Unpacking the potential of educational gaming: A new tool for gaming research

Herbert H. Wideman  
York University

Ronald D. Owston  
York University

Christine Brown  
York University

Andre Kushniruk  
University of Victoria

Francis Ho  
University of Victoria

Kevin C. Pitts  
Seneca College

The article begins by reviewing the theoretical bases for the contention that advanced computer-based educational gaming can provide powerful learning experiences, and overviews the limited research on the use of such games. Although studies to date have generally supported their value, most of the published investigations have methodological limitations. Critical process data are typically not collected, and unreliable student and teacher self-reports are heavily relied on in evaluating the educational efficacy of many games. To address these and other limitations, the authors have developed research software that can remotely and unobtrusively record screen activity during game play in classroom settings together with synchronized audio of player discussion. A field trial of this data collection system in which 42 college students were studied as they played a coursework-related Web-based learning game is described, and the article discusses how the trial outcomes concretely demonstrate the methodological advantages the tool offers researchers.

KEYWORDS: educational games; field trials; game evaluation; gaming research; research methodology; research software; task analysis

Over the past two decades, competition for market share in the computer gaming business has spurred the development of ever more complex recreational computer games.1 To capture and hold player interest, games are now being created that engage players in a wide range of potentially rewarding activities and challenges, requiring them to actively investigate the game environment and apply different...

AUTHORS’ NOTE: This research was supported by a grant from the Social Sciences and Humanities Council of Canada to the Simulation and Advanced Gaming Environments for Learning research network.

DOI: 10.1177/1046878106297650  
© 2007 Sage Publications
problem-solving strategies. Game play in genres such as role-playing, simulation, and real-time strategy now calls on considerable in-situ learning and the application of a range of cognitive and metacognitive skills. In addition, the increasingly popular genre of multiplayer games require players to employ social learning skills in support of collective problem solving, social negotiation, and distributed learning (Gee, 2003). Interest in exploiting the apparent educational potential of these recent forms of advanced gaming in formal educational contexts is surging (Becta, 2001; Kirriemuir & McFarlane, 2003; Shaffer, Squire, Halverson, & Gee, 2004). In this article, we first briefly review the theory-based arguments that have been offered in support of advanced educational gaming to develop an understanding of the main assertions of gaming proponents that require empirical investigation. We then discuss the state of the research literature exploring computer gaming in education, summarizing central findings, and examining the strengths and limitations of the various methodologies and designs employed. Finally, we describe a new software tool for gaming research and evaluation we have developed to address some of the limitations of current investigative methodologies and present results from a successful feasibility study of the tool.

**Rationales for educational gaming**

Gee (2003) argued that complex recreational computer games must embody effective learning principles if they are to be learned and played without high levels of frustration—a prerequisite for their commercial success. His analysis of the task demands and designs of several contemporary computer games belonging to different genres found that successful games incorporate learning principles that are congruent with those articulated in the most broadly supported theories of learning and cognition. The key factors that facilitate game-based learning that Gee and others (e.g., Cordova & Lepper, 1996; Garris, Ahlers, & Driskell, 2002) have identified include the experiential and fully situated nature of game learning, the intrinsically motivating nature of key game attributes, and the provision of educationally rich contexts for the development of expertise through participation in communities of practice (real and virtual).

**Situated experiential learning**

The lack of student motivation evident in traditional schooling has been viewed by many educational theorists and researchers as largely a consequence of the routinized decontextualization of instruction—the presentation of knowledge to students in its most abstract forms (Lave, 1998). Learning is removed from contexts in which it is has instrumental utility and divorced from students’ intrinsic interests (Cordova & Lepper, 1996). In contrast, effective games embed learning in meaningful situations that are endogenous to the game itself. The personally meaningful and valued social and material worlds in which game learning takes place may be “virtual” from an outsider’s
perspective; however, they have a psychological reality for the player that directly mediates the player’s level of immersion, persistence in the face of challenges, and intrinsic desire to learn. Gaming can make it possible for new, situated understandings to be developed through embodied experiences in complex domains that are otherwise inaccessible (Gee, 2003). Factual learning comes more easily when learners are immersed in personally meaningful experiences and use those facts for achieving desired ends within that situated domain (Shaffer et al., 2004).

Learning in advanced gaming occurs largely as an iterative process in which concrete experience is observed and reflected on, leading to the development of mental models and inferences that are then applied to the environment to test emergent conclusions, generating further concrete experience. This “trial-and-error” approach to meeting the types of challenges common to most recreational games has been seen as supporting the development of logical thinking and problem-solving skills (Higgins, 2000; Inkpen, Booth, Gribble, & Klawe, 1995; Prensky, 2001; Whitebread, 1997). The process embodies Kolb’s (1984) four-stage model of effective experiential learning; it is how children naturally learn outside of the school context, and it forms the basis for many types of expert practice (Gee, 2003).

It is generally accepted that useful knowledge is contextualized knowledge—the learner must know when and where to use it (Bransford, Brown, & Cocking, 2000). Games can provide experiences across multiple situated contexts that enable learners to understand complex concepts “without losing the connections between abstract ideas and the real problems they can be used to solve” (Shaffer et al., 2004, p. 5). Such situated learning is considered to be critical in two central theories of learning: situated cognition (Lave, 1998; Lave & Wenger, 1991) and constructivism (Bransford et al., 2000). These theories emphasize that learning is shaped and constrained by the intentions of the learner and situational factors (Squire, 2002). Evidence of the importance of meaningful contextualization in educational games can be seen in a series of experiments by Lepper and colleagues, who found that providing concrete contextualizations for solving gamelike puzzles by embedding them in a simple fantasy narrative markedly enhanced student motivation and learning (Cordova & Lepper, 1996; Parker & Lepper, 1992).

**Game attributes and motivation**

In addition to their provision of concrete, situated contexts for learning, a number of other attributes of advanced commercial and educational games have been cited as critical in encouraging active engagement, perseverance, and high levels of immersion in game play (cf., Garris et al., 2002, for a more complete review). These include the use of high resolution multimedia to create an immersive and quasi-realistic sensory experience (Mitchell & Savill-Smith, 2004); the generation of a valued virtual identity through the inclusion of fantasy, narrative, and role-playing elements in game design (Gee, 2003; Squire et al., 2003); and the creation in the player of a sense of power through amplification of their inputs and a sense of competence and achievement through the structuring of progressively more difficult tasks that challenge players at
the “edge of their region of competence” and require the learning of new strategies (Gee, 2003). Competition, either against oneself or others, is thought to heighten a player’s sense of accomplishment and efficacy (Mitchell & Savill-Smith, 2004). The delivery of progressively richer rewards and payoffs proportional to attainment to motivate further achievement (Bonk & Dennen, 2004). Other key motivational factors in game design that have been noted widely in the literature include providing for high levels of interactivity and a wide range of user choice in the game to enhance a player’s sense of agency, delivering rapid and informative feedback so as to increase player motivation to resolve any discrepancy between the player’s current level of attainment and desired mastery (Garris et al., 2002). All of these elements collectively contribute to making games intrinsically fun to play, and play is “part of the natural learning process in human development” (Bisson & Luckner, 1996, p. 112).

Nearly all of these elements have been shown in the learning research literature to contribute directly to learner motivation, and through this mediating process to learning itself (see, e.g., Bransford et al., 2000). Providing the learner with a challenging environment, short-term goals, and immediate feedback on performance are examples of motivational factors whose importance in learning has been well grounded in research (Blumenfeld et al., 1991; Pintrich & Schunk, 1996; Stipek, 1998). The criticality of engagement and persistence to effective learning is also well established. There is considerable evidence that learning improves as the quality of cognitive engagement increases (Hannafin & Hooper, 1993); and the extent and depth of learning (and subsequent learning transfer) is directly affected by students’ willingness to persist in learning and solving complex problems (Bransford et al., 2000).

Communities of practice and the development of expertise

Many commercial games engage players as participants in what Lave and Wenger (1991) have termed communities of practice, in which knowledge, symbol systems, skills, resources, and tools are shared with other players and utilized in the pursuit of common learning and mastery goals. In some instances the community of practice is peripheral to the actual playing of the game, as participants use online forums, game-focused Web sites, e-mail and instant messaging to share game information, resolve technical problems, and address the game’s challenging tasks. In these virtual communities, players reflect, plan, and interact to build a shared understanding, value system, and culture around the game itself. In other instances—most commonly in what are termed massively multiplayer games (fantasy or historical role-playing games in which hundreds or even thousands of players participate through personal “avatars” in persistent virtual worlds over many months)—the community of practice is embedded largely in the game itself, as players messenger and chat with each other in the game, forming various alliances, groupings, and “guilds” to solve problems and achieve mutually valued game goals. In both contexts, participation in a community of practice can provide a powerful impetus for learning by activating socially oriented motivational factors such as status seeking, acknowledgment, and social affiliation (Bonk & Dennen, 2004).
Lave and Wenger (1991) described the learning trajectory taken by participants in these communities. New members begin on the far periphery by assessing the community and its value to them. On deciding to participate they then begin to move into the community more fully, learning its practices, values, roles, and symbols through observation, practice, and interaction with more experienced members. Their inbound migration reaches a peak when they achieve experienced insider status. They then help new players develop the skills and knowledge needed to perform well by modeling that performance, offering feedback and advice to others, and serving as a key knowledge resource for the community (Steinkuehler, 2004; Tharp, 1993). This trajectory within a community of practice has long been the traditional route to the development of professional-level expertise in a wide range of disciplines. Gee, Squire, and others (Bonk & Dennen, 2004; Gee, 2003; Kirriemuir & McFarlane, 2003; Shaffer et al., 2004; Squire, 2002) contended that the communities that gaming generates provide powerful contexts for learning the shared values and ways of thinking common to “real-world” communities. Properly designed educational games, it is argued, can function as practice fields (Senge, 1994), which engage learners in many of the authentic tasks of a real-world community of practice and require the application of the same problem-solving and critical-thinking skills. These authentic practices are fundamental to the development of domain expertise, which is demonstrated in an advanced capacity to recognize patterns and work with deeper organizing principles and core concepts (Bransford et al., 2000). The limited research to date on the social context of game play and learning suggests that games do foster goal-directed social interaction that may facilitate collaborative learning (Kirriemuir & McFarlane, 2003).

The current state of educational gaming

To date, advanced game use has been far more prevalent in medical and business education, where simulation games for teaching medical diagnosis and business strategy are occasionally employed (Faria, 2001; Gredler, 2004; Wolfe, 1997). At the K-12 level, the use of simulation games such as SimCity has been growing (Kirriemuir & McFarlane, 2003; Squire, 2003); however, generally there has been a lack of rigor in educational simulation game design (Ruben, 1999).

Commenting on the generally poor state of educational games used in schools, Leddo (1996) noted that most extant K-12 educational games would never be voluntarily played outside of school. They typically lack the qualities that generate user immersion, such as high-quality graphics and sound, engaging narratives, polished user interfaces, and suspense (Dempsey, Lucassen, Haynes, & Casey, 1996); they generally incorporate drill-like activities that are repetitive and quickly become boring to students; and their tasks are poorly designed, focus on lower-level forms of learning, and fail to support progressive understanding (Kirriemuir & McFarlane, 2003).

Educational gaming: Research findings

At present there is limited evidence beyond anecdotal reportage to support the use of computer games in education (Bonk & Dennen, 2004; Dempsey, Haynes,
Lucassen, & Casey, 2002; Gredler, 2004; Kirriemuir & McFarlane, 2003; Washbush & Gosen, 2001). Most educational games to date have not been grounded in any coherent theory of learning, nor are they directly supported by any body of research (Shaffer et al., 2004). At the level of basic research, major gaps exist in our knowledge of the gaming process, the specific design elements that contribute to its effective use, and the influence of play context and pedagogical support (Wolfe & Crookall, 1998). There is skepticism in the educational community regarding the applicability of gaming to education, which can only be addressed through basic research, design experiments, and successful demonstration projects that are rigorously evaluated and show clearly documented benefits (Bonk & Dennen, 2004). This will require new methods and tools for unpacking the complex processes at play in advanced gaming and for investigating the wide range of possible outcomes in educational contexts.

Significant bodies of research in educational gaming exist in only two disciplines—medical education and business management studies. The most extensive research and evaluation has been conducted on strategic management simulation games. A review of these investigations uncovered a total of 11 well-designed experimental studies of business simulations in the literature (Washbush & Gosen, 2001). The authors concluded that use of the simulations improved learning by an average of 10% on pre- and postlearning assessments. It is interesting to note that no correlation was found between simulation performance (the simulation outcomes themselves) and the learning assessment score gains. They suggest that average performers may have learned a great deal as they struggled to improve. A review that included quasi-experimental research reached a similar conclusion, reporting a consistent finding across studies that simulation gaming produced better learning than the use of business case studies (Wolfe, 1997). Even in the most heavily researched area of simulation games—business simulations—a recent review found that the extant research was not extensive enough to firmly conclude that teaching via simulation games was effective due to a shortage of rigorously conducted studies (Gosen & Washbush, 2004).

The study of advanced medical-education gaming has not been as extensive; however, the observed learning outcomes appear to be largely positive. The findings of two studies of the efficacy of sex education games for adolescents provide illustrative examples. One employed a rigorous experimental design of a role-playing game and found student improvements on validated scales of understanding, attitude, and motivation. Dramatically better cognitive learning outcomes occurred when students took the role of the opposite sex in the game (Kashibuchi & Sakamoto, 2001), one indication of the powerful impact that the assumption of novel perspectives and identities can have on learning. In the second study, significant gains were found in knowledge and self-efficacy scores amongst students who used an adventure game designed to encourage students to develop safer-sex negotiation skills and a sense of self-efficacy in HIV/AIDS prevention programs (Thomas, Cahill, & Santilli, 1997).
Most of the very limited numbers of investigations of computer gaming use in K-12 education have consisted of case studies. Collectively this research has provided some initial evidence supporting the use of games for enhancing the understanding and learning of complex subject matter. Games have been effective in raising achievement levels in mathematics and language arts and have had a positive impact on student motivation (Mitchell & Savill-Smith, 2004). Squire (2004), for example, used the quasi-historical civilization-building simulation game CIVILIZATION III (2001) with secondary school students in two schools. Based on his observations of game play and discussions with students, he concluded that students “developed deeper understandings of social science concepts introduced in the game” (p. 138) and a deeper appreciation of geographical features. A few students went further to develop deeper understandings of the relationships between physical and political geography, economics, and history (Squire et al., 2003). Other studies have also reported favorably on the use of the SimCity game as a platform for expanding skills in mathematics, urban planning, economics, and engineering (Kirriemuir & McFarlane, 2003).

An evaluation of the use of educationally relevant commercial simulation and adventure games in 12 elementary and secondary schools found that according to teacher reports children engaged in deductive reasoning, collaborative problem solving, colearning, and peer tutoring. Students preferred challenging games with a rising gradient of difficulty over time. There was a universal perception on the part of teachers that quest and simulation games stimulated considerable learning of content and skills (McFarlane, Sparrowhawk, & Heald, 2002). In a similar study that used 11 strategy and simulation games in seven schools teachers reported that the combination of simulation gaming with teacher input and collaborative learning provided “powerful learning experiences” for intermediate- and senior-level students throughout the ability range. The combination of high game interactivity within an appealing narrative context that also had a considerable degree of novelty, together with clear a priori aims for the learning, proved very effective at promoting learning (Becta, 2001). Student motivation, collaboration, and persistence were found to be very high, social construction of knowledge was evident, and students benefited from the immediate feedback provided by the games in testing and revising strategies of play.

In a series of controlled experiments with elementary schoolchildren, Lepper and colleagues have examined the effects of several central game attributes on computer-based learning performance, student attitudes, and perceived efficacy (Cordova & Lepper, 1996; Parker & Lepper, 1992). Embellishing abstract graphical performance tasks within very simple meaningful narrative contexts substantially enhanced performance, and heightened engagement and persistence (Parker & Lepper, 1992). In a second experiment, three factors—the provision of a fantasy contextualization for the tested learning activity, opportunities for incidental personalization of the activity, and opportunities to make choices regarding desired incidental fantasy features—produced marked increases in student motivation and depth of engagement in
learning, attitudes to the task, perceived task efficacy, and amount of learning (Cordova & Lepper, 1996).

**Methodological issues in gaming research**

The large majority of published studies found in the educational gaming research literature consist of case studies of game implementations undertaken at the K-12 or postsecondary level. These have largely relied on teacher and student self-reports as the sole source of data. Some have used open-ended surveys or interview schedules to probe perceptions about the game and its efficacy as a learning tool whereas others have collected more quantitative data using Likert-type rating scales. A few have employed standardized evaluation forms for user assessments (e.g., Becta, 2001). Although data of this type is of value in exploring usability issues and in determining attitudes and perceptions, it cannot be relied on to provide an adequate measure of any learning outcomes, for self-reports of all types are known to be subject to “halo effects”—when participants enjoy an experience, they are more likely to report having learned from it regardless of actual learning (Gosen & Washbush, 2004). One validation study found no correlations at all between perceived and objectively measured learning in a business simulation game for each of 10 different types of learning (Gosen & Washbush, 1999).

Case studies of the kind described have a second major limitation in that they can do relatively little to illuminate the specific cognitive practices and learning strategies students employ when playing the game. An understanding of game play and its relationship to the cognitive processes it evokes in users is essential for answering the question of how games succeed or fail, and it plays a critical part in untangling the complex relationships between various game attributes, the learning process, and learning outcomes.

A few researchers have employed experimental and quasi-experimental designs in combination with quantitative learning and attitudinal outcome measures to assess the learning impact of games or certain game attributes (e.g., Cordova & Lepper, 1996; Sherer, 1998). To the extent the outcome measures used are reliable and valid, the use of control groups and pre- and postgame testing makes possible stronger causal claims about the impact of games and game attributes. However, many of the potential educational benefits of advanced educational game play discussed above are not readily amenable to paper-and-pencil assessment. Quantifying and measuring such dimensions as the development of student self-management skills, the capacity to collaborate, or to abstract and transfer problem-solving strategies continues to be problematic. As these studies included no observation component they cannot capture any changes in student practice that might provide evidence for these desired developments.

A major threat to the ecological validity of lab-situated experimental studies of educational gaming lies in the short periods of game play that are typically studied. In a “normal” classroom, a complex simulation or role-playing game might be used over several weeks, interspersed with other related learning activities. This provides
a far different learning context from concentrated but short-term use in a usability lab or experimental classroom; the trajectories of knowledge growth, expertise development, and learning community evolution are likely to be substantially different in such divergent contexts. The strengths or limitations of a particular game or game design element may become apparent only after many hours of play. In addition, the constraints on data collection imposed by the use of a limited range of predetermined instruments reduces the likelihood of capturing any data on unintended or unanticipated game practices or outcomes.

A few case studies of educational gaming have been undertaken that incorporate observation of the game-play process either directly or indirectly (through audio and video recordings) in naturalistic settings such as classrooms and school computer labs (e.g., Henderson, Klemes, & Eshet, 2000; Squire, 2005). This methodology has a number of strengths. It allows the direct observation of the effects of a range of advanced gaming attributes (the provision of authentic learning challenges, powerful narratives, the assumption of new identities and roles, immersion in high-resolution virtual environments, etc.) and interface features on user behavior and learning in a setting in which the game would typically be used for learning, ensuring a high level of ecological validity. Case studies of games can document the shift away from novice understanding to the development of expertise by mapping changes in knowledge acquisition and application, pattern recognition, strategy deployment, and metacognitive functioning. Individual differences in game experience and performance can be examined in detail. The impact of game players’ social and educational world on their gaming experiences and outcomes can be explored in a way not possible with other designs—critical factors such as cooperation or competition with peers, interactions with the teacher, the curricular and instructional frames in which the game is presented, and the game-related pedagogical supports and learning opportunities (such as debriefing) provided by the teacher.

One variant of gaming process case study has employed the use of a cognitive task analysis technique known as the “think-aloud protocol” in which participants verbalize their thinking processes (e.g., Hong & Liu, 2003). This technique has proven effective in understanding situated cognition and decision making (Gordon & Gill, 1997) and can be used to map the growth of expertise and unpack the strategies and misconceptions that game players work with during game sessions. It has also been employed in the usability assessment and the formative evaluation of games and simulations to understand how the learner interacts with the user interface and other design elements (Kushniruk & Patel, 2005).

However, intensive case studies of the kind just described have two significant (and related) disadvantages. They are extremely costly in terms of the researcher’s time, requiring extensive on-site observation during game play and long periods for the qualitative coding and analyses of the data collected. Consequently the sample size for a given study tends to be very small—one classroom observed over a period of several weeks is typical. Although game play is studied in depth for a relatively long period of time, the unique nature of the specific class of students working with a specific teacher makes generalizing from the conclusions of one case problematic. This is especially
true in an educational context, as it is well established that the impact and outcomes of any pedagogical innovation are very dependent on the nature and quality of the specific implementation, including the teacher practices employed and a range of other contextual variables (Bransford et al., 2000). In addition, the obtrusive nature of the researcher’s presence in the classroom or lab setting—especially in conjunction with the use of video cameras for recording—can also reduce the naturalism of the setting and impact student behaviors in unpredictable ways.

The Virtual Usability Laboratory

To address many of the methodological limitations of the game study designs discussed above, we have developed a tool for collecting the kinds of rich and “thick” game-play process data we have argued is needed to answer some of the most important research questions about advanced educational gaming. In its initial formulation the tool was used for evaluating the usability of health care information systems—hence its name, Virtual Usability Laboratory (VULab; Kushniruk & Patel, 2004). It has since been modified and extended for use in researching and evaluating educational gaming environments. More specifically, VULab is designed to work with any computer game provided the computer is simultaneously connected to the Internet. The tool makes it possible to remotely capture process data in the everyday classroom context using unobtrusive techniques that do not require researcher presence or the use of extra devices such as video cameras or specially modified computers. This capability makes possible the remote administration of naturalistic field trials or experiments in situ, maintaining a study’s ecological validity and virtually eliminating the potential for the confounding of results due to Hawthorne effects or other artifacts generated by the presence of researchers or external recording equipment. Data can be collected from game players before, during, and after game play without any researcher intervention. The tool captures a “virtual videotape” of game play using screen recording software on the server together with an optional synchronized digital audio recording of player verbalizations. The latter makes possible the recording of talk-aloud protocols during gaming. The use of a microphone also permits the recording of student–student and student–teacher interactions that comprise an important part of the student’s gaming experience.

VULab resides on a Web server and does not require any special software to be installed on the remote game player’s computer. The tool is made up of four interacting system components: a recording component, a user presentation component, a researcher component, and a relational database component.

The recording component captures the screen and audio of the remote user in a synchronized and integrated media file. The recording is currently accomplished using the Macromedia Breeze software application (2006), although the authors plan to adopt an open-source solution or develop one of their own if an open-source solution cannot be found. The user component controls the presentation of sequenced online questionnaire forms that users are to complete. Typically these forms are
presented to the user when a user logs into VULab before playing a game; however, they can also be presented on completion of a game (or at researcher-selected points during game play for Web-based games).

The researcher component allows the researcher to set the parameters of a study without having to know the details of the implementation of VULab. The researcher can define the online forms that the user will see or choose from several predefined forms within VULab and determine when they will appear in the game-playing session. These forms may contain demographic questions, questions about pregame expectations, baseline knowledge or skills, postgame usability questions, content or skill tests seeking evidence of game learning, and/or other evaluation questions. The questions asked can be open ended, short answer, or multiple choice. This component also provides an interface to the relational database component to allow the researcher to query the database using predefined or custom Structured Query Language (SQL) commands. It is important to note that VULab does not do any analysis for the researcher. Instead, VULab provides the researcher with a raw data set that can be exported for subsequent analysis with external tools. The videos can be analyzed with qualitative data analysis software such as Transana (2006) or Atlas.ti (2005). These tools allow the researcher to add time-based codes to the video, labeling and categorizing events of interest for the data analysis using standard qualitative research coding techniques (see, e.g., Bogdan & Biklen, 1998). For example, a researcher might assign a “peer reinforce” code to all observed instances of peer reinforcement (as when one student compliments another’s solution of a game-presented problem). Qualitative coding of open-ended questionnaire responses can also be developed. Quantitative questionnaire responses can be analyzed using any common statistical package such as SPSS.

The relational database component stores the data from online questionnaires and links this data to the screen recording video files (which also reside in the database) by participant identifier. Using the researcher interface, it is possible for the researcher to retrieve the videos of users based on their responses to any combination of answers on the questionnaires. For example, to conduct an analysis of gender differences in gameplay patterns, the researcher could query the database to retrieve all of the videos for females and then all of the videos for males (assuming the researcher asked players their gender in a questionnaire). Or one could query the database to retrieve the videos of all users who answered particular postgame questions incorrectly. In addition, video segments coded using qualitative analysis software can be easily retrieved from the database for further comparative study and theory building.

VULab offers a number of other advantages to the researcher. Its capacity for remote deployment and administration makes it much more feasible to evaluate game use outside of major urban centers and in remote areas where travel costs may be prohibitive, greatly expanding the repertoire of naturalistic settings available for study and potentially the generalizability of study findings. In addition, by not requiring the researcher’s presence on-site during data collection, it greatly reduces the cost per site of collecting rich data, making it possible for more sites to be researched within a given budget—also potentially increasing study generalizability. Finally, the rich data it
provides makes it possible to uncover unanticipated game practices and outcomes and will be of great value in understanding the relationship of gaming processes and practices to outcomes that are assessed either via VULab itself or using additional measures.

**VULab field testing**

A field test of VULab was conducted to assess its functionality, usability, and potential for gathering game-play data of value to game researchers in a realistic educational context (a university course). An important goal of the trial was to test the feasibility of the tool in an authentic setting, using computers of the type and capacity typically found in university and school labs to play a game that was directly integrated with course curriculum. To meet these requirements at a reasonable cost and in a timely manner, a Web-based educational game shell package to which custom course-related content could readily be added was used to develop the test game. The game developed could be played through virtually any computer with an Internet connection and sound capability; no 3D video functionality was required (as this is often not found in school computers). Although the constraints of the game shell resulted in the creation of a simpler genre of game with a shorter period of play than would be optimal for most educational purposes, our focus in this project was primarily on the development of VULab itself, and resources were not available to develop or acquire multiple copies of an advanced game. It was felt that despite its educational limitations this simpler game would provide the basis for a meaningful (if preliminary) test of VULab’s potential as a gaming research tool.

**Procedure**

The field testing was done in a computer laboratory at a university in a large Canadian city over two separate sessions in which students played an educational game, and the VULab was used to record the gaming episode and present pre- and postsession questionnaires. All of the students approached to participate in the trials were enrolled in a first year School of Information Management course. The classes from which 42 volunteers were solicited had a total of 80 students. The students had previously completed lab exercises on using the Internet and common software packages, so no instruction on using the computer (navigating, links, etc.) was given. The field tests were conducted in two of the students’ regularly scheduled lab periods. In the first period, students played in competitive groupings of two to four with each group playing on one computer, and 10 game play sessions were recorded using VULab. Because of technical issues with microphone use in this first test session, the audio was too faint to be usable. The second test provided audio and video data and yielded eight game sessions from groups of two to three students each.

The Web-based game shell employed, called TRIVIA, is modeled after the popular board game, Trivial Pursuit (2002; see Figure 1). This game shell is one
component of a set of Web-based frame game shells called EDUCATION GAMES CENTRAL (2005). The categories and multiple-choice questions presented to the students in the game were written and entered into TRIVIA by one of the research team members who has taught this course. The content was based on the material that the class had been learning in the weeks leading up to the test. The 60 questions created were ranked as “easy,” “intermediate,” and “difficult” and were presented in order of difficulty, with easier questions given first. The game shell allowed for the easy modification of all of the instructions the users would see prior to starting the game, the questions, the responses that would be supplied to the players for correct and incorrect answers, and to some extent the order that questions would appear during the game.

Play was commenced by students taking a turn “rolling the dice” by clicking on the dice icon. The game then automatically advanced the game token around a multicolored game board. A question relating to the category indicated by the color of the square they landed on was presented. The player had 40 seconds to correctly answer the question; if successful, they would receive a token corresponding to the color of their square. The winning player was the first one to gather all of the six colored tokens and then answer one additional question from the category of his or her choice.
The students were presented with short questionnaires immediately before and after playing the game. These questionnaires were remotely “popped up” on the students’ computers by VULab. The first questionnaire asked for basic demographic information, such as sex, age, and computer skills. The last questionnaire queried players about the clarity of the instructions given, whether or not they thought the game supported the course material, and asked for suggestions for improvements. The students’ responses were collected and triangulated with the other data gathered by VULab. Students were instructed to use a think-aloud protocol during game play to verbalize their experiences with and reactions to the gaming experience. Student talk was captured by VULab by means of small microphones connected to the PCs.

Data analysis

A grounded theory approach was used to develop emergent themes from the game play and audio recordings. In this approach analytic findings and theory are “inductively derived from the study of the phenomenon. . . . That is, [they are] discovered, developed, and provisionally verified through systematic data collection and analysis of data pertaining to that phenomenon” (Strauss & Corbin, 1998, p. 23). To facilitate data coding and analysis, all of the files (including the video and audio recordings) were analyzed using Atlas.ti v.5 qualitative software using procedures described by Pandit (1996).

An initial case was selected for analysis, and using emergent codes developed from this case, we examined all of the available cases, with codes being added and modified as appropriate. The codes developed covered three broad dimensions of user experience: technical issues, usability issues, and play patterns. In the last stage of the analysis, relationships between the questionnaire data and the different dimensions of game play were explored.

Results

Analysis of the data illuminated several technical and usability issues that players had with the game. Although the students indicated in questionnaire responses that their game-playing experience was an enjoyable and interesting way to review class material, the game-play video and/or audio recordings revealed that they were confused and/or frustrated at different points throughout the game session. Furthermore, these frustrations typically occurred at the same points in the game for the different teams.

Technical issue. VULab allowed us to pinpoint a technical issue related to a scripting error in the game that three teams experienced while playing. To win the game, the winning player had to correctly answer one additional question, in a category of his or her choosing. In all three failing games, the screen froze 7 s after the player had selected the final category, and the same error message was displayed each time. A review of the information in the “Script Error” code helped to identify and explain the
reasons for the difficulties common to the three teams who were unable to finish the
game. The error message indicated that the problem was an incompatibility with the
early version of Internet Explorer being used on these machines, and the game froze.
Through our recordings programmers are able to see all of the steps leading up to the
error message and can easily recreate the environment under which the error will
occur. Although 30% of the users experienced this error in the first test, none experi-
enced it in the second test as they were in a different lab.

Usability issues. There were three usability issues that emerged from the data. They
are discussed here in the order of their occurrence as users progressed through the game:

1. Load button access. The button to load the game was at the far bottom of the screen,
and users sometimes failed to scroll down far enough to see it, causing some confusion
about how to start the game. The recorded discussions verified that the players were
confused about what they should do to start the game.
2. Start button selection. The game’s onscreen startup instructions appeared to be unclear
to students because of the presence on screen of two potential “Start” buttons. The
instructions asked players to “select the Start button to commence play”; however, the
correct button was labeled “Click to Start,” which added to the players’ confusion.
3. Game instructions access. Onscreen information presented to players by the game
explained the context of the game and not how to play it. A link to the playing instruc-
tions was present; however, only one of the 18 teams reviewed the instructions prior to
playing the game.

When the first game question had been navigated successfully, the teams did not
experience many usability problems as they continued playing the game, and in fact,
most teams quickly mastered the procedures for rolling the dice, presenting ques-
tions, and entering answers.

Play patterns. The audio and/or video recordings proved invaluable in reconstruct-
ing users’ experiences while playing the game. We found that most players clearly
enjoyed interacting with their course material through the game; however, the usabil-
ity issues did interfere with flow of the interaction.

1. Player identification. Although the game had not been set up to offer each player the
option of entering his or her own name, the users were able to select a name to use
while playing the game from a list of the developers who had worked on the game
shell. This turned out to be significant for students, who chose a player name with con-
siderable care. They seemed to identify with the names and appeared to use their new
names to enhance their game experience. In several cases, they referred to each other
by their new, assumed names and continued to do so until the game had finished. We
further noted that one female student decided to choose a male name. The game’s
developers had not anticipated that players would choose to select names, or that
assuming a new identity would be important to their experience of the game.
2. Cooperative play. Although the game was ostensibly structured for competitive play, in
almost every group the members jointly answered the questions for each player and
shared in celebrating correct responses. It is interesting to note that, however, at the end
of the games, the players who won the game were congratulated by their teammates
and claimed ownership of the win.
Discussion

The field test results clearly demonstrate the feasibility and value of VULab for the unobtrusive capture of data central to understanding many of the key elements of educational gaming without requiring that game play be artificially isolated from the educational context for which it is designed and in which it is normally used—the lived social world of the classroom. The well-documented importance of a range of contextual factors in determining educational outcomes (e.g., Bransford et al., 2000) and the evidence for their significance in educational gaming (e.g., Becta, 2001) highlight the criticality of being able to research and evaluate educational games in conditions with high ecological validity. In the case of more advanced games that students might use for several weeks, VULab makes it possible to study the entire arc of game play as it unfolds and is integrated into classroom life.

The trial showed that VULab can operate unobtrusively. After startup, the screen and audio recording were invisible to the user and did not change the appearance or functionality of the game in any way. No interactions with the tool were required of players during game play beyond those needed to respond to questionnaires. Neither the presence of a small table microphone nor the request that students verbalize their thinking as they played their game appeared to inhibit play in any way, and the synchronized audio recordings generated valuable insights into obstacles players encountered and their patterns of competitive and cooperative play. The remote, unobtrusive, and transparent nature of VULab’s operation makes possible the collection of extremely “thick” data without significant risk of triggering Hawthorne effects or other methodological artifacts that threaten the validity of more obtrusive techniques.

The field trial also demonstrated the value of VULab for documenting unintended or unanticipated processes and outcomes during game play. The audio recordings revealed the surprising meaningfulness of identity and role in a game where role-playing was not a design element, a finding that is in accord with Cordova and Lepper’s (1996) discoveries about the positive impact of opportunities for personalization on children’s attitudes toward and success with computer puzzle games. They also showed the unanticipated predominance of a cooperative ethos during what was intended to be competitive game play. Findings of this type are crucial to research examining the basic processes of game play in different educational contexts, and to usability studies and formative evaluations focused on determining how a game’s design, architecture, and rules will be interpreted and appropriated by players. The tool’s capacity to record the conversations of a small group of players makes it possible for the researcher to capture much of the social discourse around game play that is central to situated learning and the functioning of communities of practice (Kirshner & Whitson, 1997; Lave & Wenger, 1991) and the collaborative nature of collective knowledge building (Scardamalia & Bereiter, 1992); these interpersonal aspects of game learning. Social interaction around gaming is often substantial (Mitchell & Savill-Smith, 2004), and its role in mediating learning from educational games cannot be ignored.
Through its relational database, VULab facilitated the triangulation of audio and screen recording data with survey and questionnaire data. The value of this capability was highlighted in the trial—students did not mention encountering any play problems when asked about play difficulties on the response forms presented at the end of the game play period and made no suggestions for improvements, and yet our analysis of the screen and audio recordings revealed that they had encountered several operational difficulties in the initial stages of game use. VULab’s capacity to capture real-time play data makes it possible for the researcher to transcend the well-known limitations of relying strictly on postuse surveys for illuminating user experiences and perspectives (Shneiderman & Plaisant, 2004).

VULab offers researchers and evaluators several other potential affordances that this pilot trial did not attempt to assess. By collecting rich process data it makes feasible a more thorough exploration of the causal relationships between specific game-play processes and practices and game-learning outcomes. An understanding of these relationships is crucial to improving the efficacy of educational gaming. In addition, it makes it far easier to assess the practicality and utility of a game or simulation across a range of real-world educational settings. Games that foster extensive learning in atypical demonstration sites or laboratory schools may not necessarily be practical or robust enough to produce the same exemplary outcomes in typical school cultures due to differing student backgrounds and prior experiences, or a lack of sufficient pedagogical support in the classroom. Cultural differences can also affect the interpretation of game content, design, and images (Asakawa & Gilbert, 2003). With VULab it should possible to closely study game use at reasonable expense in widely dispersed locations and highly divergent contexts.

Finally, VULab when it is more fully developed has the potential to facilitate student learning in a more direct fashion. If made available to teachers, it could be used to deliver supplementary learning materials of any type before, during, or after game play, providing a channel for scaffolding and support at key points during the game. Students could be encouraged to externalize their reasoning and problem-solving steps at critical decision points in the game having them respond to pop-up game journal questions. Similar techniques used in other learning contexts have proved effective in developing domain expertise (Bransford et al., 2000). Teachers could access this potentially rich source of formative data on student knowledge and strategies in real time through the VULab database, making it possible to quickly and precisely determine which students need support to correct important misconceptions or advance their learning strategies.

Further development of VULab is being undertaken to extend its data collecting capabilities to other game architectures and a variety of other educational computing applications, to improve the functionality of its researcher user interface, and to further simplify its setup and use on remote computers. Following more field testing, it will serve as the main data collection tool in a planned series of studies incorporating the use of a range of advanced educational games.
Notes

1. The term computer gaming is used in this article to refer to the playing of computer and video console (e.g., Xbox) video games. For the sake of brevity and clarity of writing, the broader terms game and gaming are used here refer only to computer games and gaming (unless otherwise specified).

2. Following Dempsey, Hayes, Lucassen, and Casey (2002), a game is defined here as a rule-guided set of activities involving one or more players which has goals, constraints, payoffs, and consequences; is artificial in some respects; and involves some aspect of competition (with self or others).

3. “Advanced” educational games are those that go beyond the historically predominant types of games employed in K-12 education in which exogenous educational content is entered into a game shell to provide factual or procedural drills (e.g., a racing game in which players advance a car by correctly answering factual or skill-testing questions). In advanced “endogenous” games, players learn the properties of a virtual world through interacting with its symbol systems, learning to detect relationships among these symbols, and inferring the game rules that govern the system.

References


Herbert H. Wideman has research interests in a number of areas related to technology-enhanced constructive and experiential learning, and its implementation in educational systems. He has published studies and conducted evaluations on such topics as student expert system development, research and evaluation methodologies for the study of educational technology, and online and blended forms of teacher professional development. He is currently collaborating in an investigation of the impact of student game development on literacy. He is a researcher in the Simulation and Advanced Gaming Environments (SAGE) for Learning research network (http://sageforlearning.ca/).

ADDRESS: Institute for Research on Learning Technologies, York University, TEL1029, 4700 Keele St., Toronto, Canada; Telephone: +1(416)736-5019; Fax: +1(416)736-5913; e-mail: herb@yorku.ca

Ronald D. Owston specializes in the evaluation of technology-based learning, and is domain leader for methodology and tools in the SAGE for Learning research network. He has led many major research projects sponsored by agencies such as the Social Sciences and Humanities Research Council, Health Canada, Ontario Ministry of Education, IBM Canada, and Apple Canada.

ADDRESS: Institute for Research on Learning Technologies, York University, TEL1029, 4700 Keele St., Toronto, Canada; Telephone: +1(416)736-5019; Fax: +1(416)736-5913; e-mail: rowston@edu.yorku.ca

Christine Brown is a doctoral candidate in education at York University who teaches business-related subjects at Ryerson University. Her research interests include distance and hybrid educational pedagogy, technology enhanced learning opportunities, and online support for medical patients and caregivers.

ADDRESS: Institute for Research on Learning Technologies, York University, TEL1020, 4700 Keele St., Toronto, Canada M3J 1P3; Telephone: +1 (416) 736-5019; e-mail: christine.brown@edu.yorku.ca

Andre Kushniruk has been a key researcher on a number of national and international collaborative projects in medical informatics, and is a collaborator in the SAGE for Learning research network. He conducts research in a number of areas including evaluation of the effects of technology, human-computer interaction in health care and other domains and cognitive science.

ADDRESS: School of Health Information Science, University of Victoria, PO Box 3050 STN CSC, Victoria, Canada V8W 3P5; Telephone: +1(250)721-8578; Fax: +1(250)472-4751; e-mail: andrek@uvic.ca

Francis Ho practiced family medicine in Ontario for 26 years before devoting his full time to medical informatics research. He is an experienced web master and programmer. His research interests include Internet usability studies, consumer health informatics, palliative care informatics and medical data translation.

ADDRESS: School of Health Information Science, University of Victoria, Victoria, Canada V8W 3N7; Telephone: +1(250)472-4731; Fax: +1(250)472-4751; e-mail: fho@uvic.ca

Kevin Pitts has had fun playing games (electronic or otherwise) and learning from them for a number of years. In his role as a faculty advisor in the eLearning Centres at Seneca College, he has had the opportunity to tinker with gaming-related educational technology and is a strong believer in the value of gaming in education.

ADDRESS: eLearning Centres, Seneca College, TEL1035, 70 The Pond Road, Toronto, Canada M3J 3M6; Telephone: +1(416)491-5050; Fax: +1(416)650-9169; e-mail: kevin.pitts@senecac.on.ca