

Digital Dome Versus Desktop Display in an Educational Game: Gates of Horus

Jeffrey Jacobson
PublicVR, USA

ABSTRACT

A visually immersive display can make an educational game more effective, if concepts are embodied in an information-rich space and the interaction/perception design exploits the egocentric view that the display affords. For example, ancient Egyptian temples juxtapose language and symbol in an architecture meant to be understood from the inside. Students, ages 11 through 14, played an educational game based on our (virtual) temple. In an unstructured test in which students produced their own guided tours of the temple, those who had played the game in a digital dome showed better factual recall ($P < 0.05$) than those who had used a standard desktop.

Keywords: Educational Games, Virtual Reality, Cultural Heritage, Empirical, Immersion, Museum, Egypt, Spatial, Dome

INTRODUCTION

We live in a time of rapid advances in the use of communication technologies for education. (Dede, 2009; Mihalca, 2007). This includes *virtual reality* (VR), computer-based media which employ a three-dimensional virtual world. Students can interact with this imaginary space and things in it, psychologically “inhabiting” that world, as themselves or as some other personas (Winn, 2003; Slater, 2009). Researchers are working to develop curricula and learning environments that take advantage of the opportunities provided by these new media (Takacs, 2008). One key strategy is to cast the learning activity in the form of a *game* (Dondlinger, 2007; Squire, 2008; Anderson, 2010; Dede, 2009). Nevertheless, there is still controversy over whether the choice of media makes any fundamental difference in learning (Clark, 2010; Russell, 2000). If not, the educator’s choice of media should be an economic one. Either way, research is needed to determine what works and how to use it.

The purpose of this case study is to contribute data to the debate by showing one way that a *visually immersive display* can support learning in game-like educational virtual reality. Visually immersive displays provide the user with the ability to look in most directions and see the virtual environment. The most straightforward example is the all-digital projection dome which can display a virtual landscape in a panoramic view up to 180 degrees wide (Lantz, 2007). The CAVE and similar devices also surround the user with a panoramic view, but they are much smaller, generally square, and usually focused on an individual user, providing that user with sophisticated interaction devices (Cruz-Neira, 1993). Another approach is the Head Mounted Display (HMD), a pair of glasses in which each lens is a video screen showing the virtual

environment. The computer employs a tracking device on the user's head so that it can show a stable view of the virtual environment in any direction the user looks.



Figure 1. *The Virtual Egyptian Temple in the Earth Theater*

In our study, we employed the Earth Theater at the Carnegie Museum of Natural History (CMNH, 2010). Its main screen is a section of a sphere, providing a view 210 degrees horizontal by 30 degrees vertical. Specifically, middle school students played *Gates of Horus*, an educational game based on our Virtual Egyptian Temple (Troche, 2010). In Jacobson (2009), we describe in detail *Gates of Horus* and our experimental findings, which indicate that students learned from the game and found it enjoyable and engaging ($P < 0.05$). In this article, we describe our case study, in which individual students who played *Gates of Horus* in the Earth Theater had better factual recall ($P < 0.05$) regarding the temple than those who played the game on a standard desktop computer.

In this article, we will situate our study in the literature, describe *Gates of Horus*, describe our experiments, and discuss our findings. We believe that we have found a case in which visual immersion does make a positive difference, although more research is needed to find the mechanisms at work.

Does Media Matter?

Clark (1988) challenges the idea that learning outcomes can be affected by communication media used in the learning process. He likens the media to a delivery vehicle and the information to be learned as the payload. To him (Clark, 1994), the only important difference among media is their cost, and therefore the most appropriate and efficient one should be chosen for the task at hand. Russell (2000) cites a very large number of studies that attempted but failed to find any significant differences in learning when comparing two or more media employed in the same or very similar curricula. Today, Clark (2010) continues to cite a lack of hard evidence of media-related differences in learning and a lack of good experimental studies. Instead, he focuses on theories and methods for good instructional design.

However, Russell himself states that the lack of significant media-related differences in learning is more likely due to inappropriate or incomplete use of the media (Russell, 2000, pp. xii-xvi) or weak means of measuring difference. In refuting Clark's claim, Kozma (1994) begins by stating that media and method (instructional design) are inseparable. Good design takes advantage of the affordances offered by the media, and the media exists to serve the design – one cannot simply substitute one communication media for another and hold all else equal. Mayer (2001) says much the same thing.

Kozma (1994) goes on to argue that learning is not a process of simply accepting delivery of information, but an active process of exploration. Central to that exploration is a reciprocal relationship between the learner's cognitive resources and the external environment. Winn (2003) describes learning as an act of mutual adaptation between the student and the environment. Constructivist learning theory (Jonassen, 2000) describes learning as the individual or social process of building knowledge (Vygotsky, 1978) which requires a supportive environment.

Today, most educational research focuses on learning environments (Mihalca, 2007) that support the learning process in increasingly intelligent and flexible ways. It is interesting that Clark, Kozma, and other central players in new media for education (e.g., Barab, 2009) all state that (apart from individual learner differences) it is the quality of the design of the learning experience that matters. Perhaps the main difference between Clark and the others is a difference of opinion on the extent to which media creates meaningful opportunities for better design.

Educational Games

Curriculum can be cast in the form of an educational computer game, a strategy that has gained great currency in recent years (Dondlinger, 2007; Mihalca, 2007; Dede, 2009; Anderson, 2010). At minimum, a game is an activity that (1) has some goal in mind, something that the player works to achieve, (2) has rules, and (3) is considered a form of play or competition (Oxford, 2010). While this encompasses "skill and drill" types of games, most are much more complex, providing flexible and responsive narrative in which the player must test hypotheses, synthesize knowledge, and respond to the unexpected (Dondlinger, 2007). The fundamental goal is to motivate the student to learn something (Ang, 2008), but successful educational game design is not easy (Baker, 2008).

According to Ang (2008), the student can be motivated by *ludology* and *narrative*. Ludology is the iterative and creative competition or work toward goals, focusing on the game play itself. Narrative is the story itself, which the designer hopes will capture the student's attention because it is interesting. Within narrative, the student can embody a character or persona (Barab, 2009; Dickey, 2006; Squire, 2008; Slater, 2009). The student (as the character) must understand the virtual context and ultimately transform it. In good game design, goals are structured flexibly, so that the student can rise to a level that challenges his or her abilities (Dondlinger, 2007). The goal is to keep the student in his or her "zone of proximal development," in Vygotsky's terms.

However, an educational game must motivate the student to learn what it was designed to teach, not just to play or "win" the game itself. Games can motivate through *extrinsic* or *intrinsic* rewards and goals (Dondlinger, 2007). Extrinsic rewards, such as earning points or encountering something fun or pretty, are defined in the structure of the game itself, but they do not have any direct relationship with the material the student is expected to learn. Extrinsic rewards can

engage the student, but they also may interfere with learning. According to Fisch (2005), when appealing elements are added to keep students interested, the students often remember those appealing elements and forget the content they were supposed to learn. Intrinsic rewards, on the other hand, are situated within the educational content itself. For example, in Dede's educational game, *River City*, the goal of the game is to figure out what is making the townspeople sick, a perfect union of game and content (Dede, 2005).

Gates of Horus

Our game, *Gates of Horus* (Jacobson, 2009; Anderson, 2010), is an example of the growing trend for using serious games to teach cultural heritage (Anderson, 2010; Champion, 2008). *Gates of Horus* is based on our virtual Egyptian temple. The temple has no real-world analog, being an exemplar of its type, designed to embody the key features of the typical Egyptian temple of the New Kingdom period (Troche, 2010). Its four major areas, from front to back, are the exterior (Pylon), the Courtyard, the Hypostyle Hall, and the inner Sanctuary. The formal Egyptian temple was a focus for community gatherings (McDowell, 1999, pp. 91-104) and an important cultural and economic multiplier (Kemp 1989, pp. 193-97). Egyptian temples, their decoration, and content were arranged according to a deliberate spatial plan (Watterson, 1998, pp. 35-43), which makes them appropriate topic for 3D media.

To play *Gates of Horus*, the student navigates the temple in a first-person view (Bowman, 2002) on a standard desktop or in an immersive display (Figure 2). The game works with a standard three-button mouse, but in our study all students employed a Gyromouse (Jacobson, 2008, p. 100). The latter functions like a standard mouse, except that an internal gyroscope allows the student to rotate view in the game by rotating the mouse itself, instead of moving it across the desktop. Pilot testing (Jacobson, 2008, p. 121) showed us that this feature removes the need for the desktop, making the mouse reasonably comfortable for use while the student is standing (Figure 2). The user moves forward in the immersive display by pressing the left mouse button and backward by pressing the right mouse button. Students using the Gyromouse while seated at a desktop also did not report any problem with the interface.)

When the student presses down on the mouse wheel, the game switches to "selection mode," in which the mouse controls the cursor shown in Figure 2. The student can then select an active object by moving the cursor over it in the current view and pressing the left mouse button. The student can select the same object from any direction. The cursor is seen on the surface of the display and turns green when it is over an active object or feature of the temple. To return to navigation mode, the student presses the mouse wheel again.



Figure 2. A student and the game, *Gates of Horus* in the Earth Theater. During actual play, she would usually look at the cursor, the white cross in the upper right corner.

The game begins at the front of the temple. Each time the student selects an “active” feature of the temple, such as the hawk, the priest explains its nature and meaning. When the student clicks on the priest, he asks the student a question based on what the student has selected previously. The priest asks questions in random order, never repeats a question the student has already answered correctly, and does not ask questions associated with a feature he has not yet explained to the student. For example, when the student selects the winged disk above the main gate of the temple, the priest says, “The winged disk...represents the divine life-force which flows from heaven and into all things....” This adds three questions to the priest’s pool of questions, such as, “Does the winged disk represent war?” The answer is “no.”

The student must answer all of the priest’s questions correctly about one area of the temple to progress to the next area. When the student has given correct answers to all the questions, the Gateway to the next area of the Temple opens, and the student explores that area and learns about it in the same way.

The priest asks a few additional questions from the introductory narrative of an area after the student answers all other questions associated with that area. The student “wins” the game when he or she answers correctly all of the questions from the priest in the inner Sanctuary, causing the divine image of the God to “speak,” bringing the blessings of heaven to the land of Egypt. For a complete description of the rules of interaction, see Jacobson (2009). For the complete dialogue, see Jacobson (2008, p. 104). A working copy of the game is currently available at PublicVR (2010).

Motivation in *Gates of Horus* is primarily ludic (Ang, 2008), dependent on the structure of the game, rather than on the narrative. (The student is supposed to be a young prince being educated to eventually become Pharaoh, but we barely mentioned it.) The reward for continuing play is the chance to see more of the temple and learn more about it – an intrinsic reward (Dondlinger, 2007). This appears to have worked, because all students in the study concentrated intensely on the game for nearly an hour, and their feelings toward it were generally positive (Jacobson, 2009, p. 144). Furthermore, the narrative setting parallels the Egyptian emphasis on learning and literacy as a key virtue and the role of the senior priests as instructors (Sauneron,

2000). Unifying the game and the content was key, because authenticity is important in educational games for cultural heritage (Champion, 2008).

Visually Immersive Virtual Reality

We are particularly interested in educational applications that employ both a virtual environment and a visually immersive display (as we defined them here). But to survey prior work in this area, we must make the reader aware of a shift in terminology.

Before the year 2000 (approximately), all applications called virtual reality (VR) used a visually immersive display or a flight simulator (Ellis, 1991). Since that time, the term has expanded until VR refers to nearly every application based on an interactive three-dimensional space, including most of the interactive games used in educational research (Dondlinger, 2007). This is a matter of some concern among longtime VR researchers (Jacobson, 2010) but not an issue we want to address here. In the following discussion, visually immersive virtual reality (VIVR) refers to all VR applications that employ a visually immersive display.

The 1990s were a time of great excitement about the educational potential for VIVR. Hundreds of educational projects employed visually immersive displays, and there are still many examples in the decade afterward. Despite this, we can find only 14 formal (statistical) experimental studies on the educational effectiveness of VIVR (Moreno, 2002; Winn, 1997, 2001; Jackson, 1999; Bowman, 1999; Dede, 1999; Roussos, 1999; Salzman, 1998, 1999; Osberg, 1997; Byrne, 1996; Rose, 1996; Jacobson, 2008; Bailenson, 2008; Limniou, 2007) and one notable study that employs activity theory for its analysis (Roussou, 2007). A major limiting factor has been the high cost of most immersive displays, but that has changed recently (Tacacs, 2008). Nevertheless, implementing and evaluating good VIVR educational design remains complex because of the large number of variables involved.

An important theme in VR research is presence, most commonly described as the feeling of “being there” in the virtual environment (Spring, 1992). Most articles we have read that discuss VIVR for education refer to presence, and nearly all of the 14 empirical experiments listed above test for it. They argue that presence focuses the student’s mind on the learning task and structures the interaction in a natural way. However, the concept of presence is complex, and has been poorly defined over the years, making it difficult to test for (Bailenson, 2008). Moreno (2002), Byrne (1996), and Rose (1996) all based their learning experiments with sensory presence as the independent variable, with no result.

Efforts are underway to create a more stable definition of presence. Slater (2009) makes a distinction between the type of presence achieved through sensory immersion and the engagement one develops toward a character by role-playing it in the context of some narrative. Others retain a broader definition of presence, but posit a more stable theoretical foundation (Chertoff, 2008). Slater (2009) provides a good overview of the topic. In our study we do test for presence, and we recognize its importance, but it is not our focus.

In two key experiments, a student appeared to gain an advantage from experiencing an egocentric viewpoint in the virtual environment, meaning he or she is (visually) inside something interesting (Dede, 1999). The view surrounds the viewer with information, making important relationships easier to read. Examples are a magnetic field (Salzman, 1999) or a simulation of Puget Sound with added visual representations of items, water temperature, flow, and salinity (Winn, 2001). Most of all, it is these two examples which motivate the premises of this study.

Accordingly, the hypotheses for this study come from the premises that (1) religious architecture is rich in spatially-organized information, including Egyptian temples (Watterson, 1998, pp. 35-43) and (2) studying the virtual temple is more effective with the benefit of an egocentric (inside) view.

METHOD

Participants and Protocol

All testing took place at the museum's Earth Theater (CMNH, 2010) with middle school students, ages 11 through 14, recruited from area schools, civic organizations, and individual families. The museum is located in Pennsylvania (USA), the study of ancient Egypt is part of the state education standards for middle school children, and the museum already had a regular schedule of Egyptian-themed field trips for them. Pilot testing confirmed students of that age were old enough to interact meaningfully with the game and were generally enthusiastic (Jacobson, 2008).

We used the first students for pilot testing over several weeks to debug the game and our protocols (Jacobson, 2008). After pilot testing and some loss from dropouts, we had data for 65 students. During the main study, as the students came in, we randomly assigned each student to one of three groups:

- **Theater Group** – Students played Gates of Horus using the immersive dome display.
- **Desktop Group** – Students played the game on a standard desktop computer in an area adjoining the main theater.
- **Control Group** – Students took the written posttest for basic knowledge of the temple *before* playing the learning game on a standard desktop.

The Control Group's Written Posttest takes the place of a knowledge pretest for the other two groups. This is necessary, because a question-and-answer pretest on knowledge of the temple would reveal too much about it. This could raise all student scores and mask the effect of the experimental treatment.

All students in the three groups completed the following protocol, with variations in steps 1, 3, and 10 for the Control Group:

1. **Pretest (for all students except the Control Group):** Students take a pretest for general information, such as their attitude towards Virtual Reality (VR) and Egypt.
2. **RPM:** Students complete a visual intelligence test, Raven's Progressive Matrices (RPM) (Raven, 1958; Gregory, 1999).
3. **Written Posttest (for the Control Group only):** To compensate for the knowledge the other two groups may have learned through the pretest in step 1, the Control Group takes this fairly typical multiple-choice and short-answer quiz.
4. **Playing the Game:** Students play the learning game, *Gates of Horus*, to completion. The software logs all activity for later analysis.

5. **Presence and Comfort Test:** Students are queried on their sense of presence (a single question), and their degree of motion sickness is determined through questions regarding common symptoms.
6. **Drawn Map:** Each student draws a map of the temple.
7. **Magnet Map:** Each student places small magnets representing features of the temple on an accurate map of the temple.
8. **Video:** Each student gives a 15-minute guided tour of the temple, navigating the model of the temple on a computer, explaining its purpose and features. The tour is videotaped and later scored by evaluators.
9. **Written Posttest:** Students who are *not* in the Control Group take the written posttest.
10. **Follow-Up:** One to two months later, all students complete an online quiz regarding the temple.

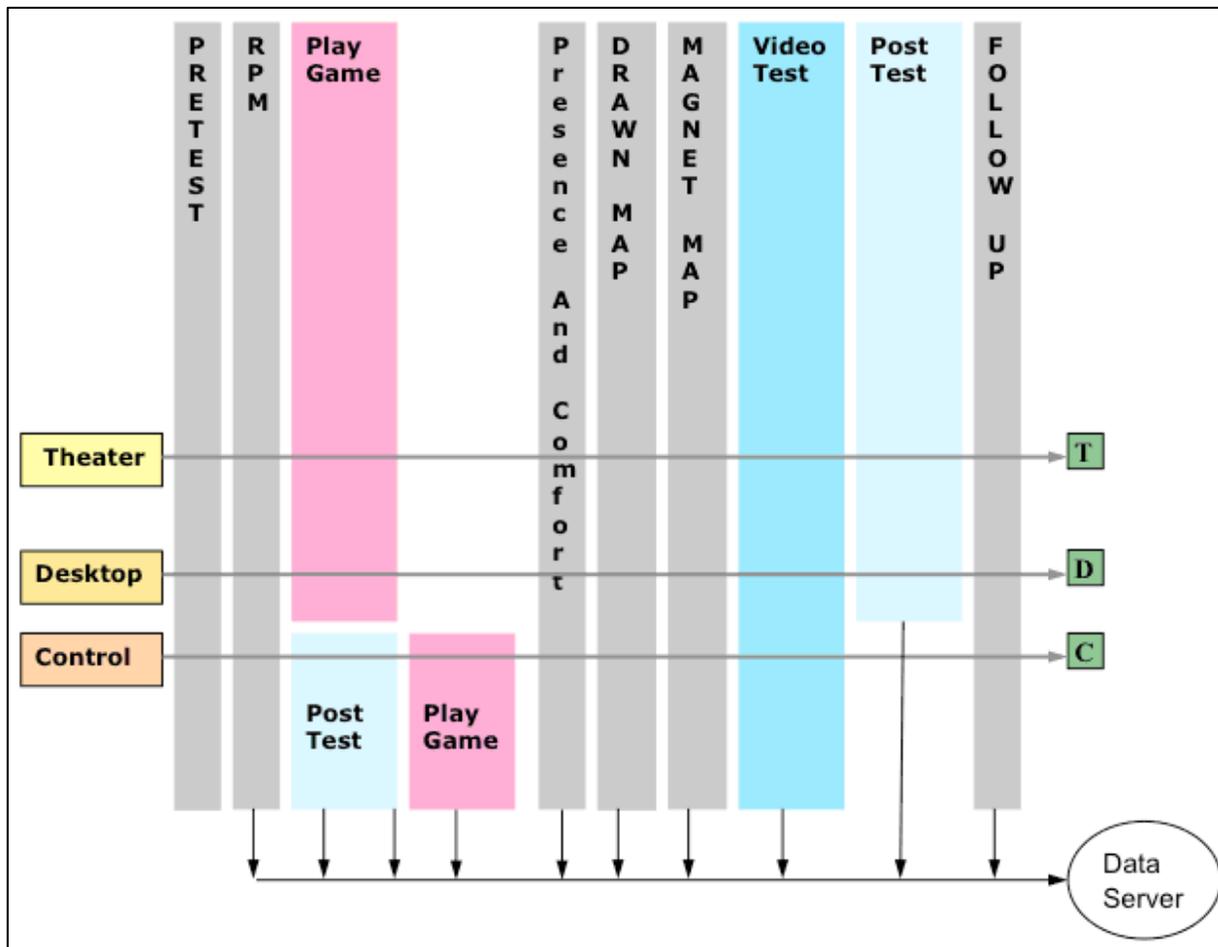


Figure 3. The experimental protocol

Hypotheses and Design

Two hypotheses, H2 and H3, embody the primary goals of the study. H1 plays a supporting role while H4, H5, and H6 refine the possible results found in H2 and H3. Many other comparisons are possible using additional measures (e.g., age, gender, computer literacy, navigation path in the virtual environment), which we gathered for later analysis, but they are beyond the scope of this discussion. In the following statements, “higher than,” “better than,” or “worse than” refer to statistically significant differences.

(H1) Students who play *Gates of Horus* to completion will demonstrate greater knowledge of the Virtual Egyptian Temple than those who have not played the game.

Specifically, scores on the written posttest for students in the Desktop Group will be better than scores on the written posttest for students in the Control Group (Figure 3). The comparison does not include members of the Theater Group, because they may gain an advantage from the immersive display.

This hypothesis is concerned only with whether the game is an effective learning tool, a necessary condition for comparing the effect of different displays. *Gates of Horus* is both a computer learning game and a VR application, and both those factors make this hypothesis risky, since good computer learning games are difficult to produce (Baker, 2008), and many VR applications fail to achieve their purpose (Jackson, 2006).

(H2) After *all* students have played the game (Figure 3), those in the Theater Group will demonstrate more factual knowledge than the others.

Scores on the video test for the Theater Group will be higher than scores produced by students in both the Desktop Group and Control Group, taken together. This strategy of combining results for the Desktop and Control Groups has the statistical advantage of comparing larger groups and is conservative with respect to this hypothesis. It gives the aggregate non-Theater Group an unfair advantage in the Video Test, because students in the Control Group had more exposure to the material by taking the written posttest first (Figure 3).

Additionally, if students in the Theater Group achieve higher scores in the written posttest than students in the Desktop group, those results will add further support to this hypothesis.

(H3) One to two months after playing *Gates of Horus*, students in the Theater Group will demonstrate more knowledge on the follow-up test than all other students (in the Desktop and Control Groups).

(H4) Students in the Theater Group will score higher on one or both of the map tests than all other students after they have all played the game (Figure 3).

(H5) One or two months after playing *Gates of Horus*, students in the Theater Group will report more motivation to learn about Egypt and a more positive view of their experience on the follow-up test than students who played the game on a desktop computer.

(H6) We expect students’ scores on the Raven’s Progressive Matrices (RPM) to interact in some way with the experimental results for the posttest, video test, or follow-up test.

Raven's Progressive Matrices (RPM) (Raven, 1958; Gregory, 1999) measures students' basic spatial and spatial-analogical reasoning ability without the use of language. One might expect those who score highly on Progressive Matrices to do better in the posttest results because of their higher spatial ability. The reverse may occur, however, with the students of lower spatial ability benefiting more from having the temple visualized for them by the software (Bricken, 1990). Nevertheless, there is some indication in a study by Winn (1997) that low-achieving students (i.e., with low grades in school) had more to gain, on average, from their VR learning application. That observation may apply in our study as well.

Threats to Validity

This section describes confounding or interfering factors and attempts to mitigate them.

Novelty Effect: At present, Immersive VR applications are rare and are usually praised by users as thrilling, fun, and exciting. This "novelty effect" may result in a higher level of user engagement (Clark, 1988), providing a transitory advantage for students in the Theater Group. However, museum staff stated that computer games, Egypt, and field trips to the museum are very popular with this age group. We hope that the experience for all students was novel and exciting enough to overwhelm any novelty effect caused by our immersive display.

John Henry Effect: (AKA Hawthorne Effect) The students in the theater group may have tried harder because they may have believed that the researchers wanted them to do better than students using desktop computers (Campbell, 1963). Our researchers were careful to present the experiment in a neutral manner.

Large Number of Variables: Most educational research in authentic settings (e.g., the museum) must contend with a very large number of potentially confounding variables, just as we did. Factors such as the weather, the attitude of the teachers towards technology, and the timing of school exam periods had potentially disruptive effects. Where a traditional experimental design is possible, the best solution is a large and randomly selected sample (Campbell, 1963) to evenly distribute interference among the test groups. The number of test subjects in this study (under 100) should smooth over enough of these interfering factors so that a statistically significant result is interesting and worth further investigation.

Motion Sickness: CAVE displays, digital domes, and similar devices provide optic flow in the user's peripheral vision, which may cause motion sickness (Kennedy, 1992; Kolasinski, 1995). In the Earth Theater, the panoramic view and the user's ability to control the selection/activation cursor each reduced the student's need to navigate to see and select objects. Less navigation reduced visual flow and therefore reduced motion sickness. Also, the navigation speed in the Earth Theater was much slower than in typical video games to reduce confusion and motion sickness and to convey a sense of the temple's grand scale. Finally, audiences see immersive movies in the Earth Theater on a regular basis with very few complaints of motion sickness. Immediately after playing the game, each student took the Presence and Comfort Test, which showed no significant discomfort in the immersive display.

Application and Interaction Design: Success of any VR application depends on the quality of its interaction design. Navigation has to be easy and comfortable, selection straightforward, and performance crisp. Many Immersive VR experiments have had major problems as a result of poor interaction design or inadequate technology (Jackson, 1999). To minimize these problems, the authors improved the interface and interaction for *Gates of Horus* according to proven design principles (Bowman, 2002) and further refined it during pilot testing (Jacobson, 2008).

Limited Quantity of Content: The content of the learning game must be finite for practical reasons. However, if there is too little to learn, a ceiling effect for learning outcomes could occur, in which all students in all treatment groups learn everything. Pilot testing showed that this was unlikely to be a problem (Jacobson, 2008).

Fixed Learning Time: For practical reasons, students must have a fixed period of time to explore the temple and learn the associated materials. This runs the risk of either forcing students to stop when they are engaged and learning or forcing them to continue when they want to stop. Recent studies have shown the learning in virtual environments can be time-consuming (Jones, 2008). Fortunately, the limited narrative structure of *Gates of Horus* allowed us to make it resolve within a limited period of time. After we did pilot testing and made adjustments to the software, most students completed the game in approximately 40 to 60 minutes.

Fatigue: Testing and treatment took approximately two hours for each student, which may have been tiring. Fortunately, students revealed a remarkable degree of energy and interest, with none of them asking for a break. Perhaps the variety of activities kept the experience interesting.

Measures and Materials

The **Pretest**, the **Written Posttest**, and the **Follow-Up** test used a standard multiple-choice and short-answer format, implemented as online (HTML) forms using SurveyMonkey® (2010), a survey authoring and administration service. When the student completed a test, the data was uploaded to a server for storage and analysis. Data gathered with the pretest could be analyzed directly, while the other two (the written posttest and the follow-up test) required scoring by evaluators. For the written posttest and the follow-up test, we created an individual grading form for each student's responses to the short-answer questions. Each evaluator indicated which expected facts the student revealed. When the grader was done, the results were also uploaded to the SurveyMonkey data server.

For example, Figure 4 shows part of the grading form for one student. The first sentence is the original question. The second sentence is the student's answer. By checking one of the three radio buttons in the next line, the grader will give full, half, or no credit based on whether the student's answer shows understanding that the Egyptians want the world to be an orderly place. Other lines (not shown) allow the grader to award points for unexpected facts.

4. What does the Pylon tell us about how the Egyptians wanted the world to be like?

They wanted the world to be organized and orderly

	none	half	full
Orderly	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ruled by Egypt	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Pharaoh rules Egypt	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Pharaoh has the support of the gods	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Figure 4. Original question, student's answer, and grader's scoring options.

The RPM test, the presence and comfort test, the drawn map test, and the magnet map test were all printed and produced by the authors. See Jacobson (2008) for the actual tests and examples of their use.

Students wrote their answers for the **RPM Test** on paper, and we had an assistant enter the data (for each student) into another HTML-based form. Raven's Progressive Matrices (RPM) (Raven, 1958; Gregory, 1999) measures students' basic spatial and spatial-analogical reasoning ability without the use of language.

The **Presence and Comfort Test** was a one-page multiple-choice test on paper, which students filled out by circling their chosen answers. The first question was taken from Hunter Hoffman's questionnaire for Presence (Hoffman, 1999), a measure of presence. It asked students to choose on what is, in effect, a five-level Likert scale how much they felt as if they were *inside* the temple. The other questions are from Kennedy's questionnaire for motion sickness (Kennedy, 1992).

We used two map tests. In the **Drawn Map Test**, each student was given a blank sheet of paper and told to draw an overhead-view floor plan of the temple, from memory. In the **Magnet Map Test**, each student was given a correct overhead floor plan of the temple with a metal backing. The student placed magnetic icons onto the map, with each icon representing an important feature of the temple. For both map tests, the administrator photographed the map and stored it for later analysis. For each map, we constructed an HTML-based evaluation form, which each evaluator filled out. When an evaluator was done with the evaluation, data would upload to the data server for later analysis.

For each **Video Test**, the tester took the student to a room containing a desktop computer, a projector, a video camera, and the same controller the student used during the game (a Gyromouse). Onto the wall, the computer projected a non-interactive copy of the Virtual Egyptian Temple, with doors removed, from the game. The tester instructed the student to produce a guided tour of the Temple, saying in effect "Tell us everything you know about the Temple, and show us as you tell us." The student could point with his or her free hand, with the hand holding the Gyromouse, or with a laser pointer (Figure 5). When the student was ready, the tester would begin recording both the student and the computer screen with a digital video camera while the student made the presentation. Students usually took 15 minutes to produce a video, as we had discovered during pilot testing (Jacobson, 2008). Each student used the Gyromouse to navigate, indicating features of the Temple in any way he or she liked, which was usually by pointing.



Figure 5. A student gives a tour of the temple

We evaluated each student-made video by having an evaluator (grader) fill out a questionnaire, which listed the facts we expected the video to include (Figure 6). The grader could assign full or half credit for how the student described each fact, in a way similar to the grading of the short answers on the written posttest. All student videos were evaluated by three evaluators. The data we collected is a measure of *factual learning* using the video test.

 A screenshot of a web browser window displaying a survey on SurveyMonkey. The browser's address bar shows the URL 'https://www.surveymonkey.com'. The survey title is 'Renee Video' and the current section is '3. Courtyard'. The survey question asks the evaluator to rate the extent to which the student repeats or paraphrases facts from a video. The facts are listed in a table with radio buttons for 'none', 'half', and 'full' credit. The student's code name 'Renee' is visible in the browser's address bar.

	none	half	full
The courtyard holds several hundred.	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>
Everybody could come there from, Farmer to Pharaoh	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>
Open and undivided.	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>
made for large religious celebrations.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Figure 6. Part of an evaluator's scoring sheet. The courtyard is part of the temple, and "Renee" is the student's code name.

When a student completed the testing at the Earth Theater, we gave her a booklet that contains more information on Egyptian life and contextualizes the temple. It also contains links to relevant external resources.

One to two months later, the student completed the online (HTML-based) **Follow-Up Test** from home or elsewhere to receive a gift certificate and a copy of her video test.

Interrater Reliability for the Video Test

To provide a broad and stable measure of student performance in the video test, three evaluators graded the student-made videos using the Evaluator's Scoring Sheet (Figure 6). One was a high-school teacher, another a medieval historian, and the third a tour-guide familiar with the temple. Each evaluator produced a complete set of scores for all student videos, which were combined into a single set of measures, using interrater reliability analysis. For reasons described in Jacobson (2008, pp.164-165), we chose Fleiss Kappa (Fleiss, 1981; Siegel, 1988, p. 284), which is designed to handle ratings with few levels (i.e., true/false, high/medium/low, etc.) and to handle any number of raters. Our data were also sensitive to grader bias, but not overly so. We employed the code in the spreadsheet produced by King (2007). Standard measures of correlation (e.g., Pearson's-R or Spearman's-R) were not appropriate in this case.

The final result was a probability score (the approximate standard error from Fleiss, 1981) for each question, indicating whether the graders were substantially in agreement on the quality of all students' performance with regard to the fact being graded. If the standard error was less than 0.05, there was a 95% chance that all three graders were truly in agreement on all students' performance (taken together) with regard to that fact, despite a small amount of random fluctuation. Student membership in an experimental treatment group was not a factor. We were concerned only with whether the graders were in agreement for each student.

In the Video Test, we found that student responses for 45 out of the 75 facts (60%) had acceptable levels of rater agreement ($P < 0.5$). However, those 45 accounted for 89% of all points awarded, because few students mentioned the remaining 30 facts, leaving too little information to determine rater agreement. Within the accepted data, the raters' three data sets can be safely averaged to produce an aggregate set of scores. All further discussion of video data analysis refers to the aggregate measures of students' descriptions of those 45 facts.

RESULTS & ANALYSIS

Table 1. Summary of results

	Test	Significant Result?
H1	Written	Yes
H2	Written	No
	Video	Yes
H3	Follow-Up	No
H4	Drawn Map	No
H5	Magnet Map	No
H6	RPM x Written	No
	RPM x Video	Yes

Written Posttest: *Gates of Horus* Probably Effective

Before we can meaningfully compare learning results for *Gates of Horus* based on display type, we must first have some assurance that the game actually functions as a learning tool. Twenty students in the no-treatment Control Group took the written test without playing the game, while twenty students in the Desktop Group played the game and took the written test. Results and analysis are only summarized, here, because they are detailed in Jacobson (2009).

For the first section of the test (the automatically stored multiple-choice questions), a two-tailed, uneven variances t-test on students' total scores yielded $P < 0.0016$, a very strong difference.

The second section of the test asked students to write responses to short answer questions, which were graded by four evaluators, who awarded full or partial credit when students stated expected facts. We combined the evaluators' grades through interrater reliability analysis similar to that used for the video tests. A two-tailed, uneven variances t-test on students' total scores yielded a strong difference of $P < 0.0044$. (One of the control group students declined to fill out the short-answer section, so we excluded his test from that comparison because evaluating his score as a zero would have been unfair to the control condition.)

Lastly, the evaluators gave subjective ratings on the students' overall performance on the test, with the difference being so strong that P was essentially zero.

Written Posttest: No Significant Difference by Display Type

Comparing results for the written test between the Desktop Group (20 students) and the Theater Group (25 students) yielded no reliable significant differences. Comparing students' total scores

in the automatically graded multiple-choice section did yield what looked like a significant difference ($P = 0.0290$) using the Mann-Whitney test. See Jacobson (2008, p. 173, p. 165) for the reasons why we chose a non-parametric test and other details of our analysis strategy. However, the Theater Group's performance showed a statistically significant advantage (Mann-Whitney, yielding $P < 0.05$) in only two of the 20 questions. This is not an adequate basis to conclude that the significant difference between the totals is meaningful.

Comparing results for short-answer questions, also using Mann-Whitney, yielded a result $P = 0.8388$. Finally, comparing evaluators' subjective ratings showed only $P = 0.6865$, and t-tests for both comparisons yielded similar results. Therefore, results of the written posttest failed to support **H2**.

Clearly, the written test (Jacobson, 2008, p. 288) was not sufficiently sensitive to detect the differences revealed in the video test. Designing an effective written test is certainly possible, but the information that is, by necessity, revealed in the questions themselves is an obstacle. This is a complex and important issue that we hope to address in a later study.

Maps, Presence, Game Logs, and Follow-Up Tests, All Negative

We had hoped that students who learned with the immersive condition would remember more facts and concepts than the others. Unfortunately, we saw no statistically significant differences, and there were not enough differences in individual scores to be convincing. Results of the follow-up test were very similar to those of the posttest, except that students gave much less informative answers on the short-answer questions. This leaves **H3** unsupported.

In data from both the drawn map test for **H4** and the magnet map test for **H5**, we saw no significant differences in performance for the Theater Group versus the other groups, and we found no trend in one direction or the other. This finding is somewhat at odds with the virtual reality training literature, which has established that Immersive VR is a good way to teach survey knowledge of a geographic area (Darken, 2002; Durlach, 2000). Perhaps the temple was too small and too simple for immersion to produce a genuine difference in students' knowledge that could be detected with a mapping test.

On the presence and comfort test, the one question that queried students' degree of presence failed to show any significant difference between the Desktop and Theater Groups. We see now that the presence question (Hoffman, 1998) was probably not an adequate measure (Slator, 2004). Fortunately, the test also failed to reveal any significant motion sickness, regardless of whether students used the immersive display or a standard desktop.

A cursory examination of the game logs showed no correlation between completion times for the game and learning outcomes.

Video Test: A Positive Result for Learning

Students' total scores for all factual learning in the Theater Group were compared to the totals for students in the other two groups (Desktop and Control) taken together (Figure 7). This comparison has the statistical advantage of comparing larger groups and is conservative with respect to **H2**: The aggregate non-Theater Group had an unfair advantage in the Video Test, because students in the Control Group had more exposure to the material by taking the posttest first (Figure 3).

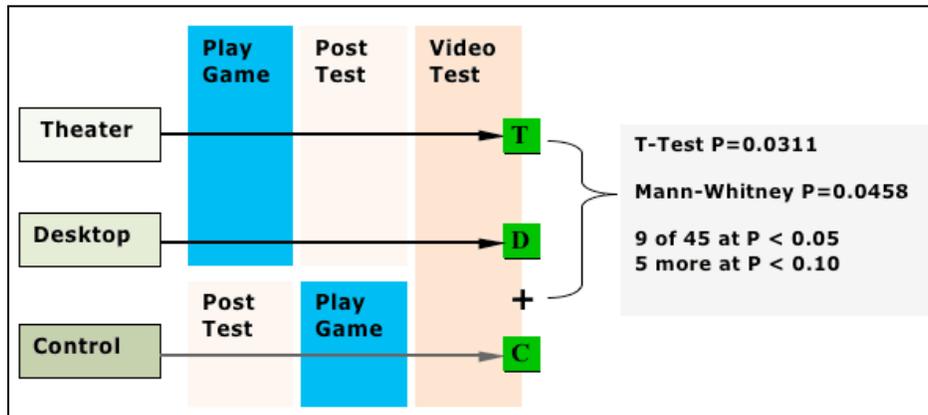


Figure 7. Video Test results.

Scheduling considerations left us with fewer completed videos than completed written posttests. Video data were collected for 18 students in the Desktop Group and 17 in the Control Group, for a total of 39, and video data were collected for 22 students in the Theater Group. A two-tailed, two-sample, unequal variance t-test comparison of total facts mentioned by students in the two groups (the Theater Group versus all others) showed a significant difference with $P = 0.0311$. For comparing the totals, the t-test is probably adequate, but the underlying data are non-continuous and highly skewed. A non-parametric alternative, the Mann-Whitney test, yielded $P = 0.0458$. See Jacobson (2008, p. 173, p. 165) for further discussion of the analysis strategy.

The effect was strongest in the inner Sanctuary of the temple, the last area where all the themes of the game come together: The t-test yielded $P = 0.0230$ and Mann-Whitney produced $P = 0.0359$. Perhaps by the time the students reach the sanctuary, they are more accustomed to the game and its controls than they were when they entered other areas earlier. The reduction of the cognitive load of playing the game may cause less noise (errors) in the data, allowing the treatment effect to surface more clearly. Alternatively, the treatment effect may be more apparent because the sanctuary is the most complex area spatially and informationally. However, this is only conjecture, and this unexpected result was not part of our a-priori experimental hypotheses.

According to the Mann-Whitney test, students in the Theater Group performed better than the other students in nine facts at the $P < 0.05$ level of certainty and in five more at the $P < 0.10$ level for a total of 14 out of 45. (Note that the main result of the study does not depend entirely upon the difference shown in just these 14, but in aggregate rating across all the facts.) Also, students in the Theater Group did better at the $P < 0.05$ level for four of the 13 facts associated with the Sanctuary and at the $P < 0.10$ level for three more, a total of seven out of 13. There were no cases where the Theater Group was statistically worse than the other groups. See Figure 7.

It is tempting to look at the data more closely and try to determine the shared characteristics of facts and ideas which seemed sensitive to immersion. Unfortunately, our test population (65) was too small and the treatment too short to support analysis on that level of detail. There would be a significant danger of “over-fitting the data,” ascribing significance to largely random events.

Video Test: A Glimpse of Conceptual Learning

Our Egyptologist (Holden) evaluated the videos for conceptual learning to complement the other evaluators' relatively rote scoring of factual and lower-level conceptual learning. Students using the immersive display generally did better at explaining the most important concepts ($P < 0.001$). The treatment group also tended to do better with the specific facts most closely linked to those key concepts. Furthermore, it appears that students with low RPM scores benefited more from the immersive display than those with high RPM scores ($P = 0.01$). The result parallels a study by Winn (1997) in which students who made poor grades benefited more from using an immersive display in Winn's educational VR application (Winn, 2001). This interaction effect was not evident in scoring for factual data.

This result is fascinating, but uncertain, because the scoring is based on the judgment of a single evaluator. In 2010, Robyn Gillam, a published Egyptologist, evaluated the student-made videos in the same way. At the time we wrote this article, she had completed her evaluation, but the data has not yet been analyzed. That analysis will be the topic of a follow-on article.

DISCUSSION

In our study, all students played *Gates of Horus* (Jacobson, 2009), a solitary-player educational game based on the Virtual Egyptian Temple (Troche, 2010). Afterward, each student produced a videotaped tour of the temple, describing its features and meaning in the student's own words. Three independent evaluators rated the completeness and accuracy of each video, yielding a composite measure of student performance on our Video Test. Some students played the game in a visually immersive display, the Earth Theater (CMNH, 2010), an all-digital partial-dome display (Figure 1); and others played it on a standard desktop computer. Students who used the visually immersive display recited more facts in their Video Test than those using a desktop ($P = 0.0458$).

This positive result for the immersive display is in line with two other studies. In Salzman (1998), students with access to both an egocentric (inside) view of a magnetic field and an exocentric view learned more than those limited to just one view. Similarly, in Winn (2001), students with access to an egocentric view of an information-rich visualization of the Puget Sound area learned more than those who had only an exocentric view. Our study is further evidence that the egocentric view provided by immersive displays is *sometimes* advantageous to the learner in educational virtual reality games. If this phenomenon is real, understanding how and why it works will require further study.

One or more of the following explanations may account for why students in the immersive display recited more facts on the video test:

- (1) They understood the material more deeply, learning more about the mythologies, philosophies, and culture expressed by and in the temple. This is possible, but given how short the game is (40 to 60 minutes), it seems less likely than other explanations.
- (2) They developed a better mental map or image of the temple, providing a visual map of the information about the temple. The game explicitly teaches facts and concepts about the temple and its features, so a better mental image of the temple itself should aid recall. This is likely, because the temple is designed to be a visual index of the information it represents.

- (3) The immersive display granted the student a physical advantage for object/feature location and selection. Because the view is wider than that provided by a desktop monitor, the student can visually search the environment more efficiently, having to rotate the view less often to find salient features. This is likely, because less time the student spent trying to search and navigate, the more time he or she could spend thinking and learning.
- (4) The immersive view inspired a higher degree of engagement in the students using it. This is very similar to the old argument that sensory immersion enhances presence and that presence enhances learning. While this is possible, presence is complex and difficult to test for (Bailenson, 2008; Slater, 2004), and our own test for presence in this study showed no significant difference between the groups. Anecdotally, all students in the study seemed highly engaged, regardless of the display type.

The *Written Test* showed that students who played the game on a desktop showed more knowledge of the temple than a no-treatment control group (Jacobson, 2009) at $P=0.0044$. Also, there appeared to be an interaction effect between students' performance on the Video Test and their visual reasoning ability as measured by Raven's Progressive Matrices (Raven, 1958). We will investigate that further in a follow-on study. The other posttest measures showed no results.

Again, the goal of our study was to add data to the larger question of when and where immersive displays can and should be used for educational virtual reality, especially games. We believe we have done that, and hope it leads to further research.

ACKNOWLEDGEMENTS

Our study was part of a larger collaboration among PublicVR (PublicVR, 2010), the Carnegie Museum of Natural History in Pittsburgh (CMNH, 2010), and the School of Information Sciences at the University of Pittsburgh (PITT, 2010). It was also this author's dissertation, where the reader can see all of the original tests we used (Jacobson, 2008).

This study and *Gates of Horus* depend on the Virtual Egyptian Temple, CaveUT, and VRGL. Each of these is a separate project with a long list of credits – too long to list here. To see them, go to <http://publicvr.org> and look under "Projects." For *Gates of Horus* and this study, specifically, we first thank the following institutions:

- PublicVR for providing the funds and open-source software to develop the game, testing software, and other testing materials (<http://publicvr.org/>).
- The Carnegie Museum of Natural History for providing the venue, the test population, payment for Kerry Handron's time, and miscellaneous resources for the game and the study (<http://www.carnegiemnh.org/>).
- The University of Pittsburgh's Department of Information Sciences for making their Virtual Theater available for software testing and general support for Jeffrey Jacobson's work on this project (<http://www.ischool.pitt.edu/>).
- Carnegie Mellon University's Studio for Creative Inquiry (<http://www.cmu.edu/studio/>) for their role in developing the Virtual Egyptian Temple.

- Epic Games for their general support for academic research using the game Unreal Tournament 2004 and its software core, the Unreal Engine (<http://www.epicgames.com/>).
- Psyonix, Inc. (contractors) for the software programming (<http://www.psyonix.com/>).
- Wrecking Crew, Inc. (contractors) for the voice recordings

Next, we would like to thank the following individuals:

- Michael Lewis (University of Pittsburgh), who generally supported the work through his role as Jeffrey Jacobson's dissertation advisor (<http://www.ischool.pitt.edu/people/lewis.php>).
- Lowry Burgess for his insights into the dramatic structure of the game, and learning and perception in virtual reality (<http://artscool.cfa.cmu.edu/~burgess/>).
- William Winn, Kurt Van Lehn, Peter Brusilovsky, and Stephen Hirtle for their advice on the educational research design, cognition, and adaptivity (Jacobson, 2008).
- Richard Graefe for his expert editing.
- Mack C. Shelley for help in selecting the appropriate statistical methods used to analyze the results (<http://www.stat.iastate.edu/directory/personal.php?id= mshelley>).
- Jane Vadnal for her help in the production of materials for the experiments.
- Kathy Bruckner, Adrienne Baker, Benjamin Getkin, Jane Vadnal, and Mary Smith for their help in evaluating the students' performance in the posttests, as well as in test subject recruitment and preparation of test materials.

REFERENCES

- Anderson, E. F., McLoughlin, L., Liarokapis, F., Peters, C., Petridis, P., and Freitas, S. (2010). Developing serious games for cultural heritage: A state-of-the-art review, *Virtual Reality*, 14, 255-275
- Ang, C. S. and Krishna, R. G. S. V. (2008). Computer game theories for designing motivating educational software: A survey study, *International Journal on E-Learning* (2008)
- Bailenson, J. N., Yee, N., Blascovich, J., Beall, A.C., Lundblad, N., and Jin, M. (2008). The use of immersive virtual reality in the learning sciences: Digital transformations of teachers, students, and social context. *The Journal of the Learning Sciences*, 17, 102-141 ISSN: 1050-8406 print / 1532-7809 online.
- Baker, C. (2008). Trying to design a truly entertaining game can defeat even a certified genius. *Wired*, 16(4). http://www.wired.com/gaming/gamingreviews/magazine/16-04/pl_games.
- Barab, S. A., Scott, B., Sinyahhan, S., Goldstone, R., Ingram-Goble, A., Zuiker, S. J., and Warren, S. (2009). Transformational play as a curricular scaffold: Using videogames to support science education. *Journal of Science Education Technology* 18, 305-320
- Barsalou, L. W. (1992). *Cognitive Psychology*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc. 0-8058-0691-1 (hard) 0-8058-0691-0 (pbk).

- Bowman, D. A. (2002). Principles for the design of performance-oriented interaction techniques. In K. M. Stanney (Ed.), *Handbook of Virtual Environments* (p. 277). Mahwah, New Jersey: Lawrence Erlbaum Associates, Inc., Publishers. 0-8058-3270-X
- Bowman, D. A., Hodges L. F., Allison, D., and Wineman, J. (1999). The educational value of an information-rich virtual environment. *Presence Teleoperators and Virtual Environments*, 8(3), 317-331.
- Byrne, C. (1996). *Water on tap: The use of virtual reality as an educational tool*. Unpublished dissertation, Department of Education, University of Washington, Seattle, USA.
- Campbell, C. T. and Stanley, J. C. (1963). *Experimental and quasi-experimental designs for research*: Houghton Mifflin Company. 0-395-30787-2
- Champion, Erik. (2008). Otherness of place: Game-based interaction and learning in virtual heritage projects. *International Journal of Heritage Studies*, 14(3), 210-228.
- Chertoff, D. B. and Schatz, S. L. (2008). Improving presence theory through experiential Design. *Presence* 17(4), 405-413
- Clark, R. E., Yates, K., Early, S. and Moulton, K. (2010). An analysis of the failure of electronic media and discovery-based learning: Evidence for the performance benefits of guided training methods. In K. H. Silber, & R. Foshay, (Eds.). *Handbook of training and improving workplace performance, Volume I: Instructional design and training delivery*. Washington, DC: International Society for Performance Improvement.
- Clark, R. E. (1994) Media will never influence learning, *Educational Technology Research and Development*, 42(2), 21-29
- Clark, R.E. and Surgrue, B. M. (1988). Research on instructional media, 1978-1988. In *Educational Media and Technology Yearbook*, 14, 19-36. Englewood, Colorado: Libraries Unlimited, Inc. D.P. Ely, B. Broadbent, and R.K. Wood (Eds.).
- CMNH (2010). Carnegie Museum of Natural History, 4400 Forbes Ave., Pittsburgh, PA 15213. <http://www.carnegiemnh.org/>.
- Cruz-Neira, C., Sandin, D., and DeFanti, T. A. (1993). Surround-screen projection-based virtual reality: The design and implementation of the CAVE. SIGGRAPH '93.
- Darken, R.P. and Peterson, B. (2002). Spatial orientation, wayfinding, and representation. *Handbook of Virtual Environment Technology*. Stanney, K. Ed., Lawrence Erlbaum Assoc.: New Jersey. pp. 493-518.
- Dede, C., Clarke, J., Ketelhut, D. J., Nelson, B., and Bowman, C. (2005). *Students' motivation and learning of science in a multi-user virtual environment*, Paper presented at the annual meeting of the American Educational Research Association.

- Dede, C., Barab, S. (2009). Emerging Technologies for Learning Science: A Time of Rapid Advances, *Journal of Science Education Technology*, 18, 301-304
- Dede, C., Salzman, M. C., Loftin, R. B., and Sprague, D. (1999). Multisensory Immersion as a modeling environment for learning complex scientific concepts. In *Computer Modeling and Simulation in Science Education*. Springer-Verlag.
- Dickey, M. D., Game design narrative for learning: Appropriating adventure game design narrative devices for the design of interactive learning environments, *Educational Technology Research and Design*, 54(3), 245-263.
- Dondlinger, M. J. (2007). Educational video game design: A review of the literature, *Journal of Applied Educational Technology*, 4(1), 21-31.
- Durlach, N., Darken, R. P., Allen, G., Loomis, J., Templeman, J. (2000). Virtual environments and the enhancement of spatial behavior: Toward a comprehensive research agenda. *Presence: Teleoperators and Virtual Environments*, 9(6), 593-615.
- Ellis, S. R. E. (1991). *Pictorial Communication in Virtual and Real Environments (2nd ed.)* London: Taylor and Francis.
- EpicGames (2010). <http://www.epicgames.com/>
- Fisch, S. M. (2005). Making educational computer games “educational.” Paper presented at the 2005 Conference on Interaction Design and Children, Boulder, CO.
- Fleiss, J. L., Levin, B., and Cho, M. (1981). *Statistical Methods for Rates and Proportions*. ISBN: 0-471-52629-0.
- Gregory, R. J. (1999). *Foundations of Intellectual Assessment*. Needham Heights, MA, Allyn and Bacon.
- Hoffman, H. G., Prothero, J., Wells, M., and Groen, J. (1998). Virtual chess: Meaning enhances users' sense of presence in virtual environments. *International Journal of Human-Computer Interaction*, 10(3), 251-263.
- Jackson, R. L. and Winn, W. (1999, Dec. 12-15). Collaboration and learning in immersive virtual environments. *Proceedings of the Computer Support for Collaborative Learning Conference (CSCL) 1999*, Stanford University, Palo Alto, California.
- Jacobson, J. and Chadwick, W.A. (2010). Reconceptualizing “virtual reality”: What is VR? *IEEE Virtual Reality 2010*, March, MA, USA (Panel, Moderators).
- Jacobson, J. (2008). Ancient architecture in virtual reality; Does visual immersion really aid learning? Unpublished doctoral dissertation, School of Information Sciences, University of Pittsburgh. <http://planetjeff.net/Jacobson2008.pdf>.

- Jacobson, J., Handron, K., and Holden, L. (2009). Narrative and content combine in a learning game for virtual heritage. *Computer Applications in Archaeology*, Williamsburg, VA.
- Jacobson, J. and Holden, L. (2005). The virtual Egyptian temple. Paper presented at the World Conference on Educational Media, Hypermedia & Telecommunications (ED-MEDIA), Montreal, Canada.
- Jonassen, D. H. and Land, S. M. (2000a). *Theoretical Foundations of Learning Environments*. Mahwah, New Jersey: Lawrence Erlbaum Associates. 0-8058-3215-7 hardback, 0-8058-3216-5 paperback.
- Jones, G. and Warren, S. (2008). The time factor: Leveraging intelligent agents and directed narratives in online learning environments. *Innovate Journal of Online Education*, 5(2). <http://www.innovateonline.info/>.
- Kennedy, R. S., Fowlkes, J. E., Berbaum, K. S., and Lilienthal, M.G. (1992). Use of a motion sickness history questionnaire for prediction of simulator sickness. *Aviation, Space, & Environmental Medicine*, 63(7), 588-593.
- King, J. E. (2007). <http://www.ccitonline.org/jking/homepage/interrater.html>
- Kemp, B.J. (1989). *Ancient Egypt: Anatomy of a Civilization*. London: Routledge.
- Kolasinski, E. M. (1995). Simulator sickness in virtual environments; Technical Report 1027: Simulator Systems Research Unit, Training Systems Research Division, U.S. Army Research Institute for the Behavioral and Social Sciences.
- Kozma, R. B. (1994). Will media influence learning? Reframing the debate, *Educational Technology Research and Development*, 42(2), 7-19.
- Lantz, E. (2007). A survey of large-scale immersive displays: *Proceedings of the 2007 workshop on emerging displays technologies*, 2007, C. Cruz-Neira and D. Reiners (Eds.) pp. 1-7. New York, NY, USA, ACM, San Diego, California, ISBN 978-1-59593-669-1.
- Lewis, M. and Jacobson, J. (2002). Game engines in scientific research. *Communications of the ACM*, 45, 27-31.
- Limniou, M., Roberts, D., and Papadopoulos, N. (2007) Fully immersive virtual environment CAVE™ in chemistry education. *Computers and Education*, 51 (2008), 584-593
- Mayer, R. E. (2001). *Multimedia Learning*. Cambridge, MA: Cambridge University Press. 0-521-7839-2, 0-521-7839-1
- Mihalca, L., Miclea, M. (2007) Current trends in educational technology research, *Cognition, Brain, and Behavior*, 11(1), 115-129.
- McDowell, K. G. (1999). *Village Life in Ancient Egypt*, London/New York: Oxford University Press.

- Moreno, R. (2002b). *Pedagogical agents in virtual reality environments: Do multimedia principles still apply?* Paper presented at the World Conference on Educational Multimedia, Hypermedia and Telecommunications (ED-MEDIA), Charlottesville, VA.
- Osberg, K. M. (1997a). *Constructivism in practice: The case for meaning-making in the virtual world*. Unpublished doctoral dissertation, University of Washington, Seattle, Washington.
- Oxford English Dictionary (2010). Retrieved December 17, 2010, from Oxford Dictionaries website: <http://oxforddictionaries.com> .
- PITT (2010). University of Pittsburgh, School of Information Sciences, 135 North Bellefield Avenue, Pittsburgh, PA 15260. <http://www.ischool.pitt.edu>.
- PublicVR (2010). <http://publicvr.org>.
- Raven, J. (1958). *Standard Progressive Matrices*. San Antonio, TX: The Psychological Corporation.
- Rose, H. (1996). *Design and construction of a virtual environment for Japanese language instruction*. Unpublished Master's thesis, University of Washington.
- Roussos, M., Johnson, A., Moher, T., Leigh, J., Vasilakis, C., and Barnes, C. (1999). Learning and building together in an immersive virtual world. *Presence*, 8(3), 247-263.
- Roussou, M., Oliver, M., and Slater, M. (2007). Exploring activity theory as a tool for evaluating interactivity and learning in virtual environments for children. *Journal of Cognition, Technology & Work, Springer*, ISSN: 1435-5558 print / 1435-5566 online.
- Russell, T. L. (2000) *The No Significant Difference Phenomenon*. Publisher: IDEC. ISBN 0-9668936-0-3.
- Salzman, M. C., Dede, C., Loftin, R. B., and Ash, K. (1998). Using VR's frames of reference in mastering abstract information. *Third International Conference on Learning Sciences*, Charlottesville, VA.
- Salzman, M. C., Dede, C., Loftin, R. B., and Chen, J. (1999). A model for understanding how virtual reality aids complex conceptual learning. *Presence: Teleoperators and Virtual Environments*, 8(3), 293-316.
- Sauneron, S. (2000). *Priests of Ancient Egypt*. Ithaca: Cornell University Press.
- Slater, M., Lotto, B., Arnold, M. M., Sanchez-Vives, M. V. (2009). How we experience immersive virtual environments: the concept of presence and its measurement, *Anuario de Psicología*. 40(2), 193-210.
- Siegel, S. (1988). *Nonparametric Statistics*, McGraw-Hill, Inc. ISBN 0-07-057357-3.

- Slator, M. (2004). How colorful was your day? Why questionnaires cannot assess presence in virtual environments. *Presence*, 13(4), 484-493.
- Spring, M. (1992). Virtual reality and abstract data: virtualizing information. *Virtual Reality World*, 1(1).
- Squire, K. (2008). Video games and education: Designing learning systems for an interactive age. *Educational Technology*, 47(2), 17-25.
- SurveyMonkey® (2010). <http://www.surveymonkey.com>
- Troche, J., and Jacobson, J. (2010). An exemplar of Ptolemaic Egyptian temples. *Computer Applications in Archaeology (CAA)*, Granada, Spain, April, 2010
- Vygotsky, L. S. (1978). *Mind in Society: The Development of Higher Psychological Processes*. Cambridge, MA: Cambridge University Press.
- Watterson, B. (1998) The House of Horus at Edfu: Ritual in an Ancient Egyptian Temple. Stroud, UK: *Tempus1998*. pp. 35-43.
- Winn, W. (1997). The impact of three-dimensional immersive virtual environments on modern pedagogy (Technical Report R-97-15). Seattle: Human Interface Technology Lab, <http://www.hitl.washington.edu/publications/r-97-15/>.
- Winn, W. (2003). Learning in artificial environments: Embodiment, embeddedness and dynamic adaptation. *Tech., Inst., Cognition and Learning*, 1, 87-114.
- Winn, W., Windschitl, M., Hedly, N., and Postner, L. (2001). Learning Science in an Immersive Virtual Environment. Paper presented at the *Annual Meeting of the American Educational Research Association*, Seattle, WA.
- WreckingCrew (2010) <http://www.wreckingcrewmedia.com/>.
- Zeltzer, D. (1992). Autonomy, Interaction, and Presence. *Presence: Teleoperators & Virtual Environments*, 1(1), 127-232.