Computers & Education 70 (2014) 29–40

Contents lists available at ScienceDirect

Computers & Education

journal homepage: www.elsevier.com/locate/compedu

Effectiveness of virtual reality-based instruction on students' learning outcomes in K-12 and higher education: A meta-analysis

Zahira Merchant [a,*], Ernest T. Goetz [b], Lauren Cifuentes [c], Wendy Keeney-Kennicutt [d], Trina J. Davis [a]

[a] College of Education, Department of Teaching Learning and Culture, Texas A&M University, College Station, TX 77843-4232, USA
[b] College of Education, Department of Educational Psychology, Texas A&M University, College Station, TX 77843-4225, USA
[c] Department of Distance Learning, Texas A&M University, Corpus Christi, TX 78412, USA
[d] College of Science, Department of Chemistry, Texas A&M University, College Station, TX 77843-3225, USA

A R T I C L E I N F O

Article history:
Received 26 October 2012
Received in revised form
3 July 2013
Accepted 29 July 2013

Keywords:
Simulations
Games
Virtual worlds
Students learning outcomes
Meta-analysis

A B S T R A C T

The purpose of this meta-analysis is to examine overall effect as well as the impact of selected instructional design principles in the context of virtual reality technology-based instruction (i.e. games, simulation, virtual worlds) in K-12 or higher education settings. A total of 13 studies ($N = 3081$) in the category of games, 29 studies ($N = 2553$) in the category of games, and 27 studies ($N = 2798$) in the category of virtual worlds were meta-analyzed. The key inclusion criteria were that the study came from K-12 or higher education settings, used experimental or quasi-experimental research designs, and used a learning outcome measure to evaluate the effects of the virtual reality-based instruction.

Results suggest games (FEM = 0.77; REM = 0.51), simulations (FEM = 0.38; REM = 0.41), and virtual worlds (FEM = 0.36; REM = 0.41) were effective in improving learning outcome gains. The homogeneity analysis of the effect sizes was statistically significant, indicating that the studies were different from each other. Therefore, we conducted moderator analysis using 13 variables used to code the studies. Key findings included that: games show higher learning gains than simulations and virtual worlds. For simulation studies, elaborate explanation type feedback is more suitable for declarative tasks whereas knowledge of correct response is more appropriate for procedural tasks. Students performance is enhanced when they conduct the game play individually than in a group. In addition, we found an inverse relationship between number of treatment sessions learning gains for games.

With regards to the virtual world, we found that if students were repeatedly measured it deteriorates their learning outcome gains. We discuss results to highlight the importance of considering instructional design principles when designing virtual reality-based instruction.

Published by Elsevier Ltd.

1. Introduction

The advent of highly immersive virtual reality technology can be traced back to the 1960's in the entertainment industry with Morton Heiling's single user console called *Sensorama*, designed to captivate audience attention (Heiling, 1962). In the 1980's, a dramatic surge of interest in using virtual reality technology beyond the entertainment industry was seen in the field of professional education and training. Particularly, virtual reality technologies frequently were used for flight simulator training and exercises (Hawkins, 1995). The introduction of virtual reality technology in K-12 and higher education began in the early 1990's with projects such as Science Space, Safety World, Global Change, Virtual Gorilla Exhibit, Atom World, and Cell Biology (Youngblut, 1998). Designers of these projects used various peripheral devices such as head-mounted display gear, data gloves, and body suits for a fully immersive learning experience. The techniques employed in these virtual environments ranged from using specially designed glass cubicles called Cave Automatic Virtual Environment (CAVE) to projecting on the walls of a room (Cruz-Neira, Sandin, & DeFanti, 1993). However, the literature reports many practical concerns and limitations that restricted wide spread dissemination of this technology in K-12 and higher education settings.

* Corresponding author. Tel.: +1 979 402 5504.
E-mail addresses: zhmerchant@tamu.edu, momentszahira@gmail.com (Z. Merchant).

0360-1315/$ – see front matter Published by Elsevier Ltd.
http://dx.doi.org/10.1016/j.compedu.2013.07.033

One of the many reasons why virtual reality technology was beyond the reach of schools was financial feasibility (Andolsek, 1995; Mantovani, Gaggiolo, Castelnuova, & Riva, 2003; Riva, 2003). The cost of both procurement and maintenance of various sophisticated devices to create an immersive environment made mass use of this technology prohibitive. In addition, there were many physical and psychological discomforts that users experienced in virtual reality environments. These included strenuous posture demands, repetitive strain injuries, headset weight and fit, simulator sickness, disorientation, hallucination, and dissociation (Costello, 1993). Another significant concern identified in the literature was poor instructional design of the virtual learning environments (Chen, Toh, & Ismail, 2005; Riva, 2003; Wong, Ng, & Clark, 2000).

Despite the problems of early virtual reality technologies, the rapid increase in the processing power of the computer led to the deployment of desktop-based virtual reality technology in K-12 and higher education. The drastic reduction in the cost of technology and availability of high-speed Internet connection further increased the use of this less immersive form of virtual reality technology (Dickey, 2005; McLellan, 2004). Although desktop-based 3-D virtual environments cannot provide fully immersive experience, their photo realistic computer graphics been shown to enhance learners' engagement (Dickey, 2003). Advances in the technology have made it possible to use low cost peripheral devices such as headphones, shutter glasses, and data gloves. In addition, with the further advancement of Web technologies, new possibilities of simultaneously allowing more than one user in a virtual environment to work collaboratively have also emerged (Chen & Teh, 2000; Kamel Boulos & Wheeler, 2007).

The assumption underlying the rapid rise in the use of desktop-based virtual reality technology in instruction is the unique affordances that it offers in enhancing learners' cognitive skills. Many educators have integrated a variety of desktop-based virtual reality technologies into their instruction. For example, educators have used a very popular virtual world called Second Life®, to create replicas of real life places wherein users, who are digitally represented in form of avatars, actively engage in realistic activities that stimulate learning. In addition, educators have used Second Life's affordances to build 3-D objects to teach abstract concepts (Merchant et al., 2012). River City is an interactive computer simulation for middle school science students to learn scientific inquiry and 21st century skills (Galas & Ketelhut, 2006). Other simulations include Vfrog™, in which students dissect a virtual frog (Lee, Wong, & Fung, 2010); MAT^{3D} which teaches high school students mathematical concepts (Pasqualotti & Freitas, 2002). DimensionM™ is a 3D video game in which students embark on a journey where they accomplish series of mission applying mathematics principles (Kebritchi, Hirumi, & Bai, 2010). Another video game designed by students of mechanical engineering is to race a simulated car on the track. In this designing process students write a computer program and learn about the concepts such as thermo dynamics (Coller & Shernoff, 2009).

1.1. Defining simulations, games, and virtual worlds

Simulations are interactive digital learning environments that imitate a real-life process or situation. Simulations allow learners to test their hypotheses of the effects of input variables on the intended outcomes (De Jong, 1991; Lee, 1999; Tobias & Fletcher, 2010). Simulations can provide cost-effective practice of procedures using virtual apparatus that in real life could be cost prohibitive. For example, frog dissection is a commonly used procedure to teach anatomy in high school biology classes. Vfrog™ is a popular simulation that allows students to conduct frog dissection numerous times using virtual apparatus. Conducting dissection procedures physically in a laboratory may not only impose financial burden, but also be inconsistent with students' personal beliefs about conducting animal dissections. Simulations are also advantageous because they can allow learners to practice skills that otherwise could be dangerous to practice in the real life situation, in a safe environment. For example, in the medical field, Mr. Vetro™ is a commonly used simulation of several medical scenarios that provides students the opportunity to sharpen their skills before practicing it on real life patients. In this way, medical students can avoid the risk of applying certain procedures directly on the patient without having sufficient practice, which may endanger patients' life.

Researchers have assigned games for learning as a special category of simulation (Tobias & Fletcher, 2010). Research suggests that in order to promote learning, games must be design to provide players with a sense of autonomy, identity, and interactivity (Gee, 2003). In order for a game to provoke learners' long-lasting motivation and prolonged engagement with the learning materials (Gee, 2007), the design of the game must provide learners with the opportunities to strategize their moves, test hypotheses, and solve problem (Ang & Krishna, 2008; Dondlinger, 2007). In order to provide these experiences, game must include elements such as goals, achievement levels, and rewards systems. The game designers also consider narrative plots, which lead the player into the game as an integral aspect of an effective game design. The games may also consist of animated agents that inform the players about the context of the game, rules of the game as well as provides interactive cues and feedback. However, game designers firmly states that narrative plots must be embedded within the learning context of the game and not something, which are overlaid and disconnected from the learning goals. Csikszentmihalyi's (2000) flow theory provides a framework for interpreting the effectiveness of games to engage players and motivate them to sustain in the play. If the game is too challenging, the player will be frustrated, and if it's too simple, the player will lose interest. In either case, players are very likely to become disengaged and quit the game play.

Virtual worlds, according to Dickey (2005) and Hew and Chung (2010), may contain one or more of the following features: the illusion of being in a 3-D space, ability to build and interact with the 3D objects, digital representation of learners in form of avatar, and ability to communicate with other learners in the virtual worlds. Contrary, to the structured environment of simulations and games, virtual worlds are open-ended environments in which users design and create their own objects.

The rapid increase in the technological sophistication, diversity, and pervasiveness of 3D virtual learning environments, along with the proliferation of research on their effectiveness in educational settings, necessitates frequent systematic analytical syntheses of their effectiveness. Few meta-analyses or other reviews have been conducted to date.

1.2. Summary of previous reviews and need for the current meta-analysis

A search of the literature revealed three meta-analyses (Lee, 1999; Sitzmann, 2011; Vogel et al., 2006) and a systematic review summarizing qualitative research on 3D virtual worlds (Hew & Cheung, 2010). Lee conducted a meta-analysis of 19 studies and found a positive impact of using simulation on learning outcomes but a negative impact on students' attitudes towards using this form of technology for learning. The major focus of Lee's meta-analysis was on moderator variables such as mode of simulation (presentation or practice), presence of expository instructional features, and guided versus unguided simulations. According to the results of this meta-analysis, simulations are effective for both presentation and practice if used in conjunction with other methods of instruction. Lee also found that students' performed better when some form of guidance was provided even in practice mode compared to those where there was no guidance provided.

More recently, Sitzmann (2011) and Vogel et al. (2006) conducted meta-analyses in which they analyzed the effects of interactive computer-based games and simulations and found statistically significant positive impacts on learning outcomes. Vogel et al. studied the moderation effects of gender, learner control, age, realism, and learner collaboration on learning outcomes. According to their report, students performed better when they were in control of their navigation through the virtual learning environment compared to when the teacher controlled the learning environment. In addition, students in the traditional group outperformed the students in the virtual learning environment when, the sequence of learning activities was controlled by the computer programs compared to when students could select the sequence.

Sitzmann (2011) focused on the effects of games and simulations in enhancing work-related knowledge and skills, examining variables such as entertainment value, type of control group treatment, access level, mode of instruction, and methodological quality. According to the outcome of this study, Sitzmann reported the highest gain in the measure of self-efficacy (20%) as compared to procedural knowledge (14%), declarative knowledge (11%), and retention (9%). The virtual environmental characteristics such as active presentation of materials, unlimited access level to the learning materials, and presentation of the materials in a supplemental format were more effective.

Hew and Cheung (2010) conducted a systematic literature review on the use of virtual worlds in the context of K-12 and higher education (primarily, university or polytechnic settings) in which 14 out of the 15 studies included were descriptive in nature. Their review examined virtual worlds' literature in three areas: uses of virtual worlds by students and teachers, types of research methods applied to study the effects of 3-D virtual worlds, and kinds of topics researched in 3-D virtual worlds. The studies reviewed were mostly descriptive in nature. The results of this review indicated that 3D virtual worlds are used as communication spaces, simulation spaces, and experiential spaces. Several different kinds of topics are researched in 3-D virtual worlds categorizes into participants' affective domain, learning outcomes, and social interaction.

Our study contributes to the field of desktop-based virtual reality technologies for instructional use in several ways. First, Lee's (1999) meta-analysis focused on assessing the effectiveness of simulations. Moreover, Sitzmann (2011) collapsed both simulations and games into a single category and called it "simulation games." This may pose some concerns because simulations and games have different design features, and it is important to study possible differences in their effects on the learning outcomes. Unlike, Sitzmann, Vogel et al. (2006) identified simulations and games into separate categories. Like Vogel et al., we also differentiated between simulations and games. In addition, we expanded the scope of this meta-analysis to include virtual worlds, which is one of the most rapidly emerging and popular forms of desktop-based virtual reality technology.

Second, Sitzmann (2011) focused on synthesizing the effects of games and simulations in the area of enhancing work-related knowledge and skills. On the other hand, Vogel et al. (2006) included studies related to both work place and educational settings; however, their study did not decompose the effects of each setting separately. We believe that both work-related training and education training differ and should be studied independently. Therefore, our meta-analytical examination focused on instructional effectiveness in K-12 and higher education settings. Third, we also analyzed the moderating effects of variables central to the field of instructional design and are discussed in the following section. These design features are not analyzed in some of the previous meta-analysis such as feedback, students' level of collaboration, teacher access, and novelty effect. In addition, we also examined the relationship between studies research design quality and reported effect sizes.

Finally, the most recent studies included in the previous meta-analyses were published in 2009 (Sitzmann, 2011), and one of the meta-analyses is more than a decade old (Lee, 1999). Our review included studies until 2011. This will not only provide the insight about the current literature on desktop-based virtual reality technologies but will also serve as a comparative analysis for examining the rapid changes in the power of computer technology and the enhancement of learning effectiveness afforded by the technology power.

2. Purpose

We undertook a meta-analysis to address some of the limitations of the previous reviews. The primary purposes were (a) to examine the overall effectiveness of desktop-based virtual reality technology in K-12 or higher education settings and (b) to identify key instructional design principles in the context of desktop-based virtual reality instruction on the learning outcomes. In order to achieve this purpose we conducted three separate analyses for games, simulation, and virtual worlds. Thirteen variables were coded for each study to answer the following seven research questions:

1. (a) Which kind of learning outcomes are more suitable for desktop-based virtual reality instruction? (b) Are the effects of virtual reality-based instruction on learning outcome measures moderated by the type of learning task? (c) Did the testing conditions impact the learning outcome gains?
2. Are the learning outcome gains higher in desktop-based virtual reality instruction as compared to the other methods of instruction?
3. Did forms of instruction impact the learning outcome gains in desktop-based virtual reality learning environment?
4. (a) Did the availability of teachers enhance the learning outcome gains in desktop-based virtual reality learning environment? (b) Did students' collaboration impact the learning outcome gains in desktop-based virtual reality learning environment?
5. Did the novelty effect impact the learning outcome gains in desktop-based virtual reality learning environment?
6. Which kind of feedback is more suitable for a particular type of learning tasks in desktop-based virtual reality instruction?
7. Did the methodological rigor moderate the learning outcomes gains?

3. Method

In the current meta-analysis, we integrated available studies that assessed the relationship between desktop-based virtual reality instruction and learning outcomes in K-12 and higher education. We followed the meta-analytical procedure suggested by Glass, McGaw, and Smith (1981). Their procedure requires a meta-analyst to (a) collect studies, (b) code characteristics of studies, (c) calculate effect sizes of each study's outcome measure on a common scale, and (d) investigate moderating effects of study's characteristics on the outcome measure.

3.1. Data sources and search strategies

The following strategies were employed to identify empirical studies to include in the meta-analyses:

1. Electronic searches were performed on the following databases: PsycINFO (EBSCO), Medline (Pub Med), Dissertation and Theses, Eric (EBSCO), Education Full Text, PaperFirst, and CINHAL (The Cumulative Index to Nursing and Allied Health).
2. Manual searches were performed in relevant journals including Educational Technology, Research and Development, British Journal of Educational Technology, Australian Journal of Educational Technology, Computers & Education, Educational Technology and Society, International Journal of Computer-Supported Collaborative Learning, Journal of Computer Assisted Learning, VR in Schools, Virtual Reality, Presence, Journal of Technology Education, and Journal of Virtual World Research.
3. Web searches were conducted using the Google Scholar search engine.
4. Branching searches were performed using forward and backward search procedures from the reference lists of the empirical studies that were located in earlier stages of the review.
5. Complied reference lists available online on the topic of virtual reality were searched. This includes Youngblut (1998), Emerson and Revere (1997), and Fallman (n.d.) as well as relevant reviews found during the electronic database search.
6. The first author personally contacted the scholars who have conducted extensive research in the field of virtual reality technologies.
7. Search terms for empirical studies included *virtual reality*, *virtual worlds*, *virtual learning environments*, *computer assisted learning*, *artificial intelligence*, *mixed reality*, *synthetic environment*, *virtual classrooms*, *augmented reality*, immersive learning environment, computer games, game-based learning environment, serious games, simulations; these were combined with other terms such as *education*, *learning*, *instruction*, and *instructional design.*

3.2. Inclusion and exclusion criteria

Studies were either included or excluded based on their consistency with the following criteria.
The following criteria were used to include studies in the meta-analysis:

1. Studies found until November 2011.
2. Studies that used samples from a population of K-12 or higher education settings.
3. Studies that used virtual reality-based instruction in form of games, simulation, or virtual worlds.
4. Studies that measured learning gains as an outcome variable using test instruments, observation of student's performance, and student's work samples.
5. Studies that used experimental control group research design to measure relationships between desktop-based virtual reality instructions with learning gains.

The following criteria were used to define the set of studies to be excluded from the meta-analysis:

1. Studies that were published in languages other than English.
2. Studies that used desktop-based virtual reality technologies as an assessment, diagnostic, or therapeutic tool.
3. Studies that did not provide sufficient data for effect size calculation.

3.3. Study sample

An initial search yielded an outcome of 7078 articles that matched the key word searches criteria. After judging the abstract of these articles, 102 were included for further consideration in the study. The first author read each full-text article to conclude the process of selecting the qualifying studies. Finally, a total of 67 studies qualified to be included in the meta-analysis study. To ascertain the reliability of the coded variables, the first author coded all the studies and the second author coded 20% (67) of all the studies included in this meta-analysis. The inter-rater reliability of the studies coded by both coders ranged between 80 and 100% on the coded variables. Any disagreements on the coded variables were discussed until the coders reached to a mutually acceptable decision.

3.4. Dependent variable and effect size calculation

The dependent variable in all 67 studies was a learning outcome measure. A two-step procedure described by Hedges and Olkin (1985) was used: first, effect size per study was calculated, and second, optimal weights based on the standard error of the effect sizes were computed. As a result, the effects sizes are comparable across all the studies included in the meta-analysis. We primarily selected means and standard deviations; when those were not available, effect sizes were calculated based on F tests, t tests, Chi-square, or p values.

In calculating the meta-analysis effect sizes, we included only one effect size per study. According Lipsey and Wilson (2001), when a study contributes more than one effect size in the analysis, it leads to statistical dependence, resulting in a biased overall effect size. The following rules were used in deciding which effect size to include in the analyses.

1. When a study assessed the same construct using more than one outcome measure (e.g., Ainge, 1996; Antonietti & Cantoia, 2000; Hauptman, 2010; Michael, 2001; Rafi & Samsudin, 2009; Sun, Chan, & Meng, 2010), we averaged the effect sizes. For example, Rafi and Samsudin (2009) used mental rotation accuracy and mental rotation speed tests as measures of students' spatial ability.
2. When a study reported effect sizes for multiple variables selected for coding and analysis, we selected the effect size based on the moderator variable needed in order to conduct the analyses for each type of virtual reality environment. For example, Copolo and Hounshell (1995) compared the effects of desktop-based virtual reality treatment against three different control group treatments. We selected the control group that was given "combination treatment" using both computer-based 3D models and 3D concrete models of molecular structures because we wanted to include different varieties of the control group treatment. Hu et al. (2009) conducted a study of virtual worlds in which they reported effect sizes for test of theory and work samples, respectively. We included the effect size for work samples because we wanted to include the category of skill-based measures under the variable of type of learning outcome measures.

3.5. Moderator variables

In order to answer the research questions raised in this study, we coded for the following moderator variables. To answer the first research question, we coded studies on the variables of learning outcome measures, type of learning tasks, and testing condition. Sitzmann (2011) coded studies on the variable of "learning outcomes" to categorize based on declarative knowledge, procedural knowledge, retention or transfer. We sought to provide a more fine-grained analysis. Therefore, we coded the studies on learning outcome measures delineating three categories of "knowledge-based", "abilities-based", or "skill-based measures". On the variable of learning tasks, studies were coded either as "declarative" or "procedural". Studies that imparted instruction based on factual information were coded as declarative, and studies of instruction in which learners acquired knowledge about performing a task were treated as procedural.

We coded the study based on the "testing condition". Studies on this variable were coded into four categories of "immediate", "delayed", "repetitive", or "transfer". Studies were coded as immediate when the learning outcome measure was administered immediately after the intervention. Studies were categorized as delayed when there was a time interval between the instructional activity and the administration of measures of learning outcomes. This time interval ranged between a class period, next day, end of semester or 40 days later. We categorized the studies as repetitive when studies administered the measure of learning outcome twice (i.e. immediate and delayed). Studies were categorized as transfer when the context of applying the concept was different than the one presented in desktop-based virtual reality instruction.

To answer the question of whether desktop-based virtual reality instruction was better compared to the other forms of instruction, we coded the studies on the variable of "control group treatment". Both Lee (1999) and Sitzmann (2011) coded their studies on the variable of control group treatment, but our categorization covered more forms of control groups. The categories created for coding the control group treatment were "traditional", "multimedia", "combination", or "no treatment". Studies were classified as using traditional instruction when they employed one or more form of these methods: lecture, textbook, paper-based exercise, 3D concrete models, or physical lab sessions. Studies were assigned to multimedia when they used instructional modalities such as videos, graphics, or computer-based tutorials. Studies that imparted instruction partially using desktop-based virtual reality instruction and traditional or multimedia methods were assigned to the category of combination. Studies in which control group received no special treatment and were used only for the purposes of comparing the instructional effectiveness of desktop-based virtual reality instruction were assigned to no treatment.

For the third research question, we coded the studies on the variable of "mode of instruction" based on the sequence in which the desktop-based virtual reality instruction was presented. We coded the studies into three categories: presentation, practice, or stand-alone. Studies were categorized as "presentation" when desktop-based virtual reality instruction was used for introducing a concept. Studies were categorized as "practice" when learners used desktop-based virtual reality instruction to apply the concept introduced to them using other forms of instruction prior to using desktop-based virtual reality tools. Finally, studies were classified as "stand-alone" when the previous form of instructional method was completely replaced by desktop-based virtual reality instruction.

To answer the fourth research question, we coded the studies on whether teacher's access was available during the instructional activity or if it was a student directed learning activity. The studies were also coded on whether students had completed the learning task working in collaboration with each other or whether the students worked individually.

One of the key issues in the desktop-based virtual reality instruction is the amount of time spent within the computer environment. According to Clark (1989), higher learning gains may not be achieved due to the instructional methods used (i.e. computer-based instruction) but due to the presence of novelty effect. According to novelty effect proposition, learners perform better because of the new technology instituted in the instructional method and not because of enhanced teaching and learning quality. In order to answer the research question about the presence of novelty effect when comparing the learning effectiveness between desktop-based virtual reality environment and other methods of teaching, we coded the studies on three different, but related variables. These variables were number of treatment sessions, duration of each session in minutes, and amount of total time spent in minutes.

We coded the studies on the variable of "feedback" learners received during their interaction with the virtual environment to answer the sixth research question. According to McNamara, Jackson, and Graesser (2009), the type of feedback used is an important factor in determining the effectiveness of virtual learning environments designed for teaching and learning purposes. We categorized studies into three different categories. The categories were knowledge of result or response, elaborate explanation, or visual clues.

We coded the studies on the variables of research design quality and type of measure (researcher-developed vs. standardized) to assess their methodological rigor. According to Lipsey and Wilson (2001), substantive effects found by a meta-analyst may actually be artifacts of confounded methodological variables. Therefore, it is important that the studies are assessed on their methodological strength. We used the model developed by Allen, Chen, Willson, and Hughes (2009) to assess the research design quality of the studies included in the meta-analysis, with some modifications to their model to suit the context of our study. According to our revised model, a study that employed "true experimental" research design were treated as "high quality". The studies employing other forms of design (i.e., quasi-experimental or biased) were further screened on two criteria to determine the quality of their methodological design. These two criteria include "quality of control group treatment" and "quality of statistical control". In addition, we also coded the studies on the variable of whether the instruments used in the studies were researcher-developed or standardized. Studies in which measuring instrument was developed specifically for the study were categorized as "researcher-developed" and studies that used pre-validated instruments were treated as "standardized".

3.6. Data analysis

We used the Comprehensive Meta Analysis 2.0 software package (Borenstein, Hedges, Higgins, & Rothstein, 2009) for effect size synthesis and moderator analyses. According to Borenstein, Hedges, Higgins, and Rothstein (2010) random effects model is more appropriate when the effect sizes of the studies included in the meta-analysis differ from each other. However, for readers benefit, we conducted analysis using both the random effects model (REM) and the fixed effects model (FEM) to calculate the pooled effect size. Heterogeneity was calculated with the Q statistic and the I^2 statistic. A significant Q rejects the null hypothesis of homogeneity and indicates that the variability among the effect sizes is greater than what is likely to have resulted from subject-level sampling error alone (Lipsey & Wilson, 2001). We also calculated I^2, which describes the percentage of total variation across studies that is due to heterogeneity rather than chance. An I^2 value of 25% is associated with low heterogeneity, 50% is associated with moderate heterogeneity, and 75% is associated with high heterogeneity (Higgins, Thompsin, Deeks, & Altman, 2003). Post hoc subgroup analyses were conducted with the mixed effects analysis (MEA) as implemented in the Comprehensive Meta-analysis software. In the MEA, the REM is used to calculate the effect size for each subgroup, while the FEM is used to test the difference between the subgroups of studies (Hedges & Pigott, 2004).

4. Results

We conducted three distinct meta-analysis of studies based on their categorization as games, simulations, or virtual worlds. Table 1 presents the descriptive features for each category (i.e. games, simulations, and virtual worlds) of desktop-based virtual environment.

For the 13 studies that investigated the instructional effectiveness of games, a REM analysis for the relationship between game-based instruction and learning outcome gains resulted in a mean effect size of 0.51 ($SE = 0.13$; 95% confidence interval 0.25–0.77); while the FEM analysis resulted in a mean effect size of 0.77 ($SE = 0.03$; 95% confidence interval 0.69–0.85). The effect sizes ranged from −0.16 to 1.17. Eight of the studies (62%) showed statistically significant positive effects (i.e., game-based instruction increased learning outcome gains); three of studies (23%) produced statistically significant negative results, and two (15%) failed to reveal statistically significant effects between the virtual learning environments and the control groups. The hypothesis of homogeneity was rejected because a statistically significant Q value indicated presence of heterogeneity among the studies ($Q = 113.56$, $I^2 = 89.43\%$).

A careful examination of the effect sizes calculated for studies of the efficacy of simulations-based instruction revealed that a study by Papaevripidou, Constantinou, and Zacharia (2007) had an effect size that was three standard deviations below the mean (Hedges $g = -6.44$), which is likely to bias the overall meta-analysis effect size. Therefore, it was considered as an outlier and was removed from further analyses. The REM analysis for the relationship between simulation-based instruction and learning outcomes resulted of the 29 remaining studies revealed a mean effect size of 0.41 ($SE = 0.11$; 95% confidence interval 0.18–0.64); the FEM analysis resulted in the mean effect size of 0.38 ($SE = 0.04$; 95% confidence interval 0.30–0.46). The effect sizes ranged from −1.16 to 2.66, with 18 (62%) studies indicating that simulation-based instruction produced statistically significant learning outcome gains, 7 (24%) that showed negative effects, and 4 (14%) with no statistically significant effects. The hypothesis of homogeneity was rejected because a statistically significant Q value indicated presence of heterogeneity among the studies ($Q = 188.00$, $I^2 = 85.10\%$).

For studies of virtual worlds, a fixed effects meta-analysis resulted in a mean effect size of 0.36 ($SE = 0.03$; 95% confidence interval 0.29–0.44) for the relationship between virtual worlds-based instruction and learning outcome gains; the REM analysis resulted in a mean effect size of 0.41 ($SE = 0.09$; 95% confidence interval 0.23–0.59). The effect sizes ranged from −0.36 to 2.23 with 17 (68%) in the positive direction (i.e., virtual worlds-based instruction increased learning gains), 5 (20%) were negative, and 3 (12%) with no significant effects. The hypothesis of homogeneity was rejected because a statistically significant Q value indicated presence of heterogeneity among the studies

Table 1
Descriptive statistics.

Variables	Games		Simulation		Virtual worlds	
	N	# of studies	*N*	# of studies	*N*	# of studies
Grade level						
Elementary	–	–	283	4	514	4
Middle	239	2	511	4	827	6
High	518	5	318	3	287	4
Under-graduate	2324	6	1441	18	1170	11
Discipline						
Science	504	6	–	–	1755	13
Mathematics	260	2	342	2	97	1
Medicine	46	1	452	6	222	3
Others	2271	4	21	1	315	8
Publication type						
Articles	12	2165	2165	26	1743	21
Conference proceedings	–	264	264	1	975	3
Dissertation	1	124	124	2	80	1

($Q = 112.36$, $I^2 = 78.64\%$). We then conducted moderator analyses to answer our research questions. A mixed effects meta-analysis for all the subgroup main and pairwise effects was conducted.

Thus, the three meta-analyses reported in this study revealed statistically significant positive effects of games, simulations, and virtual worlds on student learning. Further, in each case, analyses of the heterogeneity of the studies indicated that the extent of the gains differed depending among the studies. Therefore, moderator analyses were conducted to examine the research questions about factors that might influence the effectiveness of virtual world instruction.

4.1. Research questions

4.1.1. Question 1: (a) Which kinds of learning outcomes are more suitable for desktop-based virtual reality instruction? (b) Are the effects of virtual reality-based instruction on learning outcome measures moderated by the type of learning task? (c) Did the testing conditions impact the learning outcome gains?

A mixed effects moderator analysis of games for learning outcome variable did not yield a statistically significant difference between the studies. We then conducted a mixed effects analysis of the interaction effects between the learning outcome measures and type of learning tasks. We found that the studies that used skill-based learning outcome measures to test learning gains for procedural tasks (REM = 0.00, FEM = 0.00, $N = 130$) differed significantly from the studies that used knowledge-based learning outcome measures to test gains in procedural tasks (REM = 0.62, FEM = 0.91, $N = 2417$) and studies that used knowledge-based learning outcome measures to test gains in declarative tasks (REM = 0.47, FEM = 0.47, $N = 346$). For simulation studies, a mixed effects meta-analysis of neither the learning outcome moderator variable nor the interaction between the learning outcome measures and type of learning tasks yielded a statistically significant Q statistics.

When we analyzed the studies of virtual worlds, a mixed effects analysis of differences in the effectiveness for different learning outcomes, we found a statistically significant difference between the studies that used knowledge-based measures (REM = 0.57, FEM = 0.43, $N = 1909$) versus those that used skill-based measures (REM = 0.002, FEM = 0.002, $N = 472$). We also found statistical significant differences between the studies when we conducted a mixed effects analysis of the studies based on the interaction effects between the learning outcome measures and type of learning tasks. Studies that used knowledge-based learning outcome measures to test students' learning gains for declarative tasks showed higher (REM = 0.68, FEM = 0.56, $N = 1006$) learning gains than studies that used knowledge-based learning outcome for procedural tasks (REM = 0.25, FEM = 0.25, $N = 903$). However, studies using knowledge-based learning outcome measures to test students learning gains for procedural tasks were more effective than the studies that used skill-based learning outcome measures to test students learning gains for procedural tasks (REM = 0.002, FEM = 0.002, $N = 88$).

A mixed effect meta-analysis of testing condition variable revealed that simulation studies differed significantly and student performed better when they were administered the learning outcome measures immediately (REM = 0.64, FEM = 0.66, $N = 1011$) compared to when the administration was delayed (REM = 0.07, FEM = 0.14, $N = 1161$). The studies that assessed students learning in a context different from the instruction did not differ either with those administered the measure immediately or at a later time. We could not assess the impact of repetitive measures, as there were no studies available for comparisons. For games, we could compare the effects of testing immediately versus delayed and we found no difference between the studies. We could not conduct the analysis of the studies for repetitive and transfer for games because there were not enough studies to conduct these comparisons. For virtual worlds, students learning gains was not different irrespective of whether the measure was administered immediately or if there was a delay.

4.1.2. Question 2: Are the learning outcome gains higher in desktop-based virtual reality instruction as compared to the other methods of instruction?

A mixed effect meta-analysis of control group treatment variable revealed that the simulation studies differed significantly based on the kind of control group treatment. The students in the studies that received a combination of treatment outperformed (REM = −0.59, FEM = −0.59, $N = 162$) those students who either received traditional (REM = 0.47, FEM = 0.40, $N = 1782$), 2-D images (REM = 0.25, FEM = 0.25, $N = 31$), or no treatment (REM = 0.72, FEM = 0.64, $N = 578$). There was statistically no significant difference between the studies in the categories of games or virtual worlds based on the control group treatment.

4.1.3. Question 3: Did the form of instruction impact the learning outcome gains in desktop virtual reality-based learning environment?

We conducted a mixed effect meta-analysis of the variable forms of instruction, which revealed that simulation studies differed significantly. Studies that used simulations to provide students, opportunity to practice a concept (REM = 0.59, FEM = 0.47, $N = 1803$) that they learnt via other instructional method were more effective than the studies that used a stand-alone format of instruction (REM = 0.09, FEM = 0.18, $N = 553$). There was statistically no significant difference between the studies based on the forms of instruction for virtual worlds. We could not conduct this analysis for games because there were not enough studies to conduct these comparisons.

4.1.4. Question 4: (a) Did the availability of teachers enhance the learning outcome gains in desktop-based virtual reality learning environment? (b) Did students' collaboration impact the learning outcome gains in desktop-based virtual reality learning environment?

Teacher availability did not produce a statistically significant difference in student learning in either simulations or virtual worlds. We could not conduct an analysis of the effect of teacher access in games because of an insufficient number of studies.

A mixed effect analysis of the effect of student collaboration in games revealed that game-based instruction was more effective when students worked individually (REM = 0.72, FEM = 0.72, $N = 553$) than when students worked collaboratively (REM = −0.004, FEM = −0.004, $N = 553$). We did not find a statistically significant difference between the effects of cooperative and individual instructional modes in studies that examined student learning in simulations or virtual worlds.

4.1.5. Question 5: Did the novelty effect impact the learning outcome gains in desktop virtual reality-based learning environment?

We conducted three mixed effect analyses to test the influence of the number of treatment sessions, duration of each session, and total amount spent in virtual learning environment. We found that games-based studies differed significantly from each other on the number of treatment session. There was a statistically significant inverse relationship between the learning outcome gains and number of treatment sessions ($\beta = -0.28$). The studies did not differ either on the basis of duration of each session or total amount of time spent in the learning environment. Simulation studies differed significantly depending on the number of treatment sessions, but the relationship was quite small ($\beta = 0.04$). The studies did not differ based on the duration of each session or the total amount of time spent in the virtual learning environment. For virtual worlds, there was no statistically significant difference between the studies for neither number of treatment sessions, duration of each session, or total amount spent in virtual learning environment.

4.1.6. Question 6: Which kind of feedback is more suitable for a particular type of learning tasks in desktop-based virtual reality instruction?

A mixed effect analysis of the interaction between feedback and type of learning tasks variable indicated that the simulation studies differed significantly. For declarative tasks, elaborate explanation (REM = 2.29, FEM = 2.29, $N = 181$) was more effective than visual clues (REM = 0.81, FEM = 0.81, $N = 299$). For procedural tasks, knowledge of correct response type of feedback (REM = 1.08, FEM = 1.08, $N = 68$) was more effective than visual clues (REM = −0.06, FEM = 0.15, $N = 464$). We could not conduct an analysis of the interaction effect between type of feedback and learning tasks for games and virtual worlds because of an insufficient number of studies.

4.1.7. Question 7: Did methodological rigor moderate the learning outcomes gains?

When we analyzed the studies for differences in effect sizes related to methodological rigor of the studies, we found that the learning gains for simulations were higher when researcher-developed instruments (REM = 0.59, FEM = 0.49, $N = 6191$) were used than when standardized instruments (REM = −0.01, FEM = 0.08, $N = 967$) were used. There was no difference between the studies in the categories of games and virtual worlds on the variable of researcher-developed versus standardized instruments. We did not find any statistically significant differences among studies for any of types of virtual reality environments related to design quality.

5. Discussion and conclusions

More and more resources in the form of time and money are being devoted to the designing and developing desktop-based virtual reality instruction for teaching K-12 and higher education curriculum. Deploying desktop-based virtual reality instruction in schools and colleges not only involves financial cost but also the efforts to train the teachers to use them effectively. Therefore, it is critical that instructional designers make careful decisions in the design and development of instructional materials utilizing desktop-based virtual reality technologies. Although previous meta-analyses shed some light on the ambiguity regarding the instructional effectiveness of desktop-based virtual reality instruction (Lee, 1999; Sitzmann, 2011; Vogel et al., 2006), our meta-analysis examined all three forms of desktop-based virtual reality technologies and also assessed the instructional effectiveness of several design features such as feedback, student collaboration, and teacher access that can guide K-12 and higher education teachers in designing instruction using desktop-based virtual reality technologies. Both REM and FEM analyses had similar pattern of results and we discuss the results below.

In general, game-based learning environments were more effective than virtual worlds or simulations, with overall effect sizes that were roughly twice as large. This is a key contribution in the field of using virtual reality technologies for instruction because there is limited evidence of their effectiveness. Although Sitzmann (2011) and Vogel et al. (2006) also found positive effects of using games, this meta-analysis added to the body research by comparing the effectiveness of games to that of virtual worlds and simulations. Our analyses also identified several factors that influence the effectiveness of virtual worlds in fostering learning for each of those formats that might guide future development and implementation of these technologies.

We found no differences between studies assessing students' achievement levels using knowledge-based, abilities-based, or skill-based measures for games or virtual worlds. This indicates that both games and virtual worlds are suitable of the three kinds of learning outcomes. The studies differed on the learning outcomes measures for simulation, where studies assessing students' knowledge level were found to be more effective than the studies assessing skill level. This is likely because it may be easier for students to recall factual information then to develop the skills that they were expected to acquire from simulation because skill acquisition is a gradual process and may require repeated practice.

On decomposing the effects of virtual worlds on learning outcome measures, the interaction between measures of learning outcomes and learning tasks for games further established that when skill-based measures were used to assess learning in procedural tasks, learning outcome gains were negligible. The underlying reason for this result could be that the time spent in doing simulations is less compared to games and virtual worlds. A game may have many levels as well as virtual worlds are open unstructured virtual spaces that affords the students greater flexibility of the amount of time to be spent in these environments. Therefore, in future studies instructional designer must consider time as an essential feature of the learning environment when designing instruction that requires skill acquisition.

Our study found promising results of virtual reality-based instruction with regard to the testing conditions. We found that the effectiveness of games was the same whether students were assessed immediately or after the passage of time. This indicates that students learning in games have retention level beyond short-term learning. We were unable to analyze whether learning in games is transferable because we did not have any studies to conduct this analysis. Therefore, it is recommended that instructional designer must consider assessment strategies while designing a game-based environments where students ability to apply a concept can be assessed in a context different than the one in which they were instructed.

We investigated whether the benefits of virtual reality instruction are maintained over time or transferred to other contexts. In these analyses, we found that the benefits of simulations were greater when students were tested immediately after the instruction than when the assessment was delayed. However, the facilitative effects of virtual worlds were not affected by retention interval. Little literature discusses the instructional effectiveness of virtual reality-based instruction in the context of retention and transfer of learning from virtual to the real environment (Bossard, Kermarrec, Buche, & Tisseau, 2008). To date, there is no systemically analyzed evidence of the instructional effectiveness virtual reality-based instruction at different levels of retention. Although Sitzmann (2011) did include retention as a category for coding her studies, we included a whole spectrum of testing conditions from immediate, delayed, repetitive, or transfer.

Our study also contributed in the area of collaborative learning environments and their effectiveness. We found that students performed better when they worked individually rather than collaboratively when learning through games. This is contrary to the results found by

Vogel et al. (2006), who found that there was statistically non-significant difference between the studies that used collaborative versus non-collaborative design for the learning environment. However, further research is warranted to examine the possible benefits of collaboration from the perspective of obtaining alternative perspectives, offering personal insights, and engaging in meaning making during collaborative learning process (Bonk & King, 1998; Wan & Johnson, 1994). Our meta-analysis results were consistent with Sitzmann's (2011) finding that students learned better when simulations were used in the form of practice sessions than when they were used in a stand-alone format. These findings can be useful to designers in designing virtual reality-based instruction in terms of how to use it in the context of other instructional modes.

Our study made a significant contribution by delineating the instructional effectiveness of different kinds of feedback on the type of learning tasks for simulation studies. None of the previous reviews discussed in this paper analyzed the effects of feedback on type of learning tasks in a virtual reality-based instruction; also, Sitzmann (2011) discussed this as a limitation of her review. According to Hattie and Timperley (2007), feedback has tremendous impact on learning gains, both positive and negative. Therefore, it is essential that teachers are made knowledgeable about the features and situations that make feedback effective. Our analysis found that when learning tasks is declarative in nature, elaborate explanation type of feedback is more effective. This is likely because students may need detailed instruction or information to complete a task, which is based on factual knowledge. This argument is supported by the results that when task is procedural in nature by merely providing knowledge of correct response is sufficient to further guide the learners on completing the instructional task. The underlying reason can be that since this is a procedural task therefore, learners can explore and figure out ways of accomplishing a procedural task. These results can provide useful guidelines to teachers in designing feedback strategies that will maximize learning gains in virtual reality-based environments. However, one of the limitations of this study is that we could not analyze the interaction effects of feedback and type of learning tasks for games and virtual worlds because lack of sufficient information. Therefore, to advance the research in the area, it is advisable to the researchers to include the information on the design on of feedback in the virtual reality-based instruction.

Our synthesis of the research design quality indicated that researcher-developed instruments yielded better learning outcomes in simulations. This result may be due to the fact that standardized instruments do not encompass the whole spectrum of constructs the researcher wants to assess. On the other hand, researcher-developed instruments have not undergone the processes designed to assess and improve reliability and validity. Therefore, it is critical that researchers, instructional designers, and teachers assess the reliability and validity of the measures they use. In our study, there was no difference between the studies that were coded as "high" "low" or "medium" on the variable of design quality. This differs from the typical finding that studies coded as high quality have the lowest effect size. Our results could be interpreted as a demonstration of the robustness of the benefits of desktop-based virtual reality instruction.

In our study, we also found evidence for novelty effect for game-based studies. The result indicated that if students spent more time playing games, the learning outcome gains starts to diminish. This result also coincides with the results from analyzing the forms of instruction, in which students performed better in practice mode than in stand-alone instruction. This might result from students spending more time in the virtual environment when it is used as stand-alone instruction than when they are using virtual environment for practice purpose only.

This meta-analysis found that overall, virtual-reality based instruction was quite effective. However, there were several limitations to these findings, many due to factors that are common to all meta-analyses. Some studies did not provide adequate statistical information to permit calculation of effect sizes. Moreover, many studies provided incomplete information on the variables coded for this study. Therefore, we could not analyze effects of some of the factors that might influence the effectiveness of virtual reality-based instruction, limiting the information available to guide their design. For example, we had only a single study in games and virtual worlds that provided information on the kind of feedback provided during the virtual reality-based instruction. Therefore, we could not compare the effects between different kinds of feedback. This could be an interesting area to explore for deeper insight into the design of virtual reality applications. According to Hattie and Timperley (2007), feedback has tremendous impact on learning gains, both positive and negative. Therefore, it is essential that teachers are made knowledgeable about the features and situations that make feedback effective. Hence, researchers need to be more explicit in designing and describing the feedback mechanism embedded in the design of the learning environment.

The literature presents numerous advantages of using virtual reality-based instruction for learning. The results of this meta-analysis are encouraging in that they provide evidence that virtual reality-based instruction is an effective means of enhance learning outcomes. Educational institutions planning to invest time and financial resources are likely to see the learning benefits in their students. This meta-analysis also sheds light on the effectiveness of several instructional design principles that improve the effectiveness of the learning environments. Future studies can be designed to test more design variables and interesting interaction effects of design features to further inform about the design of virtual learning environments.

References

Allen, C., Chen, Q., Willson, V., & Hughes, J. N. (2009). Quality of design moderates effects of grade retention on achievement: a meta-analytic, multi-level analysis. *Educational Evaluation and Policy Analysis, 31*, 480–499.

Andolsek, D. L. (1995). Virtual reality in education and training. *International Journal of Instructional Media, 22*(2), 145–155.

Ang, C. S., & Krishna, R. G. S. V. (2008). Computer game theories for designing motivating educational software: a survey study. *International Journal on E-Learning, 7*(2), 181–199.

Bonk, C. J., & King, K. S. (1998). *Electronic collaborators: Learner-centered technologies for literacy, apprenticeship, and discourse*. Mahwah, NJ: Lawrence Erlbaum.

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. West Sussex: John Wiley & Sons, Ltd.

Borenstein, M., Hedges, L. V., Higgins, J., & Rothstein, H. (2010). A basic introduction to fixed-effect and random-effects models for meta-analysis. *Research Synthesis Methods, 1*, 97–111.

Bossard, C., Kermarrec, G., Buche, C., & Tisseau, J. (2008). Transfer of learning in virtual environments: a new challenge? *Virtual Reality, 12*, 151–161.

Chen, C., & Teh, S. (2000). An affordable virtual reality technology for constructivist learning environments. In *The 4th global Chinese conference on computers in education Singapore* (pp. 414–421).

Chen, C., Toh, S., & Ismail, W. (2005). Are learning styles relevant to virtual reality? *Journal of Research on Technology in Education, 38*(2), 120–128.

Clark, R. (1989). Reconsidering research on learning from media. *Review of Educational Research, 53*(4), 445–459.

Coller, B. D., & Shernoff, D. J. (2009). Video game-based education in mechanical engineering: a look at student engagement. *International of Engineering Education, 25*, 308–317.

Costello, P. (1993). *Health and safety issues associated with virtual reality – A review of current literature*. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.6.3025&rep.
Cruz-Neira, C., Sandin, D. J., & DeFanti, T. A. (1993, August). Surround-Screen projection-based virtual reality: the design and implementation of the CAVE. In *Paper presented at the annual meeting of SIGGRAPH* (pp. 135–142).
Csikszentmihalyi, M. (2000). *Beyond boredom and anxiety. Experiencing flow in work and play*. San Francisco: Jossey-Bass.
De Jong, T. (1991). Learning and instruction with computer simulations. *Education & Computing, 6*, 217–229.
Dickey, M. D. (2003). Teaching in 3D: pedagogical affordances and constraints of 3D virtual worlds for synchronous distance learning. *Distance Education, 24*(1), 105–121.
Dickey, M. (2005). Three-dimensional virtual worlds and distance learning: two case studies of Active Worlds as a medium for distance education. *British Journal of Educational Technology, 36*, 439–451.
Dondlinger, M. J. (2007). Educational video game design: a review of the literature. *Journal of Applied Educational Technology, 4*(1), 21–31.
Emerson, T., & Revere, D. (1997). *Virtual reality in training and education: Resource guide to citations and online information*. Retrieved from http://vr.coe.ecu.edu/otherpgs.htm.
Fallman, D. (n.d.). Virtual reality in education: On-line survey. Retrieved from http://www8.informatik.umu.se/~dfallman/projects/vrie/.
Galas, C., & Ketelhut, D. J. (2006). River City, the MUVE. *Learning and Leading with Technology, 33*(7), 31–32.
Gee, J. P. (2003). *What video games have to teach us about learning and literacy*. New York: Palgrave/Macmillan.
Gee, J. P. (2007). *Good video games + good learning*. New York, NY: Peter Lang.
Glass, G. V., McGaw, B., & Smith, M. L. (1981). *Meta-analysis in social research*. Beverley Hills, CA: Sage.
Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77*, 81–112.
Hawkins, D. G. (1995). Virtual reality and passive simulators: the future of fun. In F. Biocca, & M. R. Levy (Eds.), *Communication in the age of virtual reality* (pp. 159–189). Hillsdale, NJ: Lawrence Erlbaum.
Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. Orlando, FL: Academic Press.
Hedges, L. V., & Pigott, T. D. (2004). The power of statistical tests for moderator variables in meta-analysis. *Psychological Methods, 9*(4), 426–445.
Heiling, M. (1962). *Sensorama simulator*. Retrieved from http://www.mortonheilig.com/.
Hew, K. F., & Cheung, W. S. (2010). Use of three-dimensional (3-D) immersive virtual worlds in K-12 and higher education settings: a review of the research. *British Journal of Educational Technology, 41*, 33–55.
Higgins, J. P. T., Thompsin, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *British Medical Journal, 327*, 557–560.
Kamel Boulos, M. N., & Wheeler, S. (2007). The emerging Web 2.0 social software: an enabling suite of sociable technologies in health and healthcare education. *Health Information and Libraries Journal, 24*, 2–23.
Lee, J. (1999). Effectiveness of computer-based instructional simulation: a meta-analysis. *International Journal of Instructional Media, 26*(1), 71–85.
Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. Thousand Oaks, CA: Sage.
Mantovani, F., Gaggiolo, A., Castelnuovo, G., & Riva, G. (2003). Virtual reality training for health-care professionals. *CyberPsychology & Behavior, 6*(4), 389–395. http://dx.doi.org/10.1089/109493103322278772.
Merchant, Z., Goetz, E., Keeney-Kennicutt, W., Kwok, O., Cifuentes, L., & Davis, T. J. (2012). The learner characteristics, features of desktop 3d virtual reality environments, and college chemistry instruction: a structural equation modeling analysis. *Computers & Education, 59*, 551–568.
McLellan, H. (2004). Virtual realities. In D. H. Jonassen (Ed.), *Handbook of research for educational communications and technology* (pp. 461–497). Mahwah, NJ: Lawrence Erlbaum Associates.
McNamara, D. S., Jackson, G. T., & Graesser, A. (2009). Intelligent tutoring and games (ITAG). In *AIED 2009 14th* (pp. 44–65).
Pasqualotti, A., & Freitas, C. M. D. S. (2002). MAT3D: a virtual reality modeling language environment for the teaching and learning of mathematics. *CyberPsychology & Behavior, 5*(5), 409–422.
Riva, G. (2003). Applications of virtual environments in medicine. *Methods of Information in Medicine, 42*, 524–534.
Sitzmann, T. (2011). A meta-analytic examination of the instructional effectiveness of computer-based simulation games. *Personnel Psychology, 64*, 489–528.
Tobias, S., & Fletcher, J. D. (2010). Introduction. In S. Tobias, & J. D. Fletcher (Eds.), *Computer games and instruction* (pp. 3–16). Charlotte, NC: Information Age.
Vogel, J. J., Vogel, D. S., Cannon-Bower, J., Bowers, C. A., Muse, K., & Wright, M. (2006). Computer gaming and interactive simulations for learning: a meta-analysis. *Journal of Educational Computing Research, 34*(3), 229–243.
Wan, D., & Johnson, P. M. (1994). Experiences with CLARE: a computer-supported collaborative learning environment. *International Journal of Human-Computer Studies, 41*(3), 851–879.
Wong, B. L. W., Ng, B. P., & Clark, S. A. (2000). Assessing the effectiveness of animation and virtual reality in teaching operative dentistry. *Journal of Dentistry: Educational Technology Section*. Retrieved May 24, 2010, from http://www1.elsevier.com/homepages/sab/jdentet/contents/wong2/wong2.html.
Youngblut, C. (1998). *Educational uses of virtual reality technology*. Alexandria, VA: Institute for Defense Analyses (IDA Document D-2128).

Further reading[1]

*Ainge, D. J. (1996). Upper primary students constructing and exploring three dimensional shapes: a comparison of virtual reality with card nets. *Journal of Educational Computing Research, 14*(4), 345–369.
*Akpan, J. P., & Andre, T. (2000). Using a computer simulation before dissection to help students learn anatomy. *Journal of Computers in Mathematics and Science Teaching, 19*(3), 297–313.
*Akpan, J., & Strayer, J. (2010). Which comes first: the use of computer simulation of frog dissection or conventional dissection as academic exercise? *Journal of Computers in Mathematics and Science Teaching, 29*(2), 113–138.
*Anderson, M. B. (Ed.). (2000). *Research in medical education: Proceedings of the annual conference (39th, Chicago, Illinois, October 30–November 1, 2000)*.
*Annetta, L. A., Minogue, J., Holmes, S. Y., & Cheng, M. (2009). Investigating the impact of video games on high school students' engagement and learning about genetics. *Computers & Education, 53*(1), 74–85. http://dx.doi.org/10.1016/j.compedu.2008.12.020.
*Antonietti, A., & Cantoia, M. (2000). To see a painting versus to walk in a painting: an experiment on sense-making through virtual reality. *Computers & Education, 34*(3–4), 213–223. Retrieved from http://vnweb.hwwilsonweb.com.lib-ezproxy.tamu.edu:2048/hww/jumpstart.jhtml?recid=0bc05f7a67b1790e8b8fd978b1f353c32290247e61231ff46e65c033af0013b462043749adf530c4&fmt=C.
*Barnea, N., & Dori, Y. J. (1999). High-school chemistry students' performance and gender differences in a computerized molecular modeling learning environment. *Journal of Science Education and Technology, 8*(4), 257–271.
*Başer, M., & Durmuş, S. (2010). The effectiveness of computer supported versus real laboratory inquiry learning environments on the understanding of direct current electricity among pre-service elementary school teachers. *Eurasia Journal of Mathematics, Science and Technology Education, 6*(1), 47–61.
*Bayrak, C. (2008). Effects of computer simulations program on university students' achievement in physics. *Turkish Online Journal of Distance Education, 9*, 53–62.
*Blanton, W. E., Moorman, G. B., Hayes, B. A., & Warner, M. L. (1997). Effects of participation in the Fifth Dimension on far transfer. *Journal of Educational Computing Research, 16*, 371–396.
*Blunt, R. (2007). Does game-based learning work? Results from three recent studies. In *Interservice/industry training, simulation & education conference (I/ITSEC)*. Orlando, Florida, USA: NTSA.
*Bonnetain, E., Boucheix, J., Hamet, M., & Freysz, M. (2010). Benefits of computer screen-based simulation in learning cardiac arrest procedures. *Medical Education, 44*(7), 716–722.
*Brewster, J. (1996). Teaching abnormal psychology in a multimedia classroom. *Teaching of Psychology, 23*(4), 249–252.
*Brom, C., Preuss, M., & Klement, D. (2011). Are educational computer micro-games engaging and effective for knowledge acquisition at high schools? A quasi-experimental study. *Computers & Education, 57*(3), 1971–1988.
*Burgess, M. (2010). *Optimal experience and reading achievement in virtual environments among college level developmental readers* (Ed.D., Sam Houston State University). Available from ProQuest dissertations and theses database. (UMI No. 3470190).

[1] References marked with an asterisk indicate studies included in the meta-analysis.

*Cheng, Y., & Wang, S. (2011). Applying a 3D virtual learning environment to facilitate student's application ability – the case of marketing. *Computers in Human Behavior, 27*(1), 576–584.
*Choi, B. S., & Gennaro, E. (1987). The effectiveness of using computer simulated experiments in junior high students' understanding of the volume displacement concept. *Journal of Research in Science Teaching, 24*, 539–552.
*Chou, C.-H. (1998). *The effectiveness of using multimedia computer simulations coupled with social constructivist pedagogy*. Unpublished doctoral dissertation. New York: Columbia University.
*Codd, A. M., & Choudhury, B. (2011). Virtual reality anatomy: is it comparable with traditional methods in the teaching of human forearm musculoskeletal anatomy? *Anatomical Sciences Education, 4*(3), 119–125.
*Coller, B. D., & Scott, M. J. (2009). Effectiveness of using a video game to teach a course in mechanical engineering. *Computers & Education, 53*, 900–912.
*Copolo, C. F., & Hounshell, P. B. (1995). Using three-dimensional models to teach molecular structures in high school chemistry. *Journal of Science Education and Technology, 4*(4), 295–305.
*Engum, S. A., Jeffries, P., & Fisher, L. (2003). Intravenous catheter training system: computer-Based education versus traditional learning methods. *American Journal of Surgery, 186*, 67–74.
*Farrokhnia, M. R. (2010). A study on the impact of real, virtual and comprehensive experimenting on students' conceptual understanding of DC electric circuits and their skills in undergraduate electricity laboratory. *Procedia: Social Behavioral Sciences, 2*(2), 5474–5482.
*Farynaiarz, J. V., & Lockwood, L. G. (1992). Effectiveness of microcomputer simulations in stimulating environmental problem solving by community college students. *Journal of Research in Science Teaching, 29*(5), 453–470.
*Finkelstein, N. D., Adams, W. K., Keller, C. J., Kohl, P. B., Perkins, K. K., Podolefsky, N. S., et al. (2005). When learning about the real world is better done virtually: a study of substituting computer simulations for laboratory equipment. *Physical Review Special Topics – Physics Education Research, 1*(1), 010103-1–010103-8.
*Foreman, N., Boyd-Davis, S., Moar, M., Korallo, L., & Chappell, E. (2008). Can virtual environments enhance the learning of historical chronology? *Instructional Science: An International Journal of the Learning Sciences, 36*(2), 155–173.
*Frear, V., & Hirschbuhl, J. J. (1999). Does interactive multimedia promote achievement and higher level thinking skills for today's science students? *British Journal of Educational Technology, 30*, 323–329.
*Gibbons, N. J., Evans, C., Payne, A., Shah, K., & Griffin, D. K. (2004). Computer simulations improve university instructional laboratories. *Cell Biology Education, 3*(4), 263–269.
*Gordoni, J. A., Shaffer, D. W., Raemer, D. B., Pawlowski, J., Hurford, W., & Cooper, J. (2006). A randomized controlled trial of simulation-based teaching versus traditional instruction in medicine: a pilot study among clinical medical students. *Advances in Health Sciences Education, 11*, 33–39. http://dx.doi.org/10.1007/s10459-004-7346-7.
*Hauptman, H. (2010). Enhancement of spatial thinking with virtual spaces 1.0. *Computers & Education, 54*(1), 123–135.
*Hu, J., Yu, H., Shao, J., Li, Z., Wang, J., & Wang, Y. (2009). Effects of dental 3D multimedia system on the performance of junior dental students in preclinical practice: a report from china. *Advances in Health Sciences Education, 14*(1), 123–133.
*Huppert, J., Lomask, S. M., & Lazarowitz, R. (2002). Computer simulations in the high school: students' cognitive stages, science process skills and academic achievement in microbiology. *International Journal of Science Education, 24*(8), 803–821.
*Jaakkola, T., & Nurmi, S. (2008). Fostering elementary school students' understanding of simple electricity by combining simulation and laboratory activities. *Journal of Computer Assisted Learning, 24*(4), 271–283.
*Jaakkola, T., Nurmi, S., & Veermans, K. (2011). A comparison of students' conceptual understanding of electric circuits in simulation only and simulation-laboratory contexts. *Journal of Research in Science Teaching, 48*(1), 71–93.
*Kaser, K. C. (1996). *The effects of virtual reality on learning office layout design*. Available from ProQuest dissertations and theses database. (UMI No. 9703783).
*Kebritchi, M., Hirumi, A., & Bai, H. (2010). The effects of modern mathematics computer games on mathematics achievement and class motivation. *Computers & Education, 55*(2), 427–443.
*Ketelhut, D. J., & Nelson, B. C. (2010). Designing for real-world scientific inquiry in virtual environments. *Educational Research Special Issue: Virtual Worlds and Education, 52*(2), 151–167.
*Keyser, D. (2010). A comparative analysis of student learning with a collaborative computer simulation of the cardiopulmonary system. In *Paper presented at the annual meeting of the American Educational Research Association, Denver, CO*. Retrieved from ERIC database. (ED509520).
*Kim, J. H., Kim, W. O., Min, K. T., Yang, J. Y., & Nam, Y. T. (2002). Learning by computer simulation does not lead to better test performance than textbook study in the diagnosis and treatment of dysrhythmias. *Journal of Clinical Anesthesia, 14*, 395–400.
*Kinzie, M. B., Strauss, R., & Foss, M. J. (1993). The effects of an interactive dissection simulation on the performance and achievement of high school biology students. *Journal of Research in Science Teaching, 30*(8), 989–1000.
*Kwon, O. N., Kim, S. H., & Kim, Y. (2002). Enhancing spatial visualization through virtual reality (VR) on the web: software design and impact analysis. *Journal of Computers in Mathematics and Science Teaching, 21*(1), 17–31.
*Lee, E. A., Wong, K. W., & Fung, C. C. (2010). How does desktop virtual reality enhance learning outcomes? A structural equation modeling approach. *Computers & Education, 55*, 1424–1442.
*Michael, K. Y. (2001). The effect of a computer simulation activity versus a hands-on activity on product creativity in technology education. *Journal of Technology Education, 13*(1), 31–43.
*Mills, S., & de Araújo, M. M. T. (1999). Learning through virtual reality: a preliminary investigation. *Interacting with Computers, 11*(4), 453–462.
*Moreno-Ger, P., Torrente, J., Bustamante, J., Fernández-Galaz, C., Fernández-Manjón, B., & Comas-Rengifo, M. (2010). Application of a low-cost web-based simulation to improve students' practical skills in medical education. *International Journal of Medical Informatics, 79*(6), 459–467.
*Nicholson, D. T., Chalk, C., Funnell, W. R. J., & Daniel, S. J. (2006). Can virtual reality improve anatomy education? A randomised controlled study of a computer-generated three-dimensional anatomical ear model. *Medical Education, 40*(11), 1081–1087.
*North, M. M., Sessum, J., & Zakalev, A. (2004). Immersive visualization tool for pedagogical practices of computer science concepts: a pilot study. *Journal of Computing Sciences in Colleges, 19*, 207–215.
*Özmen, H. (2008). The influence of computer-assisted instruction on students' conceptual understanding of chemical bonding and attitude toward chemistry: a case for turkey. *Computers & Education, 51*(1), 423–438.
*Papaevripidou, M., Constantinou, C. P., & Zacharia, Z. C. (2007). Modeling complex marine ecosystems: an investigation of two teaching approaches with fifth graders. *Journal of Computer Assisted Learning, 23*(2), 145–157.
*Petersson, H., Sinkvist, D., Wang, C., & Smedby, O. (2009). Web-based interactive 3D visualization as a tool for improved anatomy learning. *Anatomical Sciences Education, 2*(2), 61–68.
*Rafi, A., & Samsudin, K. (2009). Practising mental rotation using interactive desktop mental rotation trainer (iDeMRT). *British Journal of Educational Technology, 40*(5), 889–900.
*Ronen, M., & Eliahu, M. (2000). Simulation—a bridge between theory and reality: the case of electric circuits. *Journal of Computer Assisted Learning, 16*(1), 14–26.
*Şengel, E. (2010). The effects of computer simulated experiments on high school students' understanding of the displacement and velocity concepts. *EğitimAraştırmalarıDergisi, 39*, 191–211.
*Serrano, E. L., & Anderson, J. E. (2004). The evaluation of food pyramid games, a bilingual computer nutrition education program for Latino youth. *Journal of Family and Consumer Sciences Education, 22*(1), 1–16.
*Shim, K., Park, J., Kim, H., Kim, J., Park, Y., & Ryu, H. (2003). Application of virtual reality technology in biology education. *Journal of Biological Education, 37*(2), 71–74.
*Song, K. S., & Lee, W. Y. (2002). A virtual reality application for geometry classes. *Journal of Computer Assisted Learning, 18*(2), 149–156.
*Stern, L., Barnea, N., & Shauli, S. (2008). The effect of a computerized simulation on middle school students' understanding of the kinetic molecular theory. *Journal of Science Education and Technology, 17*(4), 305–315.
*Sun, K., Chan, H., & Meng, K. (2010, Nov–Dec). Research on the application of virtual reality on arts core curricula. In *The proceedings of the 5th international conference of computer sciences and convergence information technology (ICCIT)* (pp. 234–239).
*Sun, K., Lin, Y., & Yu, C. (2008). A study on learning effect among different learning styles in a web-based lab of science for elementary school students. *Computers & Education, 50*(4), 1411–1422.
*Urhahne, D., Nick, S., & Schanze, S. (2009). The effect of three-dimensional simulations on the understanding of chemical structures and their properties. *Research in Science Education, 39*(4), 495–513.
*Virvou, M., Katsionis, G., & Manos, K. (2005). Combining software games with education: evaluation of its educational effectiveness. *Educational Technology & Society, 8*(2), 54–65.

*Yang, J. C., Chen, C. H., & Jeng, M. C. (2010). Integrating video-capture virtual reality technology into a physically interactive learning environment for English learning. *Computers & Education, 55*(3), 1346–1356.
*Zacharia, Z. C. (2007). Comparing and combining real and virtual experimentation: an effort to enhance students' conceptual understanding of electric circuits. *Journal of Computer Assisted Learning, 23*(2), 120–132.