JRTE, *41*(1), 113–140

A MUVE Towards PBL Writing: Effects of a Digital Learning Environment Designed To Improve Elementary Student Writing

Scott J. Warren and Mary Jo Dondlinger
University of North Texas

Sasha A. Barab
Indiana University

Abstract

Two major obstacles to using problem-based learning methods with writing in elementary school classrooms are the time it takes to design the learning environment and the time required for students to interact at their own pace with ill-structured problems used to spur student writing. This study examined whether game elements could be used along with Problem Based Learning (PBL) in a digital learning environment to improve student writing. Results from this study included statistically significant decreases in teacher time spent answering procedural and directional questions, increased voluntary student writing, and improved standardized achievement scores on writing tasks. (Keywords: achievement, elementary, MUVE, game, writing.)

Leveraging the Games Children Play

Many claims have been made about the effectiveness of instructional media and software on student learning (Clark, 1991; Kozma, 1991). Further, some theorists in the field of education have begun to look towards the power of video games and other digital learning environments to improve student learning (Gee, 2003; Jenkins, Squire, & Tan, 2003; Prensky, 2001; Squire & Steinkuehler, 2005; Steinkuehler, 2004). However, the research in this area that exists is still in a nascent phase with limited findings from studies that address changes in student achievement in content areas (Dondlinger, 2007). We still do not know if the preparation and use of a digital video game learning environment intended to aid learning correlates with improved student reading and writing skills, mathematical reasoning ability, or any other academic activity that is measured by and is at the heart of the accountability movement in the United States. While we know that off-the-shelf video games like *World of Warcraft* and *Elder Scrolls IV: Oblivion* are engaging best sellers, we do not know if learning games can be designed that are equally engaging while still providing learning gains that match educational standards. However, researchers are beginning to explore these boundaries (Dickey, 2007; Squire, 2006).

Successful teachers have long co-opted the existing interests and activities of their students into their curricular materials and instructional practices. In some

instances that may have involved encouraging a young learner interested mainly in football to read an autobiography of Joe Montana as a book report choice. In others, teachers may provide optional topics for a required essay such as skateboarding, cheerleading, and favorite toys. Currently, playing video games is one of the more popular activities engaging children in their free time with a reported 35% of the most frequent players being under the age of 18 (Entertainment Software Association, 2007). While academic motivations have been shown to decline, especially during the transition from elementary to middle school (Anderman, 1996), video game usage among all age groups has been steadily increasing for the last decade with one recent study suggesting that one in five gamers are individuals over age 30 (Entertainment Software Association, 2007).

Over the course of the last two decades, student interest in video games has rapidly increased in the United States and throughout the world, leading to record sales that have nearly outstripped Hollywood movies and spurred game driven economies that have real-world links and consequences. Software developers have tried, sometimes successfully, to leverage this interest into profits by creating "edutainment" titles such as the Civilization series, Math Blaster, Oregon Trail, and others that have shown links to learning when coupled with other forms of instruction such as guided reflection and group discussion (Dede, Ketelhut, & Ruess, 2006; Squire, 2005). However, some of these links are tenuous and poorly researched, and many of the games include an impoverished narrative and uninteresting rule structures that fail to fully engage many learners. The research into the promise of video games as a means to reengage students with learning is still largely unexplored. Formal studies are needed to determine the potential value of problem-driven digital learning environments that include game-like affordances such as embedded scaffolds, nested goals, clue resources, narrative context, and explicit rules (Crawford, 2003; Salen & Zimmerman, 2004). Measuring student time spent performing on-task activities, calculating instructor time expended answering procedural questions, and examining whether students are willing to complete voluntary activities with learning components are quantifiable ways to gain a better understanding of whether games or game-like environments can engage students more actively than traditional, instructor-led teaching methods.

LITERATURE REVIEW

Games, Learning, and Research

Video games, simulations, and those that sit in the crux between the two are already being leveraged to impact learning in many spheres ranging from adult learners to students in K–12 settings, and more research is under way to validate their use in a number of spheres including business, academia, and the military. One large movement in higher education called Serious Games seeks to develop learning environments that leverage existing games, build new games and simulations, build theory about the use of game principles in education, or simply to study the work of game designers as they work to improve public

education at all levels. This work has been led through publications by James Paul Gee's (2003) work related to what children learn about their own learning and about themselves through play with off-the-shelf games, Clark Aldrich (2003) on the use of simulations in education, Justine Cassell and Henry Jenkins (2000) regarding the importance of video games in popular culture and in children's lives, Mark Prensky (2001) focusing on the use of computer games for learning, Constance Steinkuehler's (2004) focus on the importance of player literacy practices in games as they relate to learning in massively multiplayer online role playing games (MMORPG) such as Lineage, Star Wars Galaxies, and more recently Dickey's (2007) analysis of the potential uses of World of Warcraft. While efforts are currently under way to empirically study the use of video games by learners at all levels, much of the work that has been done to this point has either been through case study, anecdote, or qualitative analysis (Dondlinger, 2007).

The Motivating Power of Games

Several publications examine motivation in video games; however, not all researchers entirely agree on the source of this motivation. Some attribute the compelling nature of games to their narrative context (Dickey, 2005, 2006; Fisch, 2005; Waraich, 2004), others find that motivation is linked to goals and rewards within the game itself or intrinsic to the act of playing (Amory, Naicker, Vincent, & Adams, 1999; Denis & Jouvelot, 2005; Jennings, 2001). Nevertheless, all find that motivation to play is a significant characteristic of educational video games and that effective game design considers both intrinsic and extrinsic rewards for play. Denis and Jouvelot (2005) distinguish between the two and their absence as follows:

Intrinsic motivation pushes us to act freely, on our own, for the sake of it; extrinsic motivation pulls us to act due to factors that are external to the activity itself, like reward or threat; amotivation denotes the absence of motivation. (p. 462)

These authors see motivation as the interplay between desire and pleasure—the desire to be competent and the pleasure one feels when one is. They argue that competence, autonomy, and relatedness are factors that affect motivation. "Motivation also leads to the activation of efficient cognitive strategies for long-term memory issues like monitoring, elaborating or organizing information. On the opposite side, resignation and amotivation have negative results on memorization and personal development" (p. 463).

Dickey (2006) argues that a narrative context that promotes "challenge, fantasy, and curiosity" and that provides feedback for players is one that promotes intrinsic motivation for play (p. 2). She also finds that "Strategies of design that lead to engagement may include role-playing, narrative arcs, challenges, and interactive choices within the game as well as interaction with other players" (p. 1). In another study, Waraich (2004) agrees that narrative is essential to motivation but cautions that "intrinsic rewards are based on a high congruence between the material being taught and the motivational techniques used" (p. 98). Dissonance between the two can decrease learning.

Game Design and Learning Theory

Given the compelling nature of commercially produced games, researchers have deployed the following approaches to implementing digital games in formal learning contexts: automating drill and practice through digital games for learning, adapting off-the-shelf games to formal learning contexts, or designing games for specific curricular objectives and audiences.

Drill and practice games. A principal advantage of digital games and simulations is that they allow for repeated practice with nearly instantaneous feedback. This capability frees up teacher time spent manually assessing performance while allowing learners to test various strategies, modify actions, and practice different approaches (Dickey, 2007; Gee, 2003). One study found that this automation allowed young learners to practice more math facts problems, increasing both their speed and accuracy (Lee, Luchini, Michael, Norris, & Soloway, 2004).

Digital games and simulations also allow experimentation and practice that is free from many of the hazards of real life. Used for both child and adult learning, the free, downloadable America's Army "first-person shooter" simulation-game allows soldiers to be trained in a safe, virtual environment where their actions do not have the severe consequences of real battle or the cost of outdoor war-game simulations with real guns, rubber bullets, and smoke grenades (Nieborg, 2005). However, studies on feedback in this game have found it to have a large impact on how well the user performed simulated actions. Constantly negative feedback may impact the users' sense of self-efficacy, making them less apt to perform well in each instance of practice or interaction with the simulation (Kaplan, 2003). Conversely, if feedback is continually positive, users may develop overconfidence in their abilities, leading to carelessness in interacting with the simulation and less susceptibility to corrective feedback. For example, soldiers fighting currently in the Iraq conflict who had been trained using the America's Army digital simulation were found to have developed specific behaviors based on the feedback they received in the game (Kaplan, 2003). Namely, if they hid behind certain objects, they could jump out and kill opponents. When it came time to translate their simulated experiences into real world experiences, the simulation had not prepared them for the reality that bullets pass through wood crates or that opponents do not react in predictable ways. Nevertheless, regression analysis on data generated by online players and soldiers at Fort Leavenworth show this game-simulation to be effective at imparting knowledge and skills about tactics related to the practice of fighting a battle (Schneider, Carley, & Moon, 2005).

Off-the-shelf games. Using off-the-shelf video games that have a learning component has been one approach to using games that not only appears to improve student learning of subject matter, but also affects the ways learners process content and reflect on their own learning. One such attempt has been Squire, Giovanetto, Devane, and Durga's (2005) use of the video game Civilization III, a turn-based strategy game-simulation (RTS) that allows students to take command of a civilization that existed at some time in history. Using interviews and surveys, this group's work found that participation in game play

(a.) immerses students in historical terminology and reinforces their knowledge of existing terms, (b.) improves student interest in the content of history, (c.) encourages understanding of the game itself as a form of historical simulation, and (d.) provides a scaffold for thinking about the historical concepts and content encountered in contexts outside of the game-simulation itself, implying transfer of learning from the close context to more distal ones (Squire, Giovanetto, Devane, & Durga, 2005). Similar to the findings regarding collaboration in learning environments noted by Linn, Clark, & Slotta (2003) and Samsonov, Pedersen, and Hill (2006), Squire et al. (2005) found that students who were successful at completing game objectives tended to work with other students and shared their experiences often. This discourse among students functioned as a means of metacognitive reflection, in which students reflected upon their personal experiences with the curricular tasks as a means to better understand their own processes for learning, came to terms with cognitive difficulties encountered during the learning activities, and compared their own experiences with peers as a means of improving their future learning experiences both with the program and in other learning contexts with comparable tasks. As a whole, this research indicates that game-based simulations can be effective for encouraging student collaboration, increasing expertise in a skill or strategy based system, and overcoming failure or frustration through cognitive or metacognitive reflection, leading students to devise new strategies that may have not been apparent from the outset.

Curriculum-driven game designs. Games also provide immersive environments in which students can collaborate to solve ill-structured problems. Such designs immerse students in an unfolding narrative and life-like context that lends authenticity to their learning experiences. With these ideas in mind, designers of the Taiga learning environment developed just such an immersive world to accompany fourth grade science curriculum. Playing the role of fledgling scientists, learners are specifically asked to develop a hypothesis that explains the mysterious death of large numbers of fish in a national park. The goals of this unit included: (a) encountering new concepts such as erosion, eutrophication, water quality, and system dynamics and (b) improving student analytical skills through graph deconstruction, hypothesis generation and revision, simulated water analysis, socio-scientific reasoning, and scientific inquiry.

The design of this environment evolved through multiple iterations over a two-year period, based on qualitative and quantitative research findings collected during each iterative implementation of the treatment (Barab & Squire, 2004; Barab et al., In Press). The results of the first iteration of the Taiga treatment design showed a statistically significant increase in pre-post learning gains using standardized test items that were close, or "proximal" (Hickey & Pellegrino, 2005), to the content used in the curricular activity ($F (1, 23) = 39.73$, $p < .001$) (Barab, Sadler, Heiselt, Hickey, & Zuiker, 2006). However, a repeated measures analysis of variance on distal items, which are defined as those in a different context and with different content, presented non-significant gains [$F (1, 23) = 2.57$, $p = .122$]. Consequently, the designer-researchers embedded additional opportunities for students to encounter the key underlying formal-

isms in their more abstracted forms in a second iteration of the game (Barab et al., In Press). These opportunities included scientific diagrams that learners encountered by chance, coupled with a virtual computer that included more formal, non-contextualized descriptions of learning content, a series of game-like interactions that forced students to decode data with help from non-player characters (NPCs) that offered aid upon student request. The findings from this second study revealed significant learning gains on both proximal ($F(1,19) = 16.77$, $p < .01$) and distal items ($F(1, 19) = 9.03$, $p < .01$), supporting the conclusion that when learning in an immersive narrative context, rich with embedded scaffolds and resources, students developed understandings about the underlying content formalisms and also began to appreciate the relationship between their own experiences and how new knowledge could be transferred to other, distal contexts.

Games for Writing

While video games as a form of technology have shown some improvements in learning mathematics, science, and battle applications, little research into the use of games as a support for learning to write or as a means of allowing students to practice their skills has been conducted. Too much of the application of technology in writing instruction has been relegated to the use of the word-processor for student writing. Sadly, despite the overwhelming investment in access to technology for learning, computers in the classroom do not get used for much more (Cuban, Kirkpatrick, & Peck, 2001). Although the word-processor has done much to enhance writing performance as it relates to outcome achievement, it does little to enhance writing instruction, provide feedback, or encourage reflection. Nevertheless, constructivist and problem-based approaches to writing have been used to improve general student literacy skills using decade-old technologies such as instant messengers, e-mail pen pal programs, and hypertext embedded in Web pages (Egbert & Hanson-Smith, 1999; Englert, Manalo, & Zhao, 2004). With the visual, audio, and rapid feedback affordances of immersive learning environments, it is possible that they could perform these functions and more, addressing issues such as learner motivation (Dede, Ketelhut, & Ruess, 2006; Tuzun, 2004) and the inclusion of heavily scaffolded activities to supplement the instructor (Barab et al., In Press).

The *Anytown* Multi-User Virtual Environment

The Anytown multi-user virtual environment was created using the Active Worlds browser that is the underlying digital system for the National Science Foundation's Quest Atlantis grant project (Barab, Thomas, Dodge, Carteaux, & Tuzun, 2005). While other towns elsewhere in Quest Atlantis are fantastic, otherworldly realms in which the architecture defies physics, employs "teleports" that move students rapidly from place to place, or simply have no analog on Earth, the design of the Anytown environment was intended to create a small town feeling in which the locations, people, and other objects would be mostly familiar to the majority of participating students. The design followed this plan in order to set student learning in an authentic environment with which they

already had some background knowledge, a design element advocated by Salen and Zimmerman (2004) as the "modeling reality characteristic" that helps to situate the learner in a modeled space not radically different from his or her own experience. This design consideration was expected to allow students to readily recognize the affordances of particular locations such as the general store, the school, and the library. Moreover, this MUVE was designed to facilitate writing instruction rather than scientific discovery. The overarching narrative context of Anytown situated learners in the role of cub reporters investigating a series of mysterious events: vandalism, a burning building, and strange lights emanating from the town's river.

Applying lessons learned from studies of America's Army and Taiga, Anytown designers incorporated the use of feedback within and from the digital system, embedded scaffolds in the form of character dialogue, as well as visual and textual clue resources used to drive learning activities. Students received textual cues by clicking on objects and characters, which provided them with information about their environment and the writing process, offered positive feedback related to their progress on learning tasks when appropriate, gave additional scaffolding for learning tasks when needed, and imparted directions to learning tasks as part of the rules of their overall experience. Feedback and interactivity was provided in colloquial text responses appropriate to each character's personality and role in the environment, which meant that students sometimes had to complete related tasks in order to elicit responses from particular people in a way similar to how people respond to each other in the real world. This was expected to be somewhat disconcerting to the young learners who were used to being provided with instant answers to their questions. However, making them earn answers was expected to induce cognitive conflict within students and compel them to think more critically about how to get information they could use to solve their problems and adequately respond to the writing tasks required as "solutions" to the ill-structured problems they investigated.

Further, the teacher in the classroom played the role of editor of the newspaper and provided both positive and negative feedback to student writing tasks via the Anytown system after each session. The purpose of this design feature was to balance instructor feedback on both learning and game tasks, provide students with the perception that they received evaluative comments from someone other than the teacher, and maintain the illusion of their roles as reporters in the context of a fictional town. Figure 1 shows Irene Morningstar of the Anytown School waiting to help learners with questions about grammar.

Each of these goals was intended to leverage technology in a way that supported students in a problem-based learning context and made the method less time-consuming for the instructor in terms of planning, development of resources, and directing learners in the classroom while increasing the level of important feedback for students on their writing.

Following the mandated curriculum for this age group, Anytown included six main learning tasks called Writing Tasks that students were required to complete during their engagement with the environment and that were expected to directly impact their learning. With this in mind, students were immersed in

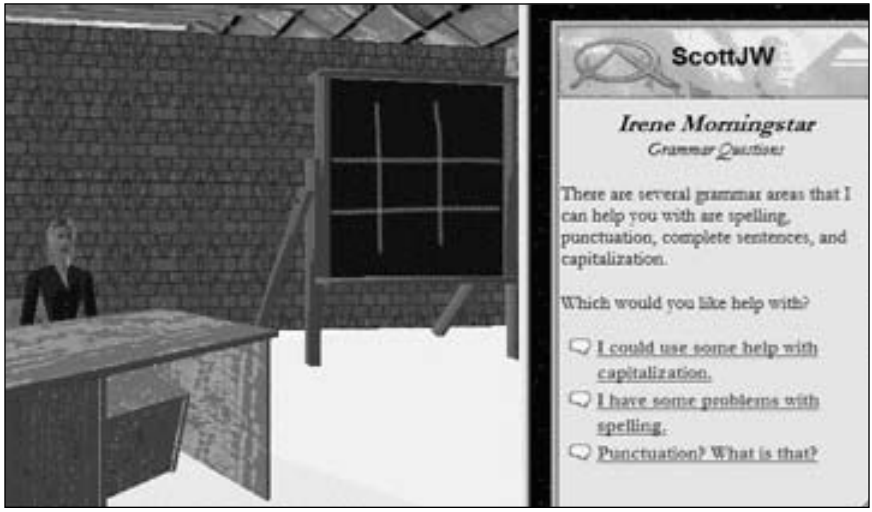

Figure 1: School teacher Irene Morningstar, an instructional support character in Anytown.

the terminology of writing through their interaction with the learning environment and its characters, engaging them with a fictional narrative tied to the meta-narrative of Quest Atlantis. Moreover, adopting the role of a newspaper reporter—a role intended to provide them with an understanding of a career that required writing and investigation—contextualized their writing activities in a meaningful way that was appropriate to their age-group.

Drawing from the research findings on metacognitive reflection presented in the Civilization III study (Squire, 2004) and the Taiga studies' results concerning knowledge transfer (Barab, Sadler, Heiselt, Hickey, & Zuiker, 2006; Barab et al., In Press), Anytown provided 22 possible non-required, free choice writing tasks: Reflection, Mystery, and Creative Writing Quests. These optional learning tasks were largely intended to engage students in higher order thinking skills such as problem solving, planning, and the use of creativity in order to overcome environmental difficulties. While the non-required tasks were more game-like and possibly more fun than the Writing Quests, students were required to manage their time so that they completed the required tasks successfully, in order to return to the more enjoyable non-required, but still educational, game tasks. This also put students into a position in which they had to take advantage of the information and discoveries of process made by themselves and peers in order to generate and develop justifications for their solutions to investigative Mystery tasks, make clear their reasoning for developing their solutions in written form, and engage in creative writing with the guidance of characters, built-in resources and their teacher. By using the chat, e-mail, and telegram functions that were part of the system itself, it was expected that students would be able to share their solutions with peers, test them for logic and defensibility, and reflect upon their experiences as a means of scaffolding for their peers who may struggle to complete a learning Task.

PURPOSES OF THE STUDY

The purposes of this study were to determine whether multi-user virtual environments that combine both strong instructional principles and basic game design principles can (a) reduce the amount of time spent by teachers answering redundant procedural and directional questions which are administrative in nature rather than educational (b) increase voluntary student writing practice which acts as an indicator of student motivation to learn and has been correlated with improvements in student writing generally, and (c) increase student writing achievement as measured by standardized writing assessments. Further, the purpose was to describe the differences between how instruction takes place in the designed learning environment when compared with instruction in a more traditional learning context.

The hypotheses addressed by this study included:

1. The amount of time that the teacher spends answering procedural and directional questions regarding the assigned and optional writing tasks in the treatment condition should be, at a statistically significant level, less than the amount of time spent by the teacher providing instruction in a face-to-face, traditional classroom.
2. The number of non-required writing activities completed by students in the treatment condition should be, at a statistically significant level, greater than the number of non-required writing activities completed for writing practice by students in a face-to-face, traditional classroom writing unit that includes the same objectives.
3. The quality of student descriptive writing achievement in the treatment condition should be, at a statistically significant level, greater than the descriptive writing achievement of students who receive instruction in a face-to-face, traditional classroom writing unit that includes the same objectives.

METHOD

This study examined the researcher-designed Anytown multi-user virtual environment in a naturalistic, classroom context. Employing a quasi-experimental, pretest-posttest comparison design, the study measured the effect of a curriculum-based, 3D learning environment on student standardized writing achievement. The design is quasi-experimental because students were randomly assigned by the school to one of the two classes that comprised the treatment and comparison groups (Gall, Borg, & Gall, 1996). The pre and posttest measures were counter-balanced by splitting the two classes and randomly assigning students to one of two writing prompts by drawing names from a hat. Whichever prompt a student did not complete for the pretest, he or she competed as a posttest. Each test was a standards-based assessment selected from released prompts by the California Achievement Program and New Jersey Assessment of Skills and Knowledge, both of which were aligned to the targeted content standards.

The independent variable in this design is the type of instruction (Anytown Language Arts Unit or Reading Curricular Unit) and the dependent variables

are student achievement on a post-test writing activity taken from released state standardized examinations, their submitted work for the instructional unit, and the amount of teacher time spent answering directional or procedural questions. The validity and reliability of the writing assessments were already established by either their respective states for standardized testing in those states and were appropriate for the age group.

Participants and Setting

Two settings were used for research in this study. The first was the school itself, and the second was the technology-supported learning environment in which the students engaged with the learning, entertainment, and metacognitive activities. The elementary school was located in a small, Midwestern city near a large, land-grant research university. The participants included 44 students in two fourth grade classrooms, split evenly between two teachers who commonly used face-to-face problem-based learning environments in their instructional methods. These students were quasi-randomly selected by the school's computer system for assignment to their respective classes; however, the classes constituted a convenience sample.

Upon conducting document analysis of federal documents related to the No Child Left Behind Act (U.S. Department of Education, 2002), fourth grade was selected as the fourth and fifth grade years are commonly targeted for state standardized testing and therefore were perceived by the researchers as a group that would benefit the most from an intervention to help improve their writing skills based on research showing that interdisciplinary, thematic and technology-enabled approaches to teaching literacy are more effective than isolated writing instruction (Englert, Manalo, & Zhao, 2004; Graham & Harris, 2000; Richards, 2002; Shanahan, 1997). The school was selected as a convenience sample because the researchers had a previous research relationship there and teachers from past studies were willing to recommend peers to participate in this study.

The teacher in the treatment condition was recruited for several reasons. While she had been part of Quest Atlantis for the past two years, she had not been an active one. During pre-recruitment discussion, she noted that she was largely uncomfortable with technology and that her classes had not been in the digital learning environment prior to the treatment. Her lack of experience with both technology and with Quest Atlantis, within which Anytown is housed, made her an excellent teacher for this condition because she and her class would not be entering the treatment with pre-set expectations about what they were to do and how to act as an instructor in such an environment. This would allow the treatment to unfold as it would for the majority of teachers who would use a video-game influenced multi-user virtual environment for the first time. The treatment class itself would be starting at the base tutorial stage learning how to navigate in the Quest Atlantis environment, use objects, begin to immerse themselves in the narrative, and approach Anytown without having spent a lot of time exploring the environment and testing the system prior to their participation in the treatment.

The comparison teacher was asked to participate based on recommendations from peers and peer researchers who had encountered her using technology in the computer lab and classroom over the previous two years. Further, during pre-recruitment discussions facilitated by a researcher familiar with her classroom instruction approaches and supported by observation of the teacher in the computer laboratory, she proved herself to be highly expert with technology as she helped students use Quest Atlantis, improvising several times as the system presented challenges to the students such as server outages and internal difficulties of navigation in the 3D space. Based on these observations, discussion, and teacher self-report, her teaching methods were determined to be based in large part on the problem-based learning (PBL) approach proposed by Savery and Duffy (1995), making her face-to-face approach to instruction comparable to the instructional methods present in the Anytown environment.

As we define it here, the core learning aspect of problem-based learning is an authentic, ill-structured problem that is posed to groups of students, which the learners must then wrestle with; they then develop a within-group, socially negotiated solution to this problem. Authentic problems stem from the local, state, and national situations of the learners and are within the learner's zone of proximal development (Vygotsky, 1978) so that the solutions that the learners generate can have real-world impact. The teacher acts as a modeler of appropriate behaviors, provider of resources, and challenger of poor knowledge constructs through cognitively-challenging questions. Outside experts, peers, and the learners themselves engage in assessment of the solution that is presented by each group, also acting to challenge the value of the solution and its practical viability.

In order to further confirm her use of PBL strategies, the researchers engaged in pre-intervention observation of her class. It was noted that this teacher provided authentic ill-structured problems for students to solve in small groups such as writing persuasive arguments to address challenges found in their local community such as water pollution and bullying. Further, she designed her learning environment to provide numerous resources for stimulating writing and critical thinking such as texts related to the problem and online resources such as science Web sites geared towards this age group. In terms of evaluation, she facilitated rubric-based student peer evaluation and, when available, invited experts to evaluate student solutions. In keeping with the PBL philosophy, she allowed students to develop solutions with little interference on her part until there was a serious flaw in their knowledge construction, which she then challenged.

In contrast with the experience of the treatment teacher, the comparison teacher reported and showed evidence in pre-implementation field observations that she had much higher levels of expertise related to teaching with innovative technologies than the treatment teacher. This ruled the teacher out for recruitment as the treatment teacher because this expertise was more likely to act as a confounding factor in any interpretation as to whether the learning environment or the teacher had been responsible for improvements in student learning found during the study. In addition, the teacher reported in pre-recruitment discourse with researchers that she already planned and developed a unit related

to descriptive writing using the senses, which was the part of the focus of the Anytown treatment. The benefit to using this teacher's existing curriculum was that there would be no need to impose an artificial curriculum designed by the researchers on this comparison teacher and would ensure her buy-in to its use. A final important factor was that both the teacher and students in the comparison condition had already used Quest Atlantis more than half a dozen times during that semester, so there would be little to report in terms of student learning challenges related to using an innovative technological curriculum because many of these would have been mitigated during the initial training session. Without observation of these challenges, it would be likely that the researcher would fail to see inherent problems in the design because students may have already developed adaptations that permitted them to succeed where a less experienced class may have been met with failure.

Conditions

Treatment condition. The treatment was student completion of a language arts and reading unit existing completely within the designed multi-user virtual environment known as Anytown. Within this unit, students completed problem-based writing activities embedded within the Anytown setting, customized to prompt the practice of descriptive writing, engagement in problem solving, and student reflection upon their own personal experiences. As described above, the Anytown learning unit contained four types of tasks. The first were termed Writing Quests; these were required of all students and focused specifically on aspects of descriptive or persuasive writing. The three other types could be chosen as part of free-choice activity while they awaited feedback on their primary, required learning tasks, which included the Mystery, Creative Writing, and Reflection Quests. Each task increased in difficulty and complexity over time, which allowed the learner the opportunity to gain competency with developmentally appropriate writing, critical thinking, and cognitive-reflective practices and receive feedback from the teacher prior to moving to the next set of tasks.

The teacher recruited to facilitate the treatment condition self-reported that she was largely uncomfortable with technology and unfamiliar with Quest Atlantis. As such, she had few pre-existing expectations about student direction and participation in the environment—factors which may have confounded some of the impact of the technology treatment. However, this selection had the benefit of allowing the treatment to unfold as it would for the majority of teachers who would use a video-game influenced multi-user virtual environment for the first time. In this way, students and teacher start in a natural fashion at the base tutorial stage in which they learn how to navigate their avatar through the environment, interact with objects, connect with the emerging narrative, and approach Anytown without prior exploration and testing.

The provision of adequate hard scaffolds within the environment in the form of in-game tutors and resources (Baylor, 1999; Baylor & Kim, 2005) was predicted to provide increased student control over of the exploration of learning environment and their own writing products as well as improve student willingness to engage in voluntary writing practice and reading activities. When com-

bined with teacher soft scaffolds in the form of verbal guidance and an immersive, authentic context with tasks linked to future work and learning goals, these scaffolds have been correlated with increased learning in other environments (Ge & Land, 2003; Hedberg, Brown, & Arrighi, 1998).

Comparison condition. The comparison classroom teacher had already developed a unit related to descriptive writing and thus provided the traditional, face-to-face instruction to which the Anytown language arts unit was to be contrasted. Because the comparison teacher used her existing descriptive writing curriculum and taught writing the way that she normally would over the course of the data collection period, researchers avoided imposing an artificial curriculum on the comparison participants. However, the teacher was apprised of the standards that would be addressed by the Anytown curriculum and what assessment measures would be used to compare the performance of her students with the performance of those students in the treatment group. Consequently, she developed a series of voluntary writing activities that paralleled those offered in Anytown. Student participation in voluntary activities served as a measure of whether or not students were motivated by the curriculum. The number completed and the amount of time that students spent working on these activities were indicators of their level of motivation (Sorensen & Maehr, 1976).

Moreover, the comparison teacher was chosen for her expertise in teaching with innovative technologies, application of the problem-based learning methods advocated by Savery and Duffy (1995), and prior experience with Quest Atlantis. These skills further controlled for such skills themselves as confounding variables between comparison and treatment. Although the comparison condition did not implement the Anytown treatment, the teacher's prior experience with instructional technology and problem-based approaches to learning made her aware of the unique role she must play when employing such innovations and offered a degree of assurance that she would provide the directional and procedural scaffolds requisite to such instructional methods, factors which were deliberately embedded in the design of treatment environment.

Instrumentation

Hickey and Pellegrino (2005) classify assessments into three categories: close, proximal, and distal. Close measures are activity-oriented and thus assessment tasks that include the same content and expected skill performance that students engaged in as part of their instructional treatment. While similar to them, they are *not* the exact same activities. Proximal level measures or curriculum-oriented assessment involves evaluation of performance in a different context and with different content than that which existed in the primary learning activities and established curriculum. Finally, distal measures or standards-oriented assessment is commonly focused on student use of learned skills in substantially different contexts or new domains, such as the substitution of social studies instead of science content (Hickey & Pellegrino, 2005). This study employed instrumentation measuring student achievement at all three assessment levels.

Activity-oriented assessment (Close measures). In this study, close measures were the writing products that students submitted through the online system

as they progressed through the *Anytown* unit, as well as those submitted to the teacher in the comparison class. These documents were analyzed to determine if students made incremental improvements in their writing based on feedback over time. They included both mandatory Writing Tasks and three forms of optional writing practice.

In the descriptive writing portion of the *Anytown* Language Arts world, three of the six required *Writing Quests* were designed to gauge student achievement on a progressive scale. The initial introduction Quest titled "Welcome to *Anytown*" allowed the teacher to establish a baseline in terms of the level of descriptive writing students were able to achieve while successive Quests were evaluated to determine progress between learning activities. Assessment rubrics and detailed directions asked evaluators to examine the level of improvement in student detail, elaboration, and extension from their first Quest to their last. The raters were chosen because other Quest Atlantis teachers referred each to us as being those that have evaluated writing using rubrics for state standardized testing in the past. Raters also evaluated optional submissions for improvement in students' writing, analyzing the following in each type of optional task:

- *Mystery Quests*—Rubrics focused on student ability to narrate their experiences, make appropriate use of evidence to support their solution to the mystery, and include a high level of detail used to describe the experiences.
- *Reflective Metacognitive Quests*—These focused on the depth of student reflection and how well students defended their responses.
- *Creative Writing Quests*—Such rubrics examined student ability to generate poems and short stories in response to past experiences within the town using appropriate levels of visual, auditory, and other sense imagery.

Each rater was trained on the specifics of evaluating the pre- and post-test writing prompts in an hour-long session by the lead researcher and a lead teacher-rater, each with more than five years of experience with evaluating writing prompts for state standardized tests. In addition, a set of directions for scoring was provided to each rater to refer to as they evaluated the written responses. Further, examples of writing that met each score level were provided to the teachers so that they could compare the writing resulting from the research project with those that had previously been evaluated for state standardized tests.

Inter-rater reliability was developed using two steps. First, each rater was provided with the same state-normed rubric and independently evaluated each written response. As a group and under the leadership of teacher with the most experience in evaluating writing, the raters then discussed their resulting scores and how they arrived at each until they arrived at 100% agreement. Overall, the optional writing products were evaluated for indicators of writing improvement by individual students rather than judged solely on a single standard or against other students' work.

By contrast, written work produced by students in the comparison classroom was largely generated in groups through Interactive Writing and Readers

Workshops. Students shared their work through peer review, group sharing, and written submission to the teacher. Much student writing work was done in multi-colored markers on large blank sheets of paper that the teacher had pinned to the wall in the Interactive Writing area. Students wrote sentences and paragraphs on these sheets, editing them with specifically colored markers that indicated spelling and grammar problems. While they were present in the classroom, researchers were also given access to student work, as it was generated and then again prior to its return to students.

Curriculum-oriented Assessment (Proximal measures). Student proximal writing achievement changes were measured by a total of two pre- and post-treatment writing activities. Namely, students in both the comparison and treatment groups responded to one of two randomly assigned writing prompts prior to engaging in either curricular unit. The pre-test prompt acted as a base line for where student writing skills and knowledge of the traits of good writing stood prior to instructional treatment. One week following the completion of their respective units, students responded to the writing prompt that they did not use as a pre-test which then served as a post-test. Each was evaluated on a rubric by multiple raters trained to evaluate student language arts and reading work. Using the resulting ratings, pre- and post-test mean scores for both classes and individual students were generated. Assessment rubrics tailored to each prompt were used to determine whether student writing improved from the beginning of the unit to the end.

Standards-oriented Assessment (Distal measures). As with the close and proximal measures, student distal writing achievement changes were measured by a pre- and a post-treatment writing response activity. To qualify them as distal measures, the prompts were not closely matched to the type of descriptive writing completed by students in either instructional treatment. As such, students in both the comparison and experimental groups responded to one of two randomly assigned standards-oriented writing prompts prior to engaging in their respective curricular activities. As with the proximal measures prompts, one of the pre-test distal measure prompts acted as a base line for where student writing skills and knowledge of the traits of good writing stood prior to treatment, the other prompt functioned as the post-test, and each was evaluated by multiple raters, past and present teachers, trained to evaluate student language arts and reading work using a rubric tailored for the prompt. Pre- and post-test mean scores for both classes and individual students were calculated using these ratings and are presented in the Results section.

Procedure

First, students in both classes were randomly assigned to either distal prompt by drawing student names from a hat, a procedure repeated for the proximal prompt. Following these assignments, students in both classes wrote in response to their respective prompts, which were administered by the researcher. Next, both groups began their writing instruction. For the treatment group, this meant that students and teacher began visiting the computer lab and engaging with the *Anytown* environment. For the comparison group, participants began

their pre-planned writing unit. Students in the comparison group continued through their normal writing process until they completed a published piece at the end of the data collection period. However, the comparison students were also provided with poster boards and teacher with written directions that gave comparison participants guidance on additional voluntary writing activities, which were as similar as possible to those available in *Anytown*. Students in the treatment group completed *Anytown* activities that directed them to create a number of published pieces related to descriptive writing.

At the end of the data collection period, each group completed the distal and proximal writing prompts not received during the pre-test, which served to ensure counter-balancing. Administered by the researcher, these prompts were collected, and analysis began. The primary comparison measures were the in-class writing activities and the student standardized writing activities. These measures were intended to determine the difference between the writing practice and achievement between the comparison and experimental groups. The difference in teacher time spent answering basic procedural and directional questions posed by students was also analyzed. Additionally, the number of voluntary writing activities completed by the students was examined as an indicator of student motivation to engage in writing. Most of these voluntary activities were available to both classes in digital and paper or poster forms and acted as free-choice activities. In both conditions, the activities did not come from a single place, but arose throughout the learner's experiences in their learning environments.

RESULTS

These quantitative findings represent the results of analysis conducted on the three separate hypotheses previously outlined in Purposes of the study.

Hypothesis One: Teacher Time On Directional Questions

The data used to test this hypothesis was identified by matching the field notes produced by four different researchers with audio recordings and transcripts of those recordings. The individual lines of these transcripts were time-stamped and coded to distinguish between two forms of dialogue:

- answering student questions about directions and task completion procedures
- all other forms of discourse between students and teacher

Following this coding, the number of minutes spent by the teacher for each student-teacher interaction was calculated.

For this hypothesis, a paired-sample t-test was conducted on the teacher time spent answering scores to see if the mean for the treatment teacher was significantly different from the mean for the comparison teacher. With the alpha set at .05, the paired-C-sample t-test showed that there were significant differences ($t(15) = 5.947$, $p = .043$) between treatment ($M = 12.118$, $SD = 6.6951$) and comparison teachers ($M = 28.413$, $SD = 3.9033$). Figure 1 presents the differences found between the amount of time spent by each teacher answering questions about task directions or procedures for completion of the task.

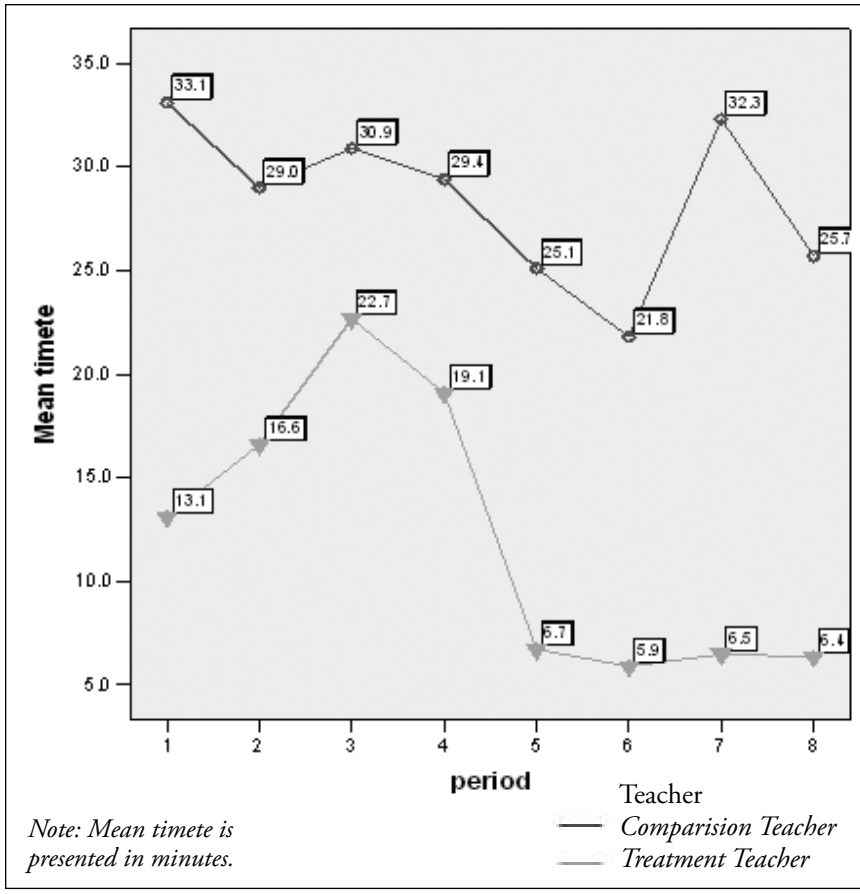

Figure 2: Time spent by teacher on directional questions

The amount of time spent by the treatment teacher shows that on the first two days of implementation, she spent nearly the same amount of time answering directional or procedural questions about the nature of the student tasks. However, by day four, the time spent answering such questions was much reduced. Further, in both instances, the comparison teacher spent more time answering such questions within each hour of instruction.

Hypothesis Two: Voluntary Writing Activity

This hypothesis was tested by collecting the voluntary writing assignments that were produced by students in both the treatment and comparison classes. In the context of this study, a voluntary assignment was defined as any writing activity that was presented by the system or teacher as an option, but was not mandated by the teacher or system to be completed as part of the student's daily work. Further, these learning tasks were not graded although students did receive written feedback and receive in-game rewards that would most closely be described as money used for purchasing useful tools, as well as experience

points that could be used to empower student actions with the larger narrative of *Quest Atlantis*. Students in the treatment group worked on or completed thirty required writing activities. In addition, they also worked on or completed 26 voluntary writing activities. The comparison class did not complete any voluntary writing activities. In addition to the four voluntary activities provided by the researcher that paralleled the *Anytown* creative writing Quests, the comparison teacher also provided several additional opportunities for students to write voluntarily. These included descriptive, comparison and contrast, and creative pieces related to the main writing trajectory of the class, which included sensory descriptive pieces. The students attempted neither the teacher nor researcher provided writing opportunities.

For this research question, a paired-sample t-test was conducted on the teacher number of voluntary writing activities completed to see if the mean for the treatment class was significantly different from the mean for the comparison class. However, because no students in the comparison class completed voluntary writing activities the outcomes were of no use and the statistical data was highly skewed towards the treatment class. Therefore, these data tell us only that students completed more free choice activities when using the computer, but not necessarily why. Given the likelihood that this result may stem from a Hawthorne effect because students in the treatment class were more highly engaged as a result of the novelty of the technology, it would be unfair to attribute the results exclusively to the treatment.

Hypothesis Three: Student Achievement Scores

The achievement scores for hypothesis three are broken up into three sections, close, proximal, and distal, depending on the relative similarity of the achievement task to the learning tasks that comprise the treatment, as described by Hickey and Pellegrino (2005).

Close level scores. Close level achievement scores were produced by collecting the Quest writing prompts that were completed by the students over the course of their time in Anytown. Scores were produced by three teachers who acted as graders for iterations of each Quest. Each written iteration was produced in response to teacher feedback that was provided through the digital agent called the Editor-in-Chief who acted as a proxy for the teacher. Close level scores were only produced for the treatment class, as the comparison class did not complete Quests.

For this question, a paired-sample t-test was conducted comparing the scores of students on the mandatory Quests with scores on the voluntary Quests. This was done to determine whether the mean for the mandatory scores was significantly different from the mean for the voluntary Quest scores. With the alpha set at .05, the paired-sample t-test showed that there were no significant differences ($t(24) = -9.505$, $p = .666$) between the scores on mandatory Quests ($M = 1.67$, $SD = .752$) and voluntary Quests ($M = 2.73$, $SD = .518$).

Proximal level scores. Proximal achievement scores on the standardized writing prompts were obtained using rubric scoring tailored to each prompt, depending on which state had validated the instrument and used it for a stan-

Table 1: Proximal Pre and Posttest Means and Standard Deviations

Teacher	Mean	Standard Deviation	Number
D-Pretest	1.86	.56	19
C-Pretest	2.16	.53	23
Total	2.02	.56	42
D-Posttest	1.79	.56	19
C-Posttest	2.49	.68	23
Total	2.18	.72	42

Table 2: Proximal Level Repeated Measures Analysis of Variance Results

Source of Variation	**SS**	**V**	**MS**	***F***	***P***
Between Subjects					
Comparison teacher (T1)	179.25	1	179.25	725.25	.000
Treatment Teacher (T2)	2.62	1	2.62	10.59	.002
Within Subjects					
Factor1 (F1)	.36	1	.36	1.84	.183
F1 * teacher	.85	1	.85	4.32	.044
Error (F1)	7.84	40	.20		

Note: Factor1 = Mean Pretest (PRD) and Mean Posttest (POD)

dardized assessment of student writing. Three graders, all teachers trained in the grading of student responses, used rubrics to independently grade the pre and post-tests on a six-point scale—one being the lowest and six the highest possible score. Inter-rater reliability was developed by providing the same normed rubric to all three graders who talked through the grades under the lead of the most experienced teacher to get 100% agreement.

For this question, a repeated-measures analysis of variance was conducted comparing the pre and posttest scores for each class to determine whether significant differences existed.

With the alpha set at .05, the repeated-measures ANOVA with a Bonferroni adjustment showed that the scores on the proximal posttest differed significantly ($F(1, 40) = 4.32$.) The Bonferroni adjustment was used to help guarantee that the use of the adjusted alpha would not raise the actual probability of family-wise type I errors above the desired level, as specified by alpha. Table 4 reports relevant distal data.

Distal level scores. Distal level achievement scores on the standardized writing prompts were also measured using rubrics that were tailored to each prompt by either state and were validated and used by these states. Responses to these prompts were graded on a four-point scale, with four being highest and one being the lowest score, by the same three graders who scored the proximal responses. As with the proximal scores, inter-rater reliability was developed by

Table 3: Distal Pre and Posttest Means and Standard Deviations

Teacher	Mean	Standard Deviation	Number
D-Pretest	2.32	.50	19
C-Pretest	2.36	.61	23
Total	2.34	.56	42
D-Posttest	2.05	.54	19
C-Posttest	2.17	.56	23
Total	2.12	.55	42

Table 4: Distal Level Repeated Measures ANOVA Results.

Source of Variation	**SS**	**V**	**MS**	***F***	***P***
Between Subjects					
Comparison teacher (T1)	206.26	1	206.26	891.2	.000
Treatment Teacher (T2)	.073	1	.073	.32	.577
Within Subjects					
Factor1 (F1)	1.06	1	1.06	6.77	.013
F1 * teacher	.029	1	.029	.186	.669
Error (F1)	6.28	40	.16		

Note: Factor1 = Mean Pretest (PRD) and Mean Posttest (POD)

providing the same normed rubric to all three graders who talked through the grades under the lead of the most experienced teacher to attain 100% agreement.

For this question, a repeated-measures analysis of variance was conducted comparing the first Quest iteration scores with the last iteration scores to determine whether the mean for the final iteration scores was significantly different from the mean for the initial Quest scores.

With the alpha set at .05, the one-way repeated-measures analysis of variance showed that the scores on the distal posttest differed significantly (F (1, 40) = 6.77, $p < .05$) by teacher. The following chart reports the relevant distal data.

Limitations

There are limitations to the generalizability and validity of the proposed study, due to both the choices made when developing the study and to unavoidable problems that could not be completely controlled. The first threat to validity was that the teacher in the comparison group may already have had a high level of ability and knowledge relevant to teaching a problem-based learning in a face-to-face learning environment. Another threat to the validity of the study was the use of only two intact classrooms, which results in limited generalizability of the results of this study to other classes and contexts. Taking a group from only one part of the state also reduced the validity and reliability of the study.

Due to the cultural differences between students within the classes themselves, the scores will be more or less valid dependent on the students' personal experience and relationship to the questions presented in the pretest and posttest.

Statistical regression toward the mean is another limitation faced by the study because students who score high on a pretest tend to earn lower scores on the posttest, and students who score lower on the pretest tend to score higher on the posttest. Therefore, those students with high writing ability prior to the treatment will be seen to have made smaller gains, though they may be at or well above their grade level in terms of ability. Further, while some generalizability to students in similar situations and school districts is warranted, drawing conclusions in terms of a larger state or national population would likely be fallacious due to the local nature of the student and teacher experience as part of the treatment sample. However, this data's use in terms of framing the quantitative findings, identifying confounding or mitigating factors, describing the learning experiences of students and teacher, and in future design and development of instruction makes it valuable.

Last, it is likely that a Hawthorne effect (Macefield, 2007) was present in the treatment class resulting from the new technology introduced to students who had previously had little interaction with other Quest Atlantis activities or a video-game influenced digital environment like that of the Anytown world. It is likely that as students continued to use Anytown, the "cool factor" of using the digital environment would fade and that at least a portion of these findings, especially related to student free choice activities would also dissipate. Repeated or longer-term studies with these students as they engaged with the environment in question here or the larger universe of Quest Atlantis should be conducted to determine whether the underlying instruction and learning activities are sufficient to sustain long-term student engagement with these environments.

DISCUSSION

During three decades of writing on the subject, Krashen (1991) has noted that increased writing practice is vital to improving general literacy skills. The need to increase student time-on-task practicing writing drove the design of the Anytown curriculum (Englert, Manalo, & Zhao, 2004). In this development, we sought to embed the main elements of problem-based learning (Savery & Duffy, 1995) by including ill-structured problems that students could work to solve using the collaborative technology tools and embedded scaffolds within the digital environment. These elements were designed as a means to reduce the normally teacher-intensive work of building a constructivist learning environment (Jonassen, 1999) and to provide continuous guidance regarding tasks, while concurrently providing freedom of activity choice that is normally not available in face-to-face problem-based learning. In doing this, we developed fully interactive stories to support learner problem-solving (Jonassen & Hernandez-Serrano, 2002), provided differentiated trajectories of experience that emerged from these stories, and included multiple characters to provide direction and other forms of guidance traditionally reserved for the teacher. Concurrently, the design leveraged elements of video games (Salen & Zim-

merman, 2004), which were encapsulated within these evolving stories such as artificial conflict to drive learner activity, a rule-based interactive system to help govern student action by providing linear experiences within each trajectory to provide a coherent narrative that would frame learner understanding of their writing experiences, and ensured that students could "win" and receive rewards for completing their problem-solving and related writing activities.

The results of this design indicate that students were motivated by the narrative structures to engage in substantially more free-choice writing practice in the treatment class at a ratio of 26:0 over the comparison class. From a curricular design viewpoint, the teacher did not have to push the optional activities in the digital environment as they emerged as choices, yet they enhanced the learner's standing within the game function of Anytown and allowed students to earn rewards and open additional content. This is in keeping with the findings in the study of the Taiga science-based environment that "teachers need to establish rich narrative contexts…this implies that the task of curricular experience is to situate students within a rich context" (Barab et al., In Press). Given that the role of the digital environment in this instance was mainly to differently contextualize the learner's writing experience as a means of motivating increased practice, teachers need not necessarily build games, but can leverage this form of reward scenario within their own face-to-face, free-choice learning activities by providing a more rich context for student writing while providing students additional external motivation which can slowly be removed to foster intrinsic motivation.

In related qualitative research (Warren, 2006), the teacher indicated that the significant decrease in time spent repeating directions, reinforcing procedure, or performing other administrative tasks in the treatment group increased the time that she was available to engage in teaching those facilitating behaviors that are commonly associated with problem-based learning such as providing improved and specific feedback on student writing, giving encouragement, empathizing with student struggles, asking cognitively challenging questions, providing soft scaffolds to struggling writers and readers, and providing tools and resources to students as they seek to become better writers. A benefit to developing the written activities in the 3D space and simulating instructional roles through pedagogical agents is that the teacher is not responsible for developing complex content, embedding multiple hard scaffolds in their classroom, or generating ill-structure problems for students to solve, thus allowing the teacher's role to evolve into that of the guide, coach or facilitator.

Given the instructional designers' focus on meeting student achievement standards, it was remarkable that students in the Anytown digital environment demonstrated improvements on writing measures in just seven treatment periods as opposed to the lack of similar gains in the comparison classroom over the same period, indicating a higher level of efficiency of the digital PBL curriculum when compared with a more traditional form. As social constructivist methods and PBL specifically are commonly faulted for requiring increased time versus more traditional, objectivist methods (Airasian & Walsh, 1997; Matthews, 2003), this finding indicates that it is not necessarily the method

that result in this difficulty, but instead it is the efficiency of the design of the instructional environment and its accompanying scaffolds that may be to blame. We believe that if digital constructivist learning environments are designed as Merrill (2002) suggests so that the scaffolding is gradually stripped away throughout a series of related problems while the authenticity of learning experience increases, moving from the protection of the simulation to problem solving with real-world consequences such as writing articles for the school newspaper will occur much more seamlessly.

Future Directions

In terms of future research, three major lines should next be explored: those involved with Anytown, those that focus on the use of games or game-like learning environments in general, and student problem-solving within game-infused problem-based learning environments. In terms of the Anytown environment, the role of peer teaching and support, as well as game incentives (such as receiving objects that permit the learner to do something special like opening a locked door) and how they contribute to student successes on Quest tasks is ripe for further exploration.

Within the parameters of the larger video game genre as it relates to instruction, whether and how games can be harnessed to more effectively train students to perform real-world tasks related to standards in other content areas. Similarly, researchers like Baylor and Kim (2005) are making strides with the use of pedagogical agents in learning environments. However, since the complexity of these characters does not reach the level of primary instructor in many cases, future research should explore whether the use of such agents should be used to replace certain instructional behaviors or only support them.

Another important topic to be explored is how the principles of game design can be leveraged to improve non-digital, face-to-face instruction. Based on findings from Steinkuehler's (2004) work with the game Lineage II combined with the findings of this, a rich area of interest is in exploring the relationships between student engagement, reading practices, and models of digital means of creating intrinsically motivating reading, writing, and problem-solving practices for K–12 students. Future studies should also examine the quality of the student experience, the specific strategies used by students in order to solve the ill-structured problems posed by the learning environment, the degree to which students work collaboratively to solve problems, and the extent to which they are able to construct valid arguments in support of their solutions.

Contributors

Scott Warren works as an assistant professor of Learning Technologies at the University of North Texas. Dr. Warren holds a PhD in Instructional Systems Technology and his research examines the use of emerging online technologies such as podcasting, immersive digital learning environments, digital video, and open source course management tools in support of developing complex instructional systems. (Address: Scott J. Warren, University of North Texas , 1300 Highland Street, Matthews Hall 322C, PO Box 310530, Denton, TX 76203-1335; swarren@unt.edu; Phone: 940.765.2799; Fax: 940.565.2185.)

Sasha Barab is an associate professor in Learning Sciences, Instructional Systems Technology and Cognitive Science at Indiana University. He also holds the Barbara Jacobs Chair of Education and Technology. His current work involves the research and development of rich learning environments, frequently with the aid of technology, that are designed to assist children in developing their sense of purpose as individuals, as members of their communities, and as knowledgeable citizens of the world. (Address: Sasha A. Barab, Indiana University-Bloomington, sbarab@indiana.edu; Phone: 812.856.1553; Fax: 812.856.0862.)

Mary Jo Dondlinger is an instructional designer at Richland College and a doctoral student in Educational Computing at the University of North Texas. Mary Jo Dondlinger, University of North Texas, mdondlinger@unt.edu; Phone: 940.369.7489; Fax: 940.565.2185.)

Funding

This research was supported in part by a CAREER Grant from the National Science Foundation, and directly by the National Science Foundation Grants #9980081 and #0092831.

Acknowledgements

The authors would like to thank several people for their hard work in support of the design of Anytown and many hours spent collecting data for this study. These include: Heather Warren, Richard Stein, Dan Hickey, Steve Zuiker, Adam Ingram-Goble, Inna Kouper, Susan Herring, Eun-ju Kwon, Robert Appelman, and Mitzi Lewison.

References

Airasian, P. W., & Walsh, M. E. (1997). Constructivist cautions. *Phi Delta Kappan, 78*(6), 444–449.

Aldrich, C. (2003). *Simulations and the future of learning*. San Francisco: Pfeiffer.

Amory, A., Naicker, K., Vincent, J., & Adams, C. (1999). The use of computer games as an educational tool: Identification of appropriate game types and game elements. *British Journal of Educational Technology, 30*(4), 311–321.

Anderman, E. (1996). The middle school experience: Effects on the math and science achievement of adolescents with LD. *Journal of Learning Disabilities, 31*(March/April), 128–138.

Barab, S. A., Sadler, T. D., Heiselt, C., Hickey, D., & Zuiker, S. (2006). *Relating narrative, inquiry, and inscriptions: A framework for socio-scientific inquiry.* Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco, CA.

Barab, S. A., Scott, B., Ingram-Goble, A., Goldstone, R., Zuiker, S., & Warren, S. J. (In press). Contextual embodiment as a curricular scaffold for transferable understanding. *Contemporary Educational Psychology*, 34.

Barab, S. A., & Squire, K. (2004). Design-based research: Putting a stake in the ground. *The Journal of the Learning Sciences, 13*(1), 1–14.

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

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

Baylor, A. (1999). Intelligent agents as cognitive tools. *Educational Technology, 39*(2), 36–40.

Baylor, A., & Kim, Y. (2005). Simulating instructional roles through pedagogical agents. *International Journal of Artificial Intelligence in Education, 15*(1), 95–115.

Cassell, J., & Jenkins, H. (Eds.). (2000). *From Barbie to Mortal Kombat*. Cambridge, MA: The MIT Press.

Clark, R. E. (1991). When researchers swim upstream: Reflections on an unpopular argument about learning from media. *Educational Technology 31*(2), 34–40.

Crawford, C. (2003). *The art of interactive design: A euphonious and illuminating guide to building successful software.* San Francisco: No Starch Press.

Cuban, L., Kirkpatrick, H., & Peck, C. (2001). High access and low use of technologies in high school classrooms: Explaining an apparent paradox. *American Educational Research Journal, 38*(4), 813–834.

Dede, C., Ketelhut, D., & Ruess, K. (2006). Designing for motivation and usability in a museum-based multi-user virtual environment. Retrieved March 11, 2006, from http://www.gse.harvard.edu/~dedech/muvees/documents/AELppr.pdf

Denis, G., & Jouvelot, P. (2005). *Motivation-driven educational game design: applying best practices to music education.* Paper presented at the 2005 ACM SIGCHI International Conference on Advances in Computer Entertainment Technology, Valencia, Spain.

Dickey, M. D. (2005). Three-dimensional virtual worlds and distance learning: Two case studies of Active Worlds as a medium for distance education. *British Journal of Educational Technology, 36*(3), 439–451.

Dickey, M. D. (2006). *"Ninja Looting" for instructional design: The design challenges of creating a game-based learning environment.* Paper presented at the ACM SIGGRAPH 2006 conference, Boston.

Dickey, M. D. (2007). Game design and learning: A conjectural analysis of how massively multiple online role-playing games (MMORPGs) foster intrinsic motivation. *Educational Technology Research & Development, 55*(3), 253–273.

Dondlinger, M. J. (2007). Educational video game design: A review of the literature [Electronic Version]. *Journal of Applied Educational Technology*, *4*, 1–11. Retrieved July 19, 2007, from http://www.eduquery.com/jaet/index.htm.

Egbert, J., & Hanson-Smith, E. (Eds.). (1999). *CALL environments: Research, practice, and critical issues*. Alexandria, VA: Teachers of English to Speakers of Other Languages, Inc.

Englert, C. S., Manalo, M., & Zhao, Y. (2004). I can do it better on the computer: The effects of technology-enabled scaffolding on young writers' composition. *Journal of Special Education Technology, 19*(1), 5–21.

Entertainment Software Association. (2007). Facts and Research. Retrieved January 7, 2007, from http://theesa.com/facts/gamer_data.php_

Fisch, S. M. (2005). *Making educational computer games "educational."* Paper presented at the 2005 conference on Interaction design and children, Boulder, Colorado.

Gall, M. D., Borg, W. R., & Gall, J. P. (1996). *Educational Research: An introduction* (6th ed. Vol. I). White Plains, NY: Longman Publishers.

Ge, X., & Land, S. (2003). Scaffolding students' problem-solving processes in an ill-structured task using question prompts and peer interactions. *Educational Technology Research & Development, 51*(1), 21–38.

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

Graham, S., & Harris, K. (2000). The role of self-regulation and transcription skills in writing and writing development. *Educational Psychologist, 35*(1), 3–12.

Hedberg, J., Brown, C., & Arrighi, M. (1998). Interactive multimedia and Web-based learning: Similarities and differences. In B. Khan (Ed.), *Web-based instruction*, (pp 47–58). Englewood Cliffs, NJ: Educational Technology Publications.

Hickey, D. T., & Pellegrino, J. W. (Eds.). (2005). *Theory, level, and function: Three dimensions for understanding the connections between transfer and student assessment.* Greenwich, CT: Information Age Publishers.

Jenkins, H., Squire, K., & Tan, P. (2003). Entering the education arcade. *ACM Computers in Entertainment, 1*(1), 1–11.

Jennings, M. (2001). *Best practices in corporate training and the role of aesthetics: Interviews with eight experts.* Paper presented at the 2001 ACM SIGCPR conference on Computer Personnel Research, San Diego, California. Available http://portal.acm.org/citation.cfm?id=371209.371239

Jonassen, D. H. (1999). Designing constructivist learning environments. In C. M. Reigeluth (Ed.), *Instructional design theories and models: A new paradigm of instructional theory* (Vol. 2, pp. 215–239). Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.

Jonassen, D. H., & Hernandez-Serrano, J. (2002). Case-based reasoning and instructional design: Using stories to support problem solving. *Educational Technology Research & Development, 50*(2), 65–77.

Kaplan, F. (2003, March 28). Why the army shouldn't be so surprised by Saddam's moves. *Slate war stories* Retrieved February 22, 2004, from http://slate.msn.com/id/2080814/

Kozma, R. (1991). Learning with media. *Review of Educational Research, 61*(2), 179–211.

Krashen, S. D. (1991). *Writing-research, theory, and applications*. Torrance, CA: Laredo Pub. Co.

Lee, J., Luchini, K., Michael, B., Norris, C., & Soloway, E. (2004). *More than just fun and games: Assessing the value of educational video games in the classroom.* Paper presented at the CHI '04 extended abstracts on Human Factors in Computing Systems, Vienna, Austria.

Linn, M. C., Clark, D., & Slotta, J. D. (2003). WISE Design for knowledge integration. *Science Education, 87*, 517–538.
Macefield, R. (2007). Usability studies and the Hawthorne Effect. *Journal of Usability Studies, 2*(3), 145–154.
Matthews, W. J. (2003). Constructivism in the classroom: Epistemology, history, and empirical evidence. *Teacher Education Quarterly, 30*(3), 51–64.
Merrill, M. D. (2002). A pebble-in-the-pond model for instructional design. *Performance Improvement, 41*(7), 39–44.
Nieborg, D. B. (2005). *Changing the rules of engagement: Tapping into the popular culture of America's Army, the official U.S. Army computer game.* Unpublished study, Universiteit Utrecht, Utrecht, NL.
Prensky, M. (2001). *Digital game-based learning*. New York: McGraw-Hill.
Richards, R. G. (2002). *Strategies for the reluctant writer*. Riverside, CA: RET Center Press.
Salen, K., & Zimmerman, E. (2004). *Rules of play: Game design fundamentals*. Cambridge, MA: MIT Press.
Samsonov, P., Pedersen, S., & Hill, C. (2006). Using problem-based learning software with at-risk students: A case study. *Computers in Schools, 23*(1/2), 111–124.
Savery, J. R., & Duffy, T. M. (1995). Problem based learning: An instructional model and its constructivist framework. *Educational Technology, 35*, 31–38.
Schneider, M., Carley, K., & Moon, I. C. (2005). *Detailed comparison of America's Army game and Unit of Action experiments*. United States Army. Available http://cos.cs.cmu.edu/publications/papers/CMU-ISRI-05-139.pdf
Shanahan, T. (1997). Reading-writing relationships, thematic units, inquiry learning...in pursuit of effective integrated literacy instruction. *The Reading Teacher, 51*(1), 12–19.
Sorensen, R. L., & Maehr, M. L. (1976). Toward the experimental analysis of "continuing motivation." *The Journal of Educational Research*, *69*, 319–322.
Squire, K. (2004). *Replaying history.* Doctoral dissertation, Indiana University-Bloomington. ISBN 9780496134908. DAI-A 65/11.
Squire, K. (2006). From content to context: Videogames as designed experience. *Educational Researcher, 35*(8), 19–29.
Squire, K., Giovanetto, L., Devane, B., & Durga, S. (2005). From users to designers: Building a self-organizing game-based learning environment. *Tech Trends, 49*(5), 34–42, 74.
Squire, K., & Steinkuehler, C. (2005). Generating cyberculture/s: The case of Star Wars Galaxies. In D. Gibbs & K. L. Krause (Eds.), *Cyberlines 2.0: Languages and cultures of the Internet* (2nd ed.). Vol. 1, pp. 177–198). Albert Park, Australia: James Nicholas Publishers
Steinkuehler, C. (2004). *The literacy practices of massively multiplayer online gaming.* Paper presented at the American Educational Research Association, San Diego, CA.
Tuzun, H. (2004). *Motivating learners in educational computer games.* Unpublished dissertation, Indiana University, Bloomington, IN.

U.S. Department of Education. (2002). *Strategic Plan, 2002–2007.* Retrieved March 10, 2004, from http://www.ed.gov/about/reports/strat/plan2002-07/index.html

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes.* Cambridge, MA: Harvard University Press.

Waraich, A. (2004). *Using narrative as a motivating device to teach binary arithmetic and logic gates.* Paper presented at the 9th annual SIGCSE conference on Innovation and technology in computer science education, Leeds, United Kingdom.

Warren, S. J. (2006). *The impact of a multi-user virtual environment (MUVE) teacher instructional time, voluntary student writing practice, and student writing achievement.* Unpublished dissertation, Indiana University, Bloomington, Indiana.