# First Things First: Design Principles for Worthwhile Educational Videogames

Daniel T. Hickey, Sasha A. Barab, Adam Ingram-Goble, Indiana University, USA
Email: dthickey@indiana.edu, aingramg@indiana.edu, sbarab@indiana.edu
Steven J. Zuiker, National Institute of Education, Singapore, steven.zuiker@nie.sg

**Abstract:** Three design principles are advanced for multi-user educational videogames. First, they should *support situative embodiment in academic knowledge,* where personally meaningful activities and coherent narratives foster collective engagement. Second, they should *offer multiple levels and forms of meaningful assessment* and the opportunity to succeed, fail, and try again. Third, they should *provide useful feedback that is used* to enhance participation, learning, and curricula. These principles were developed in three annual design-based refinements of a 15-hour ecological sciences gaming curriculum in nine upper elementary classes. Across years, the situative embodiment afforded by the curriculum was refined with informal assessment, and innovative virtual formative feedback was incorporated around a key curricular activity. Results across years revealed incremental improvements in participation, understanding of key concepts, and achievement of targeted standards. The ultimate gains in understanding and achievement were larger than those in comparison classrooms that used a conventional text-based curriculum covering the same concepts and standards.

The worldwide success of commercial multi-user virtual environments (MUVES) is arguably the result of the amount of learning those games support. While the knowledge learned is often dismissed by parents and educators, games and the gaming industry support unprecedented levels and types of learning. Games like *World of Warcraft* allow novices to quickly develop sufficient skills and understanding to become an increasingly central participant. Game-based and emergent social scaffolding defines a remarkably strong trajectory towards sophisticated collective activity, requiring thoughtful communication and sophisticated insights about the thoughts and actions of other participants. Arguably, this level of learning is unprecedented in designed learning environment. This has profound implications for advancing and supporting learning more broadly.

The success of commercial games has fostered an explosion of interest in educational relevance, fueled by books (e.g., Gee, 2003; Shaffer, 2006), articles (i.e., Squire, 2006; Schaffer, Squire, Halverson, & Gee, 2005), reports (Federation of American Scientists, 2006), and initiatives (e.g., The MacArthur Foundation's Digital Learning Initiative (1) and the European Union's ELEKTRA project (2)). While acknowledging the fundamental differences when considering educational content and contexts, most of the aforementioned scholars argue that educational researchers should examine the design features of commercial games that support learning, and study how those features are (or are not) related to learning academic knowledge. Most of these scholars also argue that newer situative views of teaching and learning are ideally suited for appreciating and enhancing the ways that videogames might be used to enhance learning of academic knowledge. This paper presents three design features that emerged across a three-year effort that used situative approaches to engagement, assessment, and feedback to enhance student participation, understanding, and achievement around on multi-user educational videogame.

## Quest Atlantis and the Taiga Ecological Sciences Curriculum

Quest Atlantis (QA) immerses learners in a virtual environment where they can experience complex situations and participate in compelling interactive narratives.(3) QA affords a game experience while including sufficient academic content to warrant use by classes of school children, typically in close collaboration with their classmates and teacher in a school computer lab. Questers are asked to help resolve virtual dilemmas on Atlantis by generating and evaluating solutions. Students navigate the three-dimensional space via their avatars, interacting with other players via text-based "chat," and with non-player characters (NPCs) via structured dialogues. Academic content is organized around "Quests" that students complete and submit electronically in-game, which are then reviewed by the teacher who inhabits the role of one of the NPCs in the underlying narrative. Quests are embedded in a mission, which make explicit the sequence of tasks Questers experience in accomplishing their goals. Authoring software makes it simple to add or remove features, and a sophisticated interface allows assignment of different features or versions to different students or classes. This makes QA an ideal laboratory for studying how different design features impact engagement and learning.

*Taiga,* one of the many worlds in QA, is a park located along a river, populated by loggers, tourists, indigenous farmers, a fishing resort, and park administration. Taiga was designed to support 15-20 hours of activity where students grades 4-6 learn standard ecological science concepts like erosion, eutrophication, and hypothesis testing, via interaction with virtual characters and data that allows students to evaluate competing explanations for declining fish populations in the Taiga River. Students learn this content while also learning to engage in complex and potentially controversial socio-scientific inquiry (SSI, Sadler, 2004). The problems go well beyond what even middle school teachers typically attempt. Players interact with 15 Non-Player Characters (NPCs) in Taiga and complete four Quests. Success in the game requires students to understand and use (1) *indicators* such as pH and dissolved oxygen, (2) the *processes* of erosion, eutrophication, and, and (3) the dynamic *relationship* between indicators and processes in socioscientific inquiry in the virtual water system.

Between 2005 and 2007, Taiga was refined across three implementation cycles implementations, in a total of 9 classes with data collected from 200 Questers (data was also collected in two non-QA comparison classrooms taught by one of the QA teachers). Each year, design-studies were conducted in order to build useful theories and practices for enhancing the quality of student experience and learning of academic content. Following is a summary of the first key design principle that emerged in these studies. This is followed by a more detailed account of the second and third principles, whose usefulness are closely bound to the first principle.

## Using Conceptual Play Spaces to Foster Situative Embodiment

One of the most important factors in the success of commercial videogames is their support of deep, extended engagement in activities that players find personally meaningful. As popularized by Gee (2004), this engagement is highly situated in the virtual space of the game; arguably 3-D virtual environments afford a type of engagement that is best captured by contemporary *embodied* characterizations of cognition (e.g., Glenberg, 1997). The title of this paper reflect our belief that the *first* step in designing worthwhile educational videogame is creating an engaging narrative situation that supports embodied experiences in which relevant academic content can play a meaningful role. Thus, the first design principle explored in this paper is that *educational videogames should to support situative embodiment in academic knowledge.* The initial implementations in this study focused on engaging questers in personally meaningful ways with an important socio-scientific dilemma, which could only be solved using scientific concepts and inscriptions. These efforts highlighted the interesting junction which we have reached. On the one hand, teachers and schools face increased pressure to prepare students for standardized tests, while on the other hand, they face a generation of students who regard the school curriculum as largely irrelevant to their own lives. The efforts to bridge this gap resulted in a related design principle known as *conceptual play spaces*. As elaborated in Barab et al. (in press), conceptual play spaces are curricular contexts that leverage what is known about how people learn, the metaphorical power of narratives, and game design principles, to establish an educational, entertaining, and personally transformative context for learning. Rather than using technology to efficiently deliver abstracted content, Quest Atlantis is designed so that content maintains its connection to the world. Rather than imparting some ready-made, expert description of a concept (a thing), the curriculum itself should constitute an experience or world-to-be-lived (an activity). Fundamentally, conceptual play spaces establish a problem situation that can only be solved if one employs disciplinary concepts as a tool. They define an immersive context that changes in response to user actions. That provides opportunities to examine one’s participation in relation to the impact of one’s choices in the immersive context.

Conceptual play is characterized by four modes of engagement: *conceptual engagement, immersive engagement, transactive engagement,* and *reflexive engagement*. Briefly, conceptual engagement involves enlisting target concepts in the service of solving a particular task. Immersive engagement then involves situating the task in the context of a larger dramatic storyline in which the learner participates. Transactive engagement occurs when one’s actions affect another, and at the same time, impacts the actor. This serves to position the self in terms of the virtual world. Reflexive engagement involves the learners examining how their participation changed the environment, and then using this understanding to interrogate the dynamics of the environment as well as their role in influencing these dynamics. In creating spaces to support this kind of learning, we have elsewhere described twelve design structures that developers can use to establish the four modes of engagement. The space should contain both real and fictional elements, all designed to embed the player and the concept within a context not typically available to the learner based on constraints of time, funding, or reality.

An essential aspect of conceptual play is that the individual is experientially situated within the play space in which she has a legitimate role. For example, a QA student becomes a scientist, examining the water quality of the green, murky water in a virtual river. Similarly, such spaces serve to situate disciplinary content. For example, the student-scientist must make use of their understanding of the related formalisms (i.e., algae blooms and eutrophication) to correctly characterize the problem and advance a successful solution. In this way, engaging in conceptual play allows for an extension of oneself into a context-of-use, establishing a new world and self—a virtual-real being (Gee, 2003). This being is nested in a virtual-real world wherein the disciplinary concepts have meaning and the learner has impact in that world. To summarize, our work focuses on using a narrative storyline and a virtual environment to *narratize* and make situationally concrete the *use value* of the to-be-learned concepts. Our theory of action states that such participation gives rise to legitimacy (established by conceptual engagement), intentionality (established by immersive engagement), consequentiality (established by transactive engagement), and accountability (established by reflexive engagement). This nascent theory of situated learning of academic content in conceptual play spaces has been deemed *situative embodiment* (Barab et al. 2007).

## Assessment for "Learning to Be" in Educational Videogames

Another reason commercial videogames support massive amounts of learning is that they present players with repeated assessments of their developing knowledge and skills. Importantly these assessments occur at multiple levels: at the more immediate levels, players are continually assessed by the game and other players regarding their enactment of specific game activities; at the more distal levels, players have opportunities to attain status among the players that takes months or years of success to attain. We contend that such assessment of developing academic knowledge is crucial for educational videogames. Thus, the second design principle examined in this study that *educational videogames should offer multiple forms and levels of meaningful assessment.* After the conceptual play space had been established in the initial design cycles, the research team set about refining the informal assessment practices in the space, along with more formal assessment practices used to refine and document learning in that space

Closely related to assessment is feedback. Assessment in videogames supports feedback that players use to enhance the success of their participation and learning. Some of this feedback is structured by the game, including the relatively immediate hints games give player who are lost or confused, while some is afforded by the game design, such as when a successful move (or mistake) is recognized and acknowledged by another player. This design study explored how feedback in educational games might be transformed to support learning of academic content. Thanks to widely cited reviews such as the one Black and Wiliam (1998), many have recognized the tremendous potential of formative feedback for supporting academic learning. However, feedback only supports learning and improvement if that feedback is useful *and* used to advance learning and instruction (Fredericksen and Collins, 1989). Unfortunately there is overwhelming evidence that (1) many teachers struggle to provide useful feedback, (2) much of the feedback that is provided goes unused, and (3) some feedback is used in ways that undermines learning and educational improvement (National Research Council, 2001a; 2001b). Hence, the third educational game design principle is the *educational videogames should offer useful feedback that is used.*

Arguably, assessments and feedback often undermine learning in educational environments. In particular, educators are increasingly using sophisticated computer-based systems to directly increase scores on external tests. Such program train students to recognize the specific associations that might appear on a targeted test. Because such practices deliver modest test gains, school computer labs are increasingly being dedicated to such practices, to the exclusion of innovations such as Quest Atlantis. This study illustrated how newer design-based research methods (e.g., Barab, 2006), newer multi-level approaches to assessment (e.g., Hickey et al., 2006), and innovative approaches to formative feedback (Hickey, Kindfield, Horwitz, & Christie, 2003) can be used to directly enhance curricular innovations like Taiga while indirectly but consistently increasing achievement on targeted content standards. Across the three annual design studies, learning was aligned across four increasingly formal assessment levels (*immediate, close, proximal,* and *distal*). These refinements aimed to (a) foster enhanced participation in socio-scientific inquiry, (b) enhance understanding of targeted scientific formalisms, and (c) increase achievement on externally developed achievement tests aligned to the targeted standards but independent of the curriculum.

At the first assessment level, "event-oriented" observations using qualitative methods were carried out around the enactment of the Taiga curricular activities. These include informal observations used to provide feedback to shape those enactments in situ, as well as more formal observations used to support and build nascent theories of participation and learning in this type of educational gaming environment. Rather than focusing on individual understanding, this "immediate-level" assessment focuses on collective participation in domain specific knowledge practices. These analyses examined how the enactment of QA activities afforded students the shared opportunities to embody the role of water quality scientists. Put differently we, our intention was for them to "try on" the identities of water quality scientists, and "try-out" the discourses and practices that define that scientific community. These analyses were conducted with an eye towards the way that many commercial game environments focus as much on learning how to "be" a particular professional as they do on how to "do" the range of skills and practices associated with a particular profession. In this sense, successful videogames provide players with ongoing opportunities to succeed (and sometimes fail) at what Gee (2005) characterized as "distributed authentic professionalism". The Taiga curriculum was designed to provide the players and their avatars distributed experiences with the authentic skills of professionals, and their authentic value systems and identities. In this sense, the research team assessed whether the design of the Taiga activities engaged students with real world problems where perspective on a particular way of knowing determines what critical knowledge is lacking. It examined the functional relations between what is significant and how one can come to know what is significant, mutually reinforcing what Scardamalia and Berieter (1994) deemed "intentional learning". Because the results of this assessment of participation are so tightly bound to the Taiga curriculum, the perspective used and the nascent theories of participation that emerged that are more broadly relevant than the actual results of the analyses.

Complementing the notion of situative embodiment detailed above, these analyses assessed how the conceptual play space of the Taiga activities supported transitional forms of participation (Beach, 1999). These transitions were framed in terms of liminality (Turner, 1967), which examines experiences in-between recognized social boundaries. Liminal transitions are activities seeded along an unfolding pathway through game play. The structure of each activity engages learners with ways of doing science and being scientific that are different from but complementary to curricular experiences. These activities simultaneously engage students in the work of transforming scientific practices and evolving their scientific roles. Designing for liminal transitions, then, represents a principled approach to assessment of "learning to be" that coordinates the QA backstory, the Taiga narrative, a player's networked assets (e.g., in-game photos, maps, data, books, and so forth), and evolving status (e.g., ranger internship, lab credential, peer and Atlantian advisor), in order to expand ongoing participation with authentic ecological sciences professionalism (elaborated in Zuiker, 2008).

As an example of design based research, liminal transitions illuminated design tensions for understanding the general consequences of learning and evolved a local theory of assessment for learning to be. Videotaped enactments of Taiga were analyzed to understand ways that this interleaving series of activities both support learning and make knowing visible. A concurrent, nested approach to data analysis featured interpretive analyses of learners' trajectories of participation with the curriculum as well as the transformations arranged by liminal transitions. The analysis revealed a range of specific episodes in which some (but not all) students recast relevant practices and roles in light of the situations they encountered during some (but not all) liminal transitions. This highlighted the importance of refining liminal transitions to help balance situational complexity and social collaboration during relatively brief curricular episodes. This guided the enactment of existing activities to provoke students to wrestle with learning as a way of being scientific (not just a process of coming to know about science). These insights contributed to the ongoing refinement of those activities and Taiga curriculum, and provided useful insights in the design of other academic play spaces in the QA world.

## Feedback in Support of Participation and Understanding

Work at the next assessment level concerned the four quests that students completed in the Taiga curriculum. Each quest was framed as a report that that the field investigators were submitting to Ranger Bartle for his review and approval. In reality, the quests were reviewed by the teacher who "played" the Ranger character. In more conventional curricular terms, they were somewhat akin to the "semi-formal" quizzes that teachers use to assess learning from the enactment of classroom activities. Such "close-level" classroom assessments afford feedback that is ideal for directly advancing student understanding and guiding the enactment of the associated curricular activities. A major focus of last design studies was refining the usefulness of the feedback that students received around the quest submissions, searching for ways to ensure that students used that feedback, and examining the impact of that feedback on individual understanding and achievement.

While the Taiga includes four quests, it is Quest 2 (Q2) that is crucial to their learning of the key ecological science concepts, which are in turn used in the socio-scientific inquiry in Quests 3 and 4. In the initial implementations, students were provided relatively modest feedback on their Q2 submissions. The responses were evaluated by a research team member, and the feedback generally encouraged students to elaborate. Given the centrality of Q2 and the well-documented potential of formative feedback, enhancement of the Q2 feedback was an obvious target for further refinement of the Taiga curriculum. The first refinement to the Q2 formative feedback was a new 3-point scoring rubric for teachers to use when evaluating their students' submissions. The rubric assigned 0-3 points based on the students' apparent understanding of water quality indicators, processes, and their interrelationship. The rubric included specific text which the teacher was invited to cut and paste into the feedback window when returning the submitted quests to the students. This rubric alone was expected to improve learning by focusing the teacher's (and therefore the students') attention on the key content of the Taiga curriculum.

The second refinement to Q2 consisted of feedback from the Lab Technician, a scientist NPC in charge of analyzing water samples for the Taiga park administration. Teachers were expected to reject most initial Q2 submissions; the feedback on the scoring rubric for Quests assigned 0-2 points advised students to visit the technician for help in revising their submission. The Technician then invited students to review 21 screens of fairly dense technical information regarding the three key Q2 elements (indicators, processes, and their relationship). At three decision points, students were asked if they wanted to see the specific feedback and wanted to learn more about each. The feedback dialogue pages used the relatively dense, technical jargon of the domain. In light of the theorizing around the first design principle, we aimed to support meaningful "embodied" participation around "embedded" scientific content. This highlighted the balancing act between making feedback *useful* (for advancing student understanding as much as possible) and ensuring it is *used* (by making it accessible and directly relevant to the task at hand). Our general design strategy is that before simplifying the feedback in order to make it more accessible and readable (and potentially less useful for maximizing participation and individual learning), we should first search for ways to maximize use of that feedback.

## Learning Measures and Research Design

At the third assessment level, a more conventional assessment was used to capture student learning from the Taiga curriculum (including the embedded assessment and formative feedback) and guide its refinement. This "proximal" assessment was intended to assess individual gains in understanding of the scientific concepts and forms of socio-scientific inquiry taught in the Taiga curriculum. Such assessments are akin to the formal classroom assessments that teachers typically use at the end of a unit or semester (often for assigning grades). They are "curriculum-oriented" in that they assess the concepts taught in particular curriculum, but in a somewhat different context. Proximal-level assessments have more modest value for directly advancing student understanding (because they are removed in time and context), but are quite useful for (a) refining the curriculum, (b) comparing versions of a curriculum, and (c) guiding formal remediation for specific topics and/or students. The project tried out a number of different formats, including open-ended essay items, short answer items, and multiple choice items that had been constructed or carefully selected. In the final implementations, the distal assessment consisted of a comprehensive set of 15 open-ended items concerning another fictional watershed system (*The Lee River*).

At the fourth assessment level was a traditional multiple-choice test that measured student achievement on the science content standards that were aligned to the Taiga curriculum. The test was akin to the externally-developed criterion referenced achievement tests. It was "standards-oriented" in that it was aligned to the standards that were aligned with the Taiga curriculum, but were otherwise unrelated to this particular curriculum. Such "distal" tests have little value for directly advancing student understanding, but are useful for comparing different curricula that target the same standards, and documenting the impact of a particular curriculum on other similarly-constructed tests. To make the test, the research team first created pools of released achievement test items that were aligned to the targeted content standards, independent of the way that standard was addressed in Taiga. After sorting each pool of items in terms of apparent difficulty, items were randomly selected from each pool to create a test of about 20 items. Twelve standards were targeted in Year 1, and four standards in Years 2 and 3.

Two teachers implemented the Taiga curriculum across three annual implementation cycles. Both teachers were in the same district that served relatively high-achieving students in a town that surrounded a university in the Midwestern US. The majority of the students (roughly 80%) were Anglo American. Few of the students (roughly 10%) qualified for free or reduced lunch. The first teacher had been working closely with the Quest Atlantis project from the start, and used Taiga in a single class of academically gifted fourth-graders in each of the three implementation years. The second teacher was new to Quest Atlantis, and taught science to four classes of sixth-grade students. During the second implementation cycle, he used Taiga in two of his classes, and used a high-quality custom-made textbook in his other classes. While the custom text covered the exact same science content as Taiga, it did so using more directly expository presentation of that content, much of it appropriated from other widely used textbooks. This was necessary because no other textbooks include such a focused consideration of water quality and socioscientific inquiry, which would have precluded a valid comparison.

## Increased Gains in Understanding and Achievement Across Years

Figures 1 and 2 summarize the gains on the proximal and distal outcome measures for the two teachers across the implementation cycles. To reiterate, the curricular refinement described in Barab et al (2007, summarized above) were carried out between the first and second implementation cycles; the virtual formative feedback routines (and other more modest refinements) were carried out between the second and third cycles. For comparability across years and measures, pre-post gains have been transformed into effect sizes using the pooled standard deviation. (4) The statistical likelihood of each gains using repeated measures analysis of variance is indicated on the graphs ($p < .05$ = *, $p < .01$ = **, $p < .001$ = ***).

In the first implementation year, the fourth-grade teacher implemented the Taiga curriculum for the first time. In this implementation, the proximal assessment consisted of two open-ended essay items that asked students quite directly about socio-scientific inquiry in the context of water quality. They were scored by two research team members who assigned up to six points on each essay for the presence of specific elements, and who attained adequate inter-rater reliability (i.e., < .80). While the proximal gains were quite substantial (over 1.5 SD), gains on the distal test (which targeted 12 standards) were quite trivial and may have occurred by chance. Following the first round of curricular refinements, gains on the new proximal assessment (with a range of items) declined some. This was not surprising as the new proximal assessment was less aligned to the Taiga curriculum. Encouragingly, gains on the distal test were substantially larger, and were (for the first time for valid distal measures) statistically unlikely to have occurred by chance. In the third year, following the inclusion of the virtual formative feedback described above, gains on the proximal assessment (the comprehensive *Lee River* assessment) were over two standard deviations. Average gains on the distal test in the third implementation were slightly smaller than year two, but were more consistent across students and therefore even less likely to have occurred by chance.

Figure 2 shows that the sixth grade teacher obtained statistically unlikely gains on both the proximal and distal measures in both conditions in Year Two. While the Taiga gains were substantially larger, the differences between the gains in the two conditions were themselves not statistically unlikely. The subsequent incorporation of the virtual formative feedback after Year 2 resulted in dramatically larger gains on the proximal test and modestly larger gains on the distal test. Like the fourth grade teacher, the sixth grade teacher showed the intended pattern of increasingly larger gains on the proximal assessment, leading to an "echo" of increased gains on the distal test. These finding show that using proximal assessments to guide implementation and refinements of game-based curriculum and embedded formative assessment practices can deliver gains on distal standards-oriented achievement tests that are independent of the particular curriculum. Because the students were never directly instructed on the associations included in the distal test, and because such tests have proven quite difficult to impact with short term curricular interventions, we contend that this is convincing evidence of the broad value of the Taiga curriculum.

### The Relationship between Engagement in Feedback and Learning Outcomes

In order to examine the impact of virtual formative feedback, additional post hoc analyses were carried out around the Quest Two submissions and student engagement in the virtual formative feedback routine. These analyses were conducted on the Year Three results for the sixth-grade teacher. Each of the initial and final (i.e., revised) Quest Two submissions were scored by two members of the research team on a detailed 5-point scale (inter-rater reliability = .85). Of the 79 students were asked to resubmit their quests, scores increased an average of 1.7 points on the five point scale, representing a gain of 1.5 SD ($F = 110$, $p < .0001$). Clearly, the revision process was resulting in improved submission quality.

The crucial question concerns the extent to which the improvement in submission quality was due the virtual formative feedback. Examination of feedback log files revealed that 18 of the 79 students did not even access the feedback, 17 quite minimally accessed the three decision screens while electing to not examine the detailed content, 16 students examined 1 of the 3 sections, 15 examined 2 of the 3 sections, and 13 examined everything in all 3 sections. As shown in Figure 3, the gains on the proximal assessment showed a clear trend, with the students who accessed all three sections gaining double what the students who did not visit the technician at all. Quite remarkably, as shown in Figure 4, the students who accessed everything on all three feedback sections were responsible for most of the entire group's gain on the distal test. Obviously these are correlation data, and so no claims of causality are supported. However, they suggest that the obvious next step is identifying the factors (e.g. extrinsic incentives) that support or undermine student engagement in formative feedback to support even greater gains in understanding and achievement. Indeed, the fourth design principle that we are exploring in subsequent studies is that *educational games should provide students incentives for using formative feedback.*

More broadly, the study shows that the general model of using close-level feedback to maximize individual participation and learning, while using proximal level feedback to guide refinements of the learning environment, while using distal-level data as feedback about the success at raising achievement, is a viable model for doing so. These results have spurred the research team to plan subsequent implementations of the Taiga curriculum in which attempt to maximize engagement in feedback using different configurations of the feedback routine (e.g., more or less embedded in the game) and motivation practices (public recognition of expertise according to Quest scores).

## Endnotes

(1) www.digitallearning.macfound.org
(2) www.elektra-project.org/
(3) www.questatlantis.org
(4) Most considerations of effect sizes follow Cohen's (1992) characterization of effect sizes of .3, .5, and .8 as *small, medium,* and *large.*

## References

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

Barab, S. A. (in press). Narratizing disciplines and discliplinizing narratives: Games as 21st century curriculum. In S. Barab and A. Ingram-Goble (Eds), *Games as 21st century curriculum.* Dordrecht, Netherlands: Springer

Barab, S. A. (2006). Design-based research: A methodological toolkit. In K. Sawyer (Ed.), *The handbook of the learning sciences* (pp. 153-170). Cambridge, MA: Cambridge University Press.
Beach, K. (1999). Consequential transitions: A sociocultural expedition beyond transfer in education. *Review of Research in Education, 24*, 101-139.
Bereiter, C. & Scardamalia, M. (1989). Intentional learning as a goal of instruction. In L. Resnick (Ed.), *Cognition and instruction: Issues and agendas* (pp. 361-379). Hillsdale, NJ: Erlbaum.
Black, P. & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education, 5*(1), 7-74.
Cohen, J. (1992). Quantitative methods in psychology: A power primer. *Psychological Bulletin 112*(1): 155.
Federation of American Scientists (2006). *Harnessing the power of video games for learning.* Washington, DC: Author.
Frederiksen, J. R., & Collins, A. (1989). A systems approach to educational testing. *Educational Researcher 18*(9), 27-32.
Gee, J. P. (2003). What video games have to teach us about learning and literacy (1st ed.). New York; Houndmills, England: Palgrave Macmillan.
Gee, J. P. (2005). What would a state of the art instructional video game look like? [Electronic Version]. *Innovate*, *1*(6). Retrieved November 16, 2007 from http://innovateonline.info/index.php?view=article&id=80.
Glenberg, A. M. (1997). What memory is for. *Behavioral and Brain Sciences 2,* 1-55.
Hickey, D. T., Kindfield, A. C. H., Horwitz, P., & Christie, M. A. (2003). Integrating curriculum, instruction, assessment, and evaluation in a technology-supported genetics environment*. American Educational Research Journal*, *40* (2) 495-538.
Hickey, D. T., Zuiker, S. J., Taasobshirazi, G., & Schafer, N. J., & Michael, M. A., (2006). Three is the magic number: A design-based framework for balancing formative and summative functions of assessment. *Studies in Educational Evaluation, 32* (180-201).
National Research Council (2001a). *Knowing what students know: The science and design of educational assessment.* J. W. Pellegrino, N. Chudowski, & R. W. Glaser (Eds.). Washington, DC: National Academy Press.
National Research Council (2001b). *Classroom assessment and the national science education standards.* J. M. Atkin., P. Black, & J. Coffey, (Eds.). Washington, DC: National Academey Press.
Sadler, T. D. (2004). Informal reasoning regarding socioscientific issues: A critical review of the literature. *Journal of Research in Science Teaching, 41,* 513-536.
Scardamalia, M., & Bereiter, C. (1994). Computer support for knowledge-building communities. *The Journal of the Learning Sciences, 3*(3), 265-299.
Shaffer, D. W. (2006). *How computer games help children learn*. New York: Palgrave Macmillan.
Shaffer, D. W., Squire, K. A., Halverson, R., & Gee, J. P. (2005). *Video games and the future of learning.* Phi Delta Kappan, 87(2), 104-111.
Squire, K. (2006). From content to context: Video games as designed experience. *Educational Researcher, 35* (8), 19-29.
Turner, V. (1969). *The ritual process: Structure and anti-structure*. Chicago: Aldine.
Zuiker, S. J. (2008, March). *Transforming practice: Designing for liminal transitions along trajectories of participation*. Paper presented at the annual convention of the American Educational Research Association, New York, NY.

## Acknowledgements

This research was supported by a grant from the US National Science Foundation to Indiana University (Grant REC 0411846). We gratefully acknowledge the many other individuals who contributed to this research effort. These include Eun Ju Kwon, Anna Barab, Ina Kouper, Scott Warren, and Tyler Dodge. Anna Barab led the Year Two grade six implementation effort, including the development of the comparison curriculum. We are especially grateful to teachers Elizabeth Piekarsky and Jake Summers for implementing Taiga, and to their students for their hard work and patience during in the data collection.

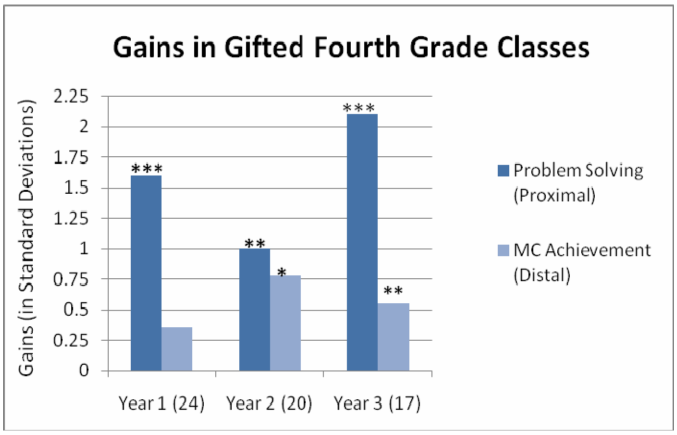


Figure 1. Gains Across Years in Fourth Grade Classes

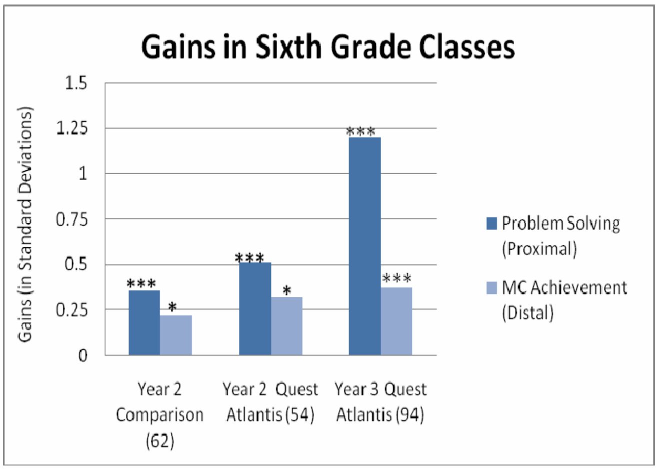


Figure 2. Gains Across Years in Sixth Grade Classes

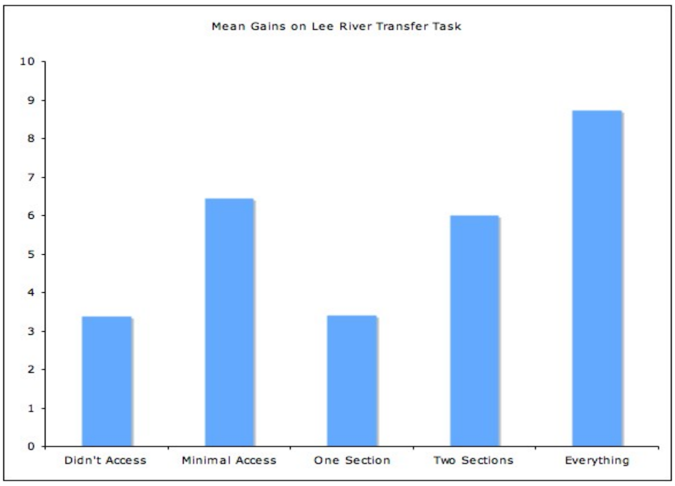


Figure 3: Differences in Proximal Gains According to Amount of Feedback Accessed

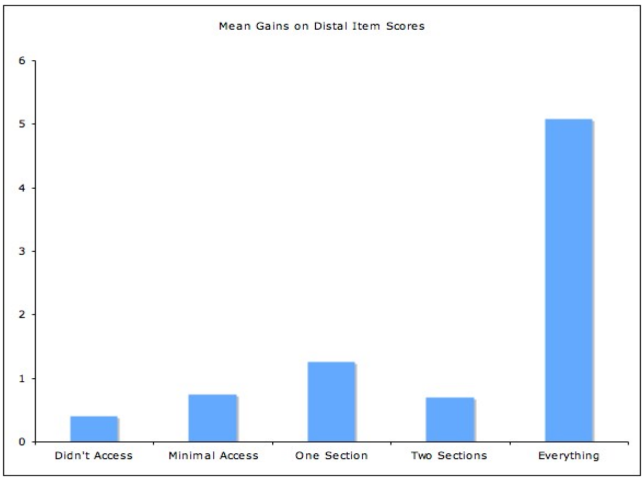


Figure 4. Differences in Distal Gains According To Amount of Feedback Accessed