. It also showed that some students understood some of the concepts *before* playing. These insights were used to refine the Taiga activities and quests. The increasingly larger gains in understanding showed that these efforts were successful. However, we did not have the teacher use the Lee River assessment to provide feedback directly to learners. While doing so might have supported more *student* learning, it would have undermined *project* learning about the effectiveness of the curriculum. This is because providing individual feedback would focus the teacher's attention too directly on the abstracted formalisms in the assessment items and would likely prepare students too directly for the post-test throughout the activity.

Principle 5: *Measure Aggregated Achievement Discreetly*

Even with our efforts to preserve the validity of the curriculum-oriented performance assessment, the iterative alignment of Taiga to the assessment still introduced an unknown (and practically unknowable) degree of "construct-irrelevant easiness" (Messick, 1994). This points to a core tension in game-based assessment. These refinements meant that some part of the improvement from one cycle to the next was the result of Taiga better familiarizing players with the problems that would appear on the Lee River. This meant that the Lee River could not yield valid evidence in comparisons with other curriculum or predict the impact on external achievement tests. An additional instrument was needed.

A "standards-oriented" achievement test was created by randomly sampling items from pools of items that were aligned to targeted standards, independent of Taiga. Because such items can be answered quickly, it was possible to include a large amount of items. As long as an individual has not been directly exposed to the specific associations on the test items, it is possible to efficiently and reliably compare how much individuals know about a domain of knowledge. Such tests should not be used to directly shape the way a curriculum is designed or enacted. As argued above, the human mind is remarkably efficient at learning information well enough to use it to recognize specific associations. It is all too easy when designing and/or teaching a curriculum to reference the specific associations that are needed for specific test items.

Putting it All Together in Iterative Refinements

What makes Participatory Assessment work is the way that the activities at one "level" are motivated and evaluated by the outcomes at the next. In Taiga, this meant focusing directly on participation in the game, less directly on understanding, and very indirectly on achievement. In educational game design, this means fostering participation before assessing understanding, and fostering understanding before measuring achievement. Learning across levels is increasingly formal and encompasses increasingly broad knowledge. It is also decreasingly contextualized and occurs over longer and longer timescales (Lemke, 2000). Rather than presenting different or more difficult problems across levels, students interact with increasingly formal representations of the same domain of knowledge across the levels. The ultimate power of this alignment comes from its potential for coordinating the activities of all of the participants. This includes the students (because activity at one level can be motivated by the desire to succeed at the next level), the designers (by providing a target for activities at each level), the teachers (by providing a goal to shape the enactment of each level), and the researchers (by providing project goals and summative evidence of success). Because the knowledge is transformed from one assessment level to the next, evidence of the transfer is obtained. By doing so over three or more levels and extending out to the level of the distal achievement test, valid evidence of achievement impact is obtained. However, this evidence is obtained without resorting to directly exposing students to specific associations that might appear on the targeted test.

The continued improvement to engagement, understanding, and achievement over implementation years suggests that this model helps deliver the diverse sorts of learning outcomes and evidence of those outcomes that have long eluded educational innovators. Additional support for the value of this model comes from similarly increased outcomes in the other programs of research in which these same design principles were refined. These studies have included high school language arts (e.g., Hickey, McWilliams, & Honeyford, 2011), online graduate education courses (Hickey & Rehak, 2013), and hybrid undergraduate lecture courses in telecommunications

(Walsh & Hickey, 2011)

References

Barab, S., Sadler, T., Heiselt, C., Hickey, D., & Zuiker, S. (2007). Relating narrative, inquiry, and inscriptions: A framework for socioscientific inquiry. *Journal of Science Education and Technology, 16,* 59-82.

Barab, S. A., Gresalfi, M., & Ingram-Goble, A. (2010). Transformational play: Using games to position person, content, and context. *Educational Researcher*, *39*(7), 525-536.

Case, R. (1996). Changing views of knowledge and their impact on educational research and practice. In Olson, D. & P. Torrance (Eds.), *The handbook of education and human development* (pp. 75–99). Cambridge, MA: Blackwell.

Collins, A., Brown, J. S., & Newman, S. E. (1989). Cognitive apprenticeship: Teaching the crafts of reading, writing, and mathematics. In L. B. Resnick (Ed.), *Knowing, learning, and instruction: Essays in honor of Robert Glaser* (pp. 453–494). Hillsdale, NJ: Erlbaum.

Engle, R. A., & Conant, F. R. (2002). Guiding principles for fostering productive disciplinary engagement: Explaining an emergent argument in a community of learners classroom. *Cognition and Instruction*, *20*(4), 399–483.

Egenfeldt-Nielsen, S. (2006). Overview of research on the educational use of video games. *Digital Kompetanse*, *1*(3), 184–213.

Greeno, J. G. (1998). The situativity of knowing, learning, and research. *American Psychologist*, *53*(1), 5-26.

Greeno, J. G., & Gresalfi, M. S. (2008). Opportunities to learn in practice and identity. In Moss, P, Pullin, D. C, Gee, J. P, Haertel, E. H., & Young, L. J. (Eds.), *Assessment, equity, and opportunity to learn* (pp. 170–199). Cambridge, MA: Cambridge University Press.

Hickey, D. T., & Anderson, K. (2007). Situative approaches to assessment for resolving problems in educational testing and transforming communities of educational practice. In P. Moss (Ed). *Evidence and decision making. The 103rd NSSE Yearbook* (pp. 269-293). National Society for the Study of Education/University of Chicago Press.

Hickey, D. T., Ingram-Goble, A., & Jameson, E. (2009). Designing assessments and assessing designs in virtual educational environments. *Journal of Science Education Technology, 18,* 187-208.

Hickey, D. T., McWilliams, J. T., & Honeyford, M. A., (2011). Reading Moby-Dick in a participatory culture: Organizing assessment for engagement in a new media era. *Journal of Educational Computing Research, 44* (4), 247-273

Hickey, D. T., & Rehak, A. (2013). *Wikifolios and participatory assessment for engagement, understanding, and achievement in online courses.* Accepted for publication the *Journal of Educational Media and Hypermedia,* January 2013.

Hickey, D. T., Tassoobshirazi, G., Cross, D. (2012). Assessment *as* learning. Enhancing discourse, understanding, and achievement in innovative science curricula. *Journal of Research in Science Teaching, 49,* 1240-1270.

Hickey, D. T., & Zuiker, S. J. (2012). Multi-level assessment for discourse, understanding, and achievement in innovative learning contexts. *The Journal of the Learning Sciences, 22,* (4) 1-65.

Kafai, Y. B. (2006). Playing and making games for learning. *Games and Culture*, *1*(1), 36.

Lave, J. (1977). Cognitive consequences of traditional apprenticeship training in West Africa. *Anthropology & Education Quarterly*, *8*(3), 177–180.

Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. Cambridge, MA: Cambridge University Press.

Lemke, J. L. (2000). Across the scales of time: Artifacts, activities, and meanings in ecosocial systems. *Mind, Culture, and Activity*, *7*(4), 273-290.

Lepper, M. R., & Hodell, M. (1989). Intrinsic motivation in the classroom. *Research on Motivation in Education*, *3*, 73–105.

Lepper, M. R., & Malone, T. W. (1987). Intrinsic motivation and instructional effectiveness in computer-based education. *Aptitude, Learning, and Instruction*, *3*, 255–286.

Messick, S. (1994). The interplay of evidence and consequences in the validation of performance assessments. *Educational Researcher*, *23*(2), 13-23.

Rieber, L. P. (2005). Multimedia learning in games, simulations, and microworlds. *The Cambridge handbook of multimedia learning*, 549–567.

Shaffer, D. W., Squire, K. R., Halverson, R., & Gee, J. P. (2005). Video games and the future of learning. *Phi Delta Kappan*, *87*(2), 104–111.

Squire, K. (2003). Video games in education. *International Journal of Intelligent Simulations and Gaming*, *2*(1), 49–62.

Walsh, J. D., & Hickey, D. T. (2011, October). *Designing for participation in a 125-student hybrid media production course.* Presentation at the Indiana University Center for Innovative Teaching and Learning faculty poster fair.