Review of Educational Research Methods in Desktop Virtual World Environments: Framing the Past to Provide Future Direction

Dennis Beck
University of Arkansas

Ross A. Perkins
Boise State University
Abstract

This article describes educational research methodologies being used in the investigation and evaluation of desktop virtual world environments (DVWEs) as used to teach a variety of subjects in higher education. Ellis’ research framework on innovations (2005) and Reeves and Hedberg’s research goals (2003) were used as lenses to examine the state of educational research methods within DVWEs. An extensive search that resulted in 127 peer-reviewed papers chosen, was carried out to locate journal articles in the fields of educational technology, computer science and information systems, virtual reality/virtual worlds/gaming, science and health education, human-computer interaction, media and communications, psychology, social science, cognitive science, and library science. Results showed that while small scale research is common, research with broader goals (i.e. program evaluation, developmental research) is lacking. Implications include the need for signature developmental researchers to make their virtual world applications open source, and for DVWE researchers to explore collaborative research opportunities with postmodern, education researchers in order to shed more light on important issues of equity, gender, politics, and culture that impact education.
1. Introduction

The purpose of this article is to describe educational research methodologies being used in the investigation and evaluation of desktop virtual world environments (DVWEs). Specifically, it examines how DVWEs are implemented in educational contexts across multiple subject areas in higher education. We will utilize Ellis’ research framework on innovations (2005) and Reeves and Hedberg’s research goals (2003) as lenses to examine the state of educational research within DVWEs. Results will provide a “map” of research methodologies being used according to the first, second, and third levels of research (Ellis, 2005) and according to the six research goals (Reeves and Hedberg, 2003), which will be further described in this paper. Results will also show methodological gaps that exist in educational research about DVWEs. Equipped with such knowledge, researchers can make more intelligent decisions about types of studies and methods of data collection and analysis, resulting in more rigorous studies that cover a broad range of research levels and goals.

2. Theoretical Framework

2.1 Framework on Educational Innovations

Ellis (2005) provides a helpful framework from which to view research in new fields of academia.

- Level I research is theory building, pure research, or both. It is most commonly conducted in experimental or laboratory settings. Its purpose is to establish a theoretical construct or idea as having some validity. It includes both qualitative and quantitative studies, and may encompass such measures as correlations, descriptive data, and case studies. An important limitation of Level I research is that it is not designed to answer applied questions directly.

- Level II research is empirical research, either quantitative, qualitative, or a combination thereof. Its purpose is to determine the efficacy of particular programs or methods in authentic settings. As applied research, it seeks to apply theories and procedures developed in Level I research and provide practical insights that cannot be derived from Level I research. An important limitation to note about Level II research is that it is severely limited in generalizability, but that can be improved by replication studies.

- Level III research determines the extent to which a program (curriculum innovation) is successful when its implementation becomes widespread. An example from the field of education would be evaluation studies that determine the overall effects on teachers and students of a particular district or school-wide innovation. Most importantly, this level of research answers the question: Do these results work on a larger scale?

2.2 Common Research Goals

Reeves and Hedberg (2003) classify research into six categories based on the goals of the investigators: theoretical, predictive, interpretivist, postmodern, action, and developmental. The focus of research with the theoretical goals is to explain “phenomena through the logical analysis and synthesis of theories, principles, and the results of other forms of research such as empirical studies” (p. 270). This type of research is relatively rare given both the requisite experience of the researchers and the longer amount of time it takes to frame this unique type of inquiry.
Researchers who have as their focus predictive goals conduct experimental or quasi-experimental studies in order to test and expand existing theories. In the field of educational technology, such studies are often carried out with a particular innovation in mind, and done so using conditions that are as controlled as possible. Studies driven by predictive goals are prevalent in the field, though their quality is often suspect (Reeves & Hedberg, 2003), and their ultimate usefulness to practice, despite claims of generalizability, tends to be questionable.

In contrast to those who approach research with a desire to predict outcomes are those who have interpretivist goals. The focus is to “[portray] how education works by describing and interpreting phenomena related to teaching, learning, performance, assessment, social interaction, innovation...” (p. 270). The methodological approach tends to be qualitative. The major criticism of this kind of inquiry is that the results are seen as “data islands,” in that they cannot be generalized outside of the context in which such studies were conducted.

Critical analysis of phenomena, agendas, practices, and so on is not common in educational technology (Reeves & Hedberg, 2003), but one can occasionally find examples of this type of inquiry. Among scholars with postmodern goals, there is a keen desire to examine equity, gender, political and cultural issues as they impact the field.

A fifth research goal is the “action goal,” where researchers look at the application of a program, product, or method with the aim to describe, improve, or evaluate its impact or effectiveness. In “action” or “evaluation” research, the “major goal is solving a particular problem in a specific place within a relatively short timeframe” (Reeves & Hedberg, 2003, p. 272). The audience tends to be either practitioners or stakeholders.

Developmental research is advocated by the authors because it uses the empirical method to examine authentic, complex problems in a realistic context. Additionally, it seeks to provide pragmatic solutions to those who implement educational technology innovations - not just in one context, but multiple contexts. Practitioners are important to collaboration because it is “focused on the dual objectives of developing creative approaches to solving human teaching, learning, and performance problems while at the same time constructing a body of design principles that can guide future development efforts” (Reeves & Hedberg, 2003, p. 271). This type of research blends “best practices” design guidance and hypothetical approaches in a manner that seeks to create solutions to the problems presented. Also known as “formative research,” a great deal of emphasis is placed on reflection to test, re-test, and “refine innovative learning environments as well as to define new design principles” (p. 272).

Ellis’ (2005) framework and Reeves and Hedberg’s (2003) research goals will be utilized in this literature review in order to show how studies might be categorized, thereby showing any gaps in types of research and methodologies.

2.3 Related works

DVWEs differ from other virtual reality technologies in that they display the visual information of a virtual world on a computer screen and provide little in the way of proprioceptive cues (Blascovich, Loomis, Beall, Swinth, Hoyth, & Bailenson, 2002; Jansen-Osmann, 2007). Literature reviews on DVWEs are currently focused in the domain of education. One subset is descriptive (Baker, Wentz, & Woods, 2009; Harris & Rea, 2009; Oliverio & Beck, 2009; Yee, 2006). Many authors, such as Baker et al. (2009), have written descriptive pieces about virtual worlds such as Second Life, Whyville, Active Worlds, and There. These articles provide basic information about virtual worlds including usage...
statistics, how they are being used in a particular academic field, advantages and disadvantages of using them in teaching, and common-sense tips on how to get the most out of a virtual world experience. The demographic information provided about virtual world environments in this type of literature review can be extremely helpful in ensuring investigators get the right mix of participants in their research study.

Another subset of DVWE literature reviews search for scaffolding that may be supplied for teaching and learning (de Byl, 2009; Kemp & Livingstone, 2006; Livingstone, Kemp, & Edgar, 2008; Salt, 2008). For example, Livingstone et al. (2008) examine virtual world environments in order to determine the level of support they may provide for learning activities. Results show that virtual world environments have not been designed or built for teaching and learning, and many of the elements of support for teaching and learning that are found in web-based learning environments are completely absent. In order to support educational activities in virtual worlds, the authors suggest the integration of a Learning Management System that is integrated with the virtual world. Salt et al. (2008) would also be considered in this category, as they draw together a number of different threads in their literature review of education in virtual world environments. The authors discuss the characteristics of the current generation of virtual worlds, a snapshot of educational activities, a discussion of the major learning theories that have been studied, consideration of the design of learning activities and an exploration of what techniques are being used to control the development of the building of relevant online resources. One common thread that emerges in this review is the focus on the potential for the creation of useful and meaningful learning activities within DVWEs. It also discovers an absence of objective measures of the efficacy and quality of learning in virtual world environments, and a relative paucity in terms of the number of peer-reviewed articles specific to virtual worlds.

Also part of the subset concerning scaffolding for teaching and learning would be studies such as Nelson and Ketelhut’s (2007) review examining the problems of implementing authentic scientific inquiry curricula in schools utilizing educational virtual world environments. Their review reveals that virtual world-based curricula supporting real-world inquiry is viable provided it is based on authentic interactivity with simulated worlds and tools. It also shows that this curricula works well with individuals of either sex. Examining the literature from the perspective of finding an answer to a particular problem is important because it applies theory to real-life problems.

In this category as well, Richter and Livingstone (2011) review the literature on multi-user games and learning, with a focus on the motivations that draw people into playing and learning with multi user games, learning outcomes related research using multi-user games, and thoughts on conducting educational research in light of emerging multi-user games--such as augmented reality or "mLearning"-that blur previously defined boundaries and operational contexts for integrating technology, research, and education. Additionally, Tobias and Fletcher (2011) compare computer games with other instructional methods, as well as review of research on a number of game design issues related to education, and recommend research needed to enhance the usefulness of games for instruction.

Literature reviews on DVWEs have been categorized as descriptive, scaffolding, discipline-specific, and methodology-based. Specifically, the latter category of reviews provides much helpful information on the types and quality of educational research being performed. However, these reviews lack the perspective and guidance provided by a theoretical framework. A theoretical framework is a series of informal or formal interrelated concepts, which provides a systematic view of a phenomenon (Creswell, 2013). The reviews by Inman et al. (2010) and Hew and Chung (2010) differ in that although they provide much helpful information, there is no theoretical framework that helps to interpret results and guide future research. To help fill this void, this article reviews the literature in order to describe research methodologies being utilized in the investigation and evaluation of DVWEs. Moreover, Ellis’
(2005) framework and Reeves and Hedberg’s (2003) research goals will be utilized in this literature review as a theoretical framework in order to help interpret results and guide future research.

A third subset of literature reviews on DVWEs analyzes literature for evidence of educational and discipline-specific opportunities (Mason & Moutahir, 2006; Sidorko, 2008; Whitton & Hollins, 2008). For example, Sidorko (2008) attempts to discover the educational and library information opportunities afforded through DVWEs. Results indicate that DVWEs have the potential to provide a fertile learning and information context. Virtual worlds can enhance the learning experience if problematic issues are addressed and if expectations are realistic. Uncovering opportunities for future research and application of educational innovations help to provide guidance to other researchers on the gaps in a particular field of research or educational innovations.

A fourth subset of literature reviews on DVWEs review past research for information on the types and quality of studies. For example, Hew and Cheung (2010) reviewed past empirical research studies on the use of three-dimensional immersive virtual worlds in education settings. They attempted to determine how virtual worlds are being used by students and teachers, the types of research methods applied, and the variety of topics studied. Hew and Cheung (2010) discovered that virtual worlds are being utilized for communication, simulations, and experiential opportunities. They discovered that most of the studies reviewed were descriptive in nature and conducted in higher education settings. Main research foci were participants’ affective domain, learning outcomes, and social interaction.

Inman et al. (2010) also reviewed the use of Second Life in K-12 and Higher Education. The methodology, findings, and recommendations of 27 research studies were analyzed. Results indicated that qualitative, quantitative, and mixed method research studies were all equally prevalent. It also found that sample sizes of the studies conducted thus far are relatively small, with approximately 74% of the identified studies having a sample size between four and 43 participants. Next, it discovered that educational research in Second Life is primarily focused on higher education utilizing traditional survey instruments for data collection. Finally, it showed that most teachers used Second Life as part of a course, engaging students with experiential learning activities such as role-play, simulations, project-based learning, and group learning. The study concludes with several recommendations of best practices for educators using Second Life.

3. Method

In order to accomplish this purpose, an extensive search was carried out to locate journal articles in multiple academic fields. Keywords used to search for these articles were combinations of the following: online worlds, virtual worlds, Multi-User Virtual Environments, Massively Multiplayer Online Role Playing Games, Virtual Reality, Second Life, Online worlds, role playing games, Cyberspace, and Immersive Worlds.

Databases searched were ERIC, Academic Search Premier, PsychInfo, Google Scholar, and WilsonWeb. Studies that resulted from this search were refined and focused by using the following criteria: a) They must deal with DVWEs, b) They had to be peer-reviewed, and c) They must be an actual study, and not a simple description of activities. Dissertations were excluded. The authors of these articles examined different audiences and presented widely differing approaches to research in DVWEs. From this selection process, 127 studies were chosen which dealt substantively with DVWEs (see Table 1 for a breakdown by academic field). An article (or presentation) was assigned to a particular field based on the field or academic area that publishes the journal or proceedings. The criteria stated above
assured that the works reviewed were a) representative of works in the different fields, b) have been central or pivotal to the topic area, and c) initiated a line of investigation or thinking.

<table>
<thead>
<tr>
<th>Academic field</th>
<th>Number of articles</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Educational Technology</td>
<td>48</td>
<td>37.8%</td>
</tr>
<tr>
<td>Computer Science or Information Systems</td>
<td>12</td>
<td>9.45%</td>
</tr>
<tr>
<td>Virtual Reality/Virtual Worlds/Gaming</td>
<td>13</td>
<td>10.24%</td>
</tr>
<tr>
<td>Science and Health Education</td>
<td>10</td>
<td>7.87%</td>
</tr>
<tr>
<td>Human-Computer Interaction</td>
<td>10</td>
<td>7.87%</td>
</tr>
<tr>
<td>Media and Communications</td>
<td>9</td>
<td>7.09%</td>
</tr>
<tr>
<td>Psychology or Educational Psychology</td>
<td>8</td>
<td>6.30%</td>
</tr>
<tr>
<td>Other fields</td>
<td>6</td>
<td>4.72%</td>
</tr>
<tr>
<td>Social Science and Educational Research</td>
<td>4</td>
<td>3.15%</td>
</tr>
<tr>
<td>Cognitive Science</td>
<td>4</td>
<td>3.15%</td>
</tr>
<tr>
<td>Library Science</td>
<td>3</td>
<td>2.36%</td>
</tr>
</tbody>
</table>

114 articles, 12 presentations, 1 book chapter

The resulting articles were reviewed in order to describe educational research methodologies being utilized in the investigation and evaluation of DVWEs. The following data were collected from each study: American Psychological Association (APA) reference for the paper itself, publication outlet, academic field (human computer interaction, computer science, educational technology, etc.), research aim, research questions, website address, data protection measures, opportunity cost of research, data collection methods, type of data collected, independent variables, dependent variables, sampling issues, data analysis method, research innovation level, and research goal level (see Table 2 for a more detailed explanation of each category).

<table>
<thead>
<tr>
<th>Type of data collected</th>
<th>Justification</th>
</tr>
</thead>
<tbody>
<tr>
<td>APA Reference</td>
<td>Categorization purposes</td>
</tr>
<tr>
<td>Website address (if any)</td>
<td>To provide resources for future researchers</td>
</tr>
<tr>
<td><strong>Type of data collected</strong></td>
<td><strong>Justification</strong></td>
</tr>
<tr>
<td>---------------------------</td>
<td>------------------</td>
</tr>
<tr>
<td>Type of entry</td>
<td>Categorized by working paper, peer-reviewed article, or peer-reviewed proceedings</td>
</tr>
<tr>
<td>Academic field of study</td>
<td>Anthropology, Biology, Education, etc.</td>
</tr>
<tr>
<td>Research Aim</td>
<td>Aims are broad statements of desired outcomes, or the general intentions of the research, which 'paint the picture' of your research proposal. Research aims address the long-term project outcomes, i.e. they should reflect the aspirations and expectations of the research topic</td>
</tr>
<tr>
<td>Research Question(s)</td>
<td>The specific questions that the study sets out to answer.</td>
</tr>
<tr>
<td>Is security mentioned? In what way?</td>
<td>Discussion involving backup of data, encryption, data access, data integrity, privacy, protection, data storage, etc are included in order to explore the security of data on the server of the virtual environment</td>
</tr>
<tr>
<td>Opportunity cost of research?</td>
<td>Is this study funded? How?</td>
</tr>
<tr>
<td>Type of data collected?</td>
<td>Chat logs, movement data, survey, interviews, etc.</td>
</tr>
<tr>
<td>Independent variables</td>
<td>What was adjusted to make the change occur?</td>
</tr>
<tr>
<td>Dependent variables</td>
<td>What was expected to change?</td>
</tr>
<tr>
<td>Does this study address sampling issues?</td>
<td>The population and sample of virtual world participants is particularly narrow - as regular users do not represent a larger population of students</td>
</tr>
<tr>
<td>Data Analysis method</td>
<td>For example, content analysis, case studies, grounded theory, experimental - ANOVA, ANCOVA, etc.</td>
</tr>
<tr>
<td>Research Innovation Level</td>
<td>According to Ellis (2005)</td>
</tr>
<tr>
<td>Research Goal Level</td>
<td>According to Reeves and Hedberg (2003)</td>
</tr>
</tbody>
</table>
Data were recorded in a spreadsheet and transferred to the NVivo 10.2 software for qualitative analysis. This software was used to help the researchers to organize, analyze, explore, and visualize the data. It helped to uncover subtle connections between data and discover themes that made these connections more explicit. After this, Ellis’ framework (2005) and Reeves and Hedberg’s research goals (2003) were utilized to show where each study “fits” in research, and to reveal any gaps in types of research and methodologies. Data were collected and categorized by multiple researchers in order to ensure trustworthiness and data quality.

4. Results

4.1 Research aims and questions.

The leading research aims and questions focused on affordances of virtual worlds, learning, and instructional strategies. Research that concentrated on affordances examined unique qualities of DVWEs that allow individuals to perform an action. Explorations that focused on learning typically examined outcomes after participants had some kind of interaction with content matter (e.g., art lessons, math lessons, etc.). Where instructional strategies were investigated, researchers tested the logistics of using one or more approaches to teaching a certain topic within the environment. In most cases, these three aims were not found in separate studies but co-occurred in a number of the investigations. The qualities of the environment (e.g., desktop 3D objects, avatars, etc.) were studied to determine their impact on learning or instructional strategies.

To a somewhat lesser extent, other research aims and questions focused on participant behavior, user attitudes, interaction, avatars, collaboration and presence. Also worth noting are the studies that focused on community, engagement, technology adoption, motivation, and teacher perceptions (Table 3). Please see Figure 1 for a visualization of how these differing research aims and questions are related within a study.

<table>
<thead>
<tr>
<th>Aims/questions</th>
<th>Articles</th>
</tr>
</thead>
<tbody>
<tr>
<td>Affordances of VW</td>
<td>56</td>
</tr>
<tr>
<td>Learning</td>
<td>35</td>
</tr>
<tr>
<td>Instructional Strategies</td>
<td>31</td>
</tr>
<tr>
<td>Behavior</td>
<td>21</td>
</tr>
<tr>
<td>User Attitudes</td>
<td>20</td>
</tr>
<tr>
<td>Interaction</td>
<td>18</td>
</tr>
<tr>
<td>Avatar</td>
<td>16</td>
</tr>
<tr>
<td>Collaboration</td>
<td>14</td>
</tr>
<tr>
<td>Presence</td>
<td>12</td>
</tr>
</tbody>
</table>
While the vast majority of studies avoided using the maligned “media comparison study” framework (comparing learning outcomes in different media conditions), at least five studies used this approach.

**Figure 1: Research Theme Connectivity.**

The graphic shows the connectivity of the various research themes as well as their representation in the literature.
4.2 How data were collected

Data collection in virtual worlds can be classified into three categories: 1) through mechanisms which capture in-world data, 2) through back-end data recording mechanisms that track user actions, and 3) through data collection tools external to the virtual world that capture users’ experiences or input about their experiences. The majority of data that were collected among the 127 studies under review did so through external tools (72%). These included the use of discussion boards, participant blogs, electronic pre- and post-tests, web-based surveys, physiological measurements (e.g., eye tracking), and interviews. A smaller number of researchers relied on in-world tools to collect data (23%). The recording of data within a virtual environment often depends on the affordance of the environment itself. For example, in Second Life, investigators can either make, modify, or use objects that accept “notecards,” or act as surveys, or do counts of avatar proximity or length of stay. An examination of the studies found that in-world data mostly was qualitative in scope, including such things as chat logs, images, machinima, or observations either by the researcher or a sensor device. The availability of these data is based on whether or not the researcher has access to the “back-end” of the software (5%). In the case of Second Life, which has proprietary protections on its data, this is not possible. However, those who run servers (e.g., ActiveWorlds) have a tremendous amount of data available to them, all of which is captured passively - meaning that the participant only needs to interact rather than actively use data collection tools.

4.3 What data were collected

Of the data collected for the various studies, a great deal of it yielded qualitative data or nominally quantitative (surveys). Few studies actually had participants submit to pre- and post-test evaluations of their skills or made objective measurements of movement (physical or avatar). As Table 4 shows, the primary source of data included surveys, interviews, or participant-created artifacts, such as blogs.

<table>
<thead>
<tr>
<th>Table 4: Mechanisms Used to Collected DVWE data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data collection mechanisms</td>
</tr>
<tr>
<td>-------------------------------</td>
</tr>
<tr>
<td>In-world mechanisms</td>
</tr>
<tr>
<td>Chat Logs</td>
</tr>
<tr>
<td>Data (In-world)</td>
</tr>
<tr>
<td>Video</td>
</tr>
<tr>
<td>Audio Recordings or Transcripts</td>
</tr>
<tr>
<td>Discussions</td>
</tr>
<tr>
<td>Images</td>
</tr>
<tr>
<td>External Data Collection mechanisms</td>
</tr>
<tr>
<td>Surveys or Questionnaires</td>
</tr>
<tr>
<td>Observations</td>
</tr>
<tr>
<td>Data collection mechanisms</td>
</tr>
<tr>
<td>---------------------------------------------------</td>
</tr>
<tr>
<td>Interview</td>
</tr>
<tr>
<td>Artefacts (External)</td>
</tr>
<tr>
<td>Content knowledge and Skills tests</td>
</tr>
<tr>
<td>Student Reflections</td>
</tr>
<tr>
<td>Measurement of Physical Movement</td>
</tr>
</tbody>
</table>

**Back-end Data Collection mechanisms**

| Data Log Files (not chat)                         | 7                      |

Of the articles reviewed, 114 appeared in journals, while 12 were conference proceedings, and one was a handbook chapter.

The journal articles were found in 67 different journals, with approximately 7% (n=9) of those appearing in the *British Journal of Educational Technology* and another 7% in *Computers & Education*. The fields represented by the journals were not limited to educational technology and virtual reality, but included fields of study such as computer and information science, psychology, language studies, and cultural studies.

### 4.4 Sampling

Looking across studies, one sees a fairly consistent theme in sampling in that most participants tend to be solicited based on their participation in a given course (or courses) and tend to be in groups of about 20 to 50 students total. A number of small, pilot studies or qualitative studies with between approximately five and 10 students are also found. When studies have implemented surveys, or where they have studied usage data (from servers), the number of participants is quite high (up to 4,000 in one case). There are practically no studies done where true random sampling has been done, and this parallels with the fact that very few empirical studies have been carried out. Interestingly, though all of a user's on-screen actions can be recorded by a server, only a few studies report using these data. In proprietary virtual worlds, such as Second Life, the researcher often does not have access to these data beyond using scripted metrics reported by in-world tools or traffic analysis services. This means that a wealth of data is not analyzable. The samples tend to be one of convenience as students in a course are "captive." Except in only a few studies, participants tend not to be diverse in age or nationality, though gender seems to be equally represented across studies.

### 5. Data Analysis

#### 5.1 Variables

Of the 127 total sources, 104 included information on the independent variables used. These variables fell into six major categories: Interface/Environment, attitudinal, demographic, cognitive, teaching method, and behavioral. Sixty of the 127 total sources included information on the dependent
variables used. These variables fell into eight major categories: Learning, in-world choices/tactics, external behaviors, interaction, perception of experience, presence, flow and engagement, and scientific inquiry.

### Table 5: Independent and Dependent Variables

<table>
<thead>
<tr>
<th>Independent Variables</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interface/Environment</td>
<td>35%</td>
</tr>
<tr>
<td>Attitudinal</td>
<td>25%</td>
</tr>
<tr>
<td>Demographic</td>
<td>21%</td>
</tr>
<tr>
<td>Cognitive</td>
<td>7%</td>
</tr>
<tr>
<td>Behavioral</td>
<td>7%</td>
</tr>
<tr>
<td>Teaching Method</td>
<td>5%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Dependent Variables</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Learning</td>
<td>23%</td>
</tr>
<tr>
<td>In World Choices/Tactics</td>
<td>22%</td>
</tr>
<tr>
<td>External Behaviors</td>
<td>12%</td>
</tr>
<tr>
<td>Interaction</td>
<td>12%</td>
</tr>
<tr>
<td>Perception of Experience</td>
<td>10%</td>
</tr>
<tr>
<td>Presence</td>
<td>8%</td>
</tr>
<tr>
<td>Flow and Engagement</td>
<td>7%</td>
</tr>
<tr>
<td>Scientific Inquiry</td>
<td>6%</td>
</tr>
</tbody>
</table>

**5.2 Privacy and Security.**

According to common human subjects’ research protocols, researchers need to describe how they have made explicit to participants the risks to data being electronically transmitted and stored. Although participant protections are required for institutional review board applications, a clear explanation of the privacy and security concerns does not appear in many research reports. Only 14% (n=18 papers) of the studies examined directly discussed data security issues outside of the required need to change participant names (or avatar names) for reporting purposes. The articles that did cover concerns related
to security indicated to their participants that public chat in virtual spaces is public, or talked about the logistics for backing-up data on another server.

5.3 Funding

A total of 29% (n=37) of the studies were supported by outside agencies, such as the National Science Foundation, private foundations, or consortia. It may seem that there should be more externally funded studies, especially if one compares research on the implementation of other, more established technologies (note: no data are available on the average percentage of peer-reviewed publications that are funded). This lower level of funding may be due to the relative newness of DVWEs in research and the overall lack of empirical studies done to date.

5.4 Virtual Environments Used

The virtual environments reviewed were comprised of commercial MMORPGs (24) such as World of Warcraft, Everquest, and Star Wars Galaxies. The largest group of studies were done using commercially available virtual worlds (92), which included Second Life (47) and Active Worlds (31). It should be noted that in several of these studies, researchers took the Active Worlds platform and customized it for the sake of their own projects (e.g., River City at Harvard and Quest Atlantis at Indiana).

Other virtual environments reviewed included MMORPGs (4) and virtual worlds (7) developed completely within a non-profit institution such as a university. Several of the studies examined two or more virtual environments: text-based virtual worlds (3), institutionally-based game (3), and CAVE-like desktop system (13).

5.5 Ellis’ Levels of Research

As explained in the theoretical framework section, Ellis (2005) explains three levels of research. Level I research is theory building, pure research, or both. Level II research is empirical research, either quantitative, qualitative, or a combination thereof. Level III research determines the extent to which a program (curriculum innovation) is successful when its implementation becomes widespread.

Seventy eight of the studies reviewed were categorized as level I research by Ellis’ (2005) research framework. This means that 61% of the studies reviewed focused on theory building, pure research, or both. They were mostly conducted in experimental or artificial lab settings. They included both qualitative and quantitative studies, and used analyses such as correlations, descriptive data, and case studies.

Forty six of the studies reviewed were level II research (Ellis, 2005), indicating that approximately 36% of the studies were either quantitative or qualitative empirical research or mixed methods. Their purpose was to determine the efficacy of particular programs or methods in authentic settings. This kind of research seeks to apply theories and procedures developed in Level I research and provide practical insights.

About 3% (3 studies) were categorized as level III research (Ellis, 2005). These studies examined a program evaluation in which the researchers examined whether the program is successful when its implementation becomes widespread.

Reeves and Hedberg (2003) hail the importance of developmental research due to its use of the empirical method to examine authentic, complex problems in a realistic context. Interestingly, 6.3%
(8/127) of the reviewed studies were categorized as developmental research, and these 8 articles represent research done using four different DVWEs (River City, Quest Atlantis and EcoMUVE). With that said, the most prevalent research is still being accomplished by researchers with interpretivist (41.7%, 53/127 studies) and predictive (35.4%, 45/127 studies) research goals. Other studies were scattered among theoretical (3.9%, 5/127 studies), action (3.9%, 6/127 studies), postmodern (2.3%, 3/127 studies), and mixed goal (6.3%, 8/127 studies) research goals.

6. Discussion

According to Ellis’ (2005) framework for research in new fields of education, Level III research is almost completely lacking from our study (3/127 studies). This means that there has been virtual no DVWE research on the level of program evaluation where it becomes possible to learn the extent to which a program is successful when its implementation becomes widespread. The three studies that were categorized at this level were led by senior educational technology researchers who had customized a DVWE for their research. These DVWEs (River City, EcoMUVE, and Quest Atlantis) could lead the way by converting their application to open source and making it freely available to anyone in exchange for participation in research. This would most likely result in a larger number of teachers, classrooms, and schools playing a part in program evaluation on a grand scale, and would help to answer the question of whether the promising results that have been discovered with less numerous samples will translate to a population of a larger magnitude.

As Reeves & Hedberg (2003) state, postmodern research goals are a rarity in the field of educational technology, and our literature review reflects this (only 2.3% of articles had a postmodern goal). Atypical to this trend, Beck (2011) examines the influence of student avatar choice on teachers’ expectations and perception in his study of over 450 teachers in the DVWE Second Life. With that said, DVWE researchers should pursue collaborative research opportunities with postmodern, education researchers in order to shed more light on important issues of equity, gender, politics, and culture that impact the field.

The presence of developmental research in DVWEs is encouraging (8/127 studies, 6.3%) because of the authentic, complex problems that can be addressed in these type of environments (Reeves & Hedberg, 2003). Moreover, the need to implement these kinds of environments in multiple contexts also lends itself to an approach that tests, retests and refines the DVWE as well as to define new design principles. Three major projects (Quest Atlantis, River City, and EcoMUVE) and approximately seven primary researchers are not enough to act as a catalyst for this important research goal. Unfortunately, the opportunity cost of this type of research is high as all of the studies were funded by grants from the National Science Foundation, setting a high bar on the amount of funding, type of institution, and quality of researcher that is able to pursue this type of grant. These researchers should consider converting the proprietary DVWEs they authored into open-source applications that can be easily utilized by other researchers and institutions who might not have access to NSF funds to create their own. This access could come with a professional development requirement that would enable the current researchers to train others in developmental research skills. Such actions would also likely have an overflow effect of increasing the number and quality of other types of research goals, such as action and critical.

Finally, research with a theoretical research goal was also minimal (5/127, 3.9%). This is concerning because it points to a lack of logical analysis and synthesis of theories in DVWE research. Theoretical research can often provide a theoretical road map that other types of research can follow. Additionally, other research with differing research goals will lack the focus necessary to make a larger
impact without theoretical research. For example, the large amount of predictive (45/127 studies, 35.4%) and interpretive (53/127 studies, 41.7%) research may “miss the target” provided by theoretical research. As a result, it is important for DVWE researchers to pursue this type of research.

7. Future Studies

Hew & Cheung (2010) suggest that future research should provide a rich, thick description of the methodology, including the duration of the study, interobserver and intraobserver agreement reliability and effect sizes so that findings can be adequately interpreted.

The APA Task Force stressed that researchers should provide some effect-size estimate such as Cohen’s $d$ when reporting a $p$-value because reporting and interpreting effect sizes are essential to high quality, empirical research (Wilkinson & APA Task Force on Statistical Inference, 1999). Kotrlik and Williams (2003) argued that reporting effect size allows a researcher to judge the magnitude of the differences present between groups, thus increasing the capability of the researcher to judge the practical significance of the results derived.

Future studies should also be longitudinal in nature, extending the duration to perhaps more than one year. Doing longitudinal studies provides researchers with the opportunity to examine not only whether students’ and teachers’ perceptions of virtual worlds undergo change but also whether there are any detrimental effects of using virtual world environments over a long period of time, which has not been investigated up to this point.

Our research showed that there are already sufficient level I research studies (78/127) and a growing number of Level II research studies (46/127). Future research should continue to focus on Level II research that determines the efficacy of particular programs or methods in authentic settings. This level of research is a prerequisite for Level III research and is one of the reasons why very few Level III studies have been conducted (3/127).

As stated previously, Level III research is on the level of program evaluation where it becomes possible to learn the extent to which a program is successful when its implementation becomes widespread. Will there ever be a situation where there are enough Level III studies to provide sufficient evidence? Or do the logistical circumstances such as large start-up costs, lack of authentic implementation context, labor, and subject matter expertise prohibit such a situation from ever occurring? Future research should strive to overcome these logistical difficulties through collaboration with the individuals who are already performing Level III research, and they should also seek to form new partnerships with other researchers who are interested in Level III research.

Our research also showed that most of the research being done with DVWEs has predictive (35.4%, 45/127 studies) and interpretive (41.7%, 53/127 studies) research goals. Researchers who are seeking to find areas that need to be addressed in this field should consider DVWE research with theoretical, action, postmodern, and developmental research goals. Because of the pressing need for innovation and thoughtful technology integration in authentic learning environments such as schools, action and developmental research goals should receive particular emphasis (Reeves and Hedberg, 2003). Implementation and evaluation studies, however, cannot be initiated simply because a gap exists as a number of antecedents must necessarily be in place for such studies to be possible.

Initiating research in a relatively new field of educational technology can be fraught with peril. Traditional methods of data collection and analysis may not automatically integrate into the new technological environment. Unproven methods may appear to work better, but usage of these methods
will lengthen the study cycle. Moreover, usage of unproven methods may be shunned by tier one journals, thus relegating the research to less impactful journals and potentially affecting the tenure-seeking process of assistant professors.

As a result, it is important to approach a new field of educational technology with as much knowledge as possible of the research methodologies being used. This is especially true in the relatively new field of DVWEs, where pioneering research is being done without much knowledge of what has been done and what works well.
References


Appendix A: List of Virtual World Research Articles Analyzed for this Paper


Barab, S., Dodge, T., Gentry, E., Saleh, A., & Pettyjohn, P. (2011). Uganda’s road to peace may run through the river of forgiveness: Designing playable fictions to teach complex values. (pp. 312-333). In K. Schrier & D. Gibson (ed.) Ethics and game design: Teaching values through play (Vol. II). IGI Global, Hershey, PA.


Casanueva, J. S. & Blake, E. H. (2001). The effects of avatars on co-presence in a collaborative virtual environment. Annual Conference of the South African...


