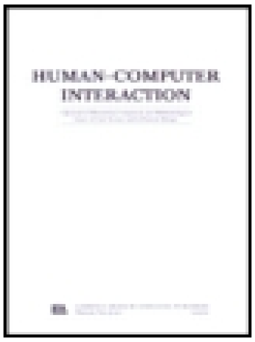


Human–Computer Interaction

ISSN: 0737-0024 (Print) 1532-7051 (Online) Journal homepage: http://www.tandfonline.com/loi/hhci20

Virtual Reality Interventions for Personal Development: A Meta-Analysis of Hardware and Software

Matt C. Howard

HUMAN–COMPUTER INTERACTION, 2018, Volume 00, pp. 1–35

ISSN: 0737-0024 print / 1532-7051 online
DOI: https://doi.org/10.1080/07370024.2018.1469408

Taylor & Francis
Taylor & Francis Group

Check for updates

Virtual Reality Interventions for Personal Development: A Meta-Analysis of Hardware and Software

Matt C. Howard
The University of South Alabama

Virtual reality (VR) has been repeatedly applied for personal development purposes, ranging from learning and training (cognitive outcomes), to psychological therapies (emotional outcomes), to physical rehabilitation (physical outcomes). Several factors lead to a successful VR intervention, most notably the hardware and software. In the current article, a meta-analysis is performed to test the effect of specialized input hardware (e.g motion sensors, floor pads, etc.), advanced output hardware (i.e., head-mounted displays, surround-screen displays, etc.), and game elements (i.e., score, competition, etc.) across and within the three noted applications of VR intervention. When analyzing the overall effects, only game elements had a significant impact on outcomes. When analyzing specific applications, input hardware did not have a notable impact on outcomes for any application; output hardware had a notable impact on cognitive and emotional outcomes but not physical; and game elements had a notable impact on cognitive outcomes but not emotional or physical. From these results, the current article provides direct suggestions for future research and practice. Particularly, certain mediating mechanisms are suggested to explain the impact of output hardware and game elements on VR intervention outcomes, sparking possible new directions for research and practice.

In recent years, the popularity of virtual reality (VR) has greatly expanded, and the technology has been implemented to improve training, therapy, rehabilitation, and a host of other applications focused on personal development (Burdea & Coiffet, 2003; Mirelman et al., 2010; Seth, Vance, & Oliver, 2011). In workplace training applications alone, the computer-administered simulation market, which includes VR, generated $2.4 billion in revenue in 2012, and it was expected to grow to $6.7 billion

Matt C. Howard (mhoward@southalabama.edu, MattCHoward.com) is a researcher with an interest in statistics, workplace technologies, personal development, and personality; he is an Assistant Professor in the Mitchell College of Business at the University of South Alabama.

in 2017 (Adkins, 2013). In regard to therapy, many authors have suggested that VR therapy programs are the *most* ideal method to address certain emotional disorders, particularly phobias (Bohil et al., 2011; Rothbaum et al., 2014). Lastly, VR rehabilitation programs are often more effective than comparable methods for developing physical abilities (Brütsch et al., 2011; Yen et al., 2011). While applications of VR for personal development have been largely successful, much remains unknown about VR interventions.[1]

All VR interventions include various hardware and software. As many have suggested, the applied hardware and software may have the largest impact of any factor in predicting VR intervention success (Rahman, 2010; Rizzo et al., 2013; Rostami et al., 2012). Unfortunately, analyses of particular hardware and software are difficult to conduct. Most often in research, a VR intervention is contrasted against a more typical intervention in a randomized control trial. While the results provide valuable information about the entire VR intervention, it is almost impossible to discern the effects of particular hardware or software through this method. To do so using an experimental design, researchers would need to compare multiple VR interventions, which is beyond the resources of most. Therefore, the answer to a central question for the study of VR interventions is still unknown—What is the effect of commonly applied hardware and software on VR intervention outcomes? Due to the importance of this research question, the current article seeks answers to guide future research and practice.

VR interventions consist of two types of hardware—inputs and outputs (Burdea & Coiffet, 2003; Jayaram, Connacher, & Lyons, 1997). Input hardware receives user commands. Currently, the standard input hardware is the keyboard, mouse, and joystick, whereas specialized input hardware are sensor gloves, motion sensors, and other technologies (O'Flynn et al., 2013; Regterschot et al., 2014). Alternatively, output hardware presents relevant visual information. Currently, the standard output hardware is the computer monitor, whereas more advanced output hardware are head-mounted displays (HMDs), surround-screen displays, three-dimensional monitors, and other technologies (Hou, Nam, Peng, & Lee, 2012; Kihara et al., 2012). Both specialized input and output hardware have been suggested to improve the physical fidelity and immersive properties of a VR intervention, which could result in better outcomes (Burdea & Coiffet, 2003; Hou et al., 2012; Jayaram et al., 1997; Kihara et al., 2012; O'Flynn et al., 2013). The current article tests whether these specialized input and advanced output hardware provide better VR intervention outcomes than standard alternatives. In doing so, the results may indicate whether researchers and developers are allocating costly resources toward creating technologies with little benefits, and future research can identify the exact mechanisms that cause these outcomes.

[1] The current study is multidisciplinary in nature and relevant to all applications of VR technology for personal development. For this reason, the current study uses the term "intervention" to refer to any effort to alter individuals' characteristics, whether it is a training to develop knowledge and skills, a therapy to alter emotions, a rehabilitation program to improve physical abilities, or any other application of VR for personal development.

Further, serious games are likely the most discussed type of VR intervention software, which are games with a primary purpose other than entertainment—most often learning. Authors have proposed that certain game elements, such as challenge and score, improve user enjoyment and motivation, which may result in better intervention outcomes (Bedwell, Pavlas, Heyne, Lazzara, & Salas, 2012; Deterding, Sicart, Nacke, O'Hara, & Dixon, 2011; Sitzmann, 2011). Likewise, these game elements have also been suggested to improve outcomes via heightened cognitive fidelity, thereby signifying that game elements may improve VR intervention outcomes via a variety of mechanisms (Kapp, 2012; Zichermann & Cunningham, 2011). The current article tests whether serious games actually provide better outcomes than typical VR interventions. Once again, the results of this analysis may determine whether resources should be allocated to other aspects of VR interventions, and future research can further analyze whether the suggested mechanisms cause these effects.

Additionally, when testing the impact of these hardware and software on VR intervention outcomes, the current article analyzes their effectiveness across all personal development applications, as well as within three specific applications: learning and training, psychological therapy, and physical rehabilitation. Through performing general and focused analyses, the overall impact of these technologies, as well as their effectiveness for developing certain outcomes (cognitive, emotional, and physical), can be determined.

Together, the current article analyses the impact of certain hardware (input and output) and software (serious games) on VR intervention effectiveness, and the results provide inferences across and within multiple types of outcomes. To perform these analyses, meta-analytic methods are applied. Meta-analyses statistically combine the results of prior studies to obtain estimates of overall effects. Through systematically analyzing the effects of particular technologies across studies, a true understanding of these technologies can be obtained— an almost impossible feat with experimental designs alone. Thus, these results may provide great insights into the effects of certain hardware and software on VR intervention effectiveness.

To achieve the objectives of the current article, the following is organized as follows: First, a review is provided of the hardware and software of interest, in which specific hypotheses and research questions are presented. Second, a meta-analysis is conducted to determine the effects of these specific hardware and software, and a meta-regression analyses each of these effects simultaneously. Third, the implications of the results are discussed, and the findings are integrated into current literature on VR interventions.

1. BACKGROUND

VR is the computer-simulation of an environment that imitates a physical presence in real or imagined worlds, and users navigate these worlds through a realistic or unrealistic digital representation of themselves called an avatar (Mirelman et al., 2010; Seth et al., 2011). Across the many applications of VR, authors regularly propose that the purpose of VR is to cognitively transport the user to a digital environment (Burdea & Coiffet, 2003; Cairns, Cox, Day, Martin, & Perryman, 2013; Mirelman et al., 2010; Schmierbach, Limperos, & Woolley, 2012). Whereas users only interact with other types of technology, they may be able to "experience" VR. For this reason, the concepts of fidelity, immersion, and presence become pivotal when discussing VR (Witmer & Singer, 1998).

Fidelity refers to the similarity of the practice environment and/or tasks to the transfer environment and/or tasks (Cairns et al., 2013; Hou et al., 2012; Slater, 2003). A VR intervention with realistic environments and tasks would be considered to have high fidelity, whereas a VR intervention with basic graphics and dissimilar tasks has low fidelity. Further, fidelity can be further differentiated as physical and cognitive (Cairns et al., 2013; Hou et al., 2012; Slater, 2003). Physical fidelity is the extent that the practice environment and/or tasks look like the transfer environment and/or tasks, whereas cognitive fidelity is the extent that the cognitive processes during practice are similar to the cognitive processes during transfer. A VR intervention to train military skills would likely have poor cognitive fidelity if it does not have consequences for failure, for example (Cairns et al., 2013; Hou et al., 2012; Slater, 2003).

Further, immersion is an objective qualification of the VR system, and it refers to "the extent to which the [VR systems] are capable of delivering an inclusive, extensive, surrounding, and vivid illusion of reality" (Slater & Wilbur, 1997, p. 3). Immersion is often synonymous with physical fidelity (Cairns et al., 2013; Hou et al., 2012; Slater, 2003). Alternatively, presence is "a psychological state or subjective perception in which … part or all of the individual's perception fails to accurately acknowledge the role of technology in the experience" (Lombard, 2000; as cited in Krijn et al., 2004; Schmierbach et al., 2012), and immersion (physical fidelity) and cognitive fidelity are often expected to contribute to feelings of presence (Cairns et al., 2013; Hou et al., 2012; Schmierbach et al., 2012; Slater & Wilbur, 1997). Most often, users of VR are cognizant that they are using the technology, but they may perceive objects, events, entities, environments, and other features as being independent of the technology.

As many authors have argued, it may be necessary for users to be immersed in their experiences and experience feelings of presence to receive the true benefits of VR (Bouchard et al., 2012; Bulu, 2012; Ling, Nefs, Brinkman, Qu, & Heynderickx, 2013; Wirth et al., 2007). When interacting with VR without experiencing presence, users may be detached from their experiences. When present, however, users may experience notable emotional reactions from their digital events, and users may also be more mentally engaged in any presented information (Bouchard et al., 2012; Bulu, 2012; Ling et al., 2013). These factors may subsequently prompt better emotional and cognitive outcomes from a VR intervention. Further, many of these notions have already been supported by studies across several applications of VR. For example, for a VR racing video game, Schmierbach and colleagues (2012) showed that immersion predicted perceptions of presence and enjoyment, suggesting that immersed users may be more emotionally engaged. For a VR educational program, McQuiggan and colleagues (2008) also showed that immersion predicted perceptions of presence and self-efficacy, suggesting that immersed users may be more cognitively engaged. Likewise, a vast amount of other studies have found similar results when analyzing the impact of physical and cognitive fidelity on personal outcomes (Bouchard et al., 2012; Cairns et al., 2013; Dow, Mehta, Harmon, MacIntyre, & Mateas, 2007; Fox, Bailenson, & Binney, 2009; Hou et al., 2012; Krijn et al., 2004; Leong, 2011; Monahan, McArdle, & Bertolotto, 2008; Salmon, 2009), and a recent meta-analysis showed that immersion indeed has a moderate effect on presence (Cummings & Bailenson, 2016). Together, these prior findings suggest that immersion and presence may impact the benefits obtained from a VR intervention.

For these reasons, developers have placed great efforts toward creating hardware and software that allows users to interact and observe with their digital environments in a realistic manner, with hopes of creating more immersive experiences that prompt feelings of presence (Diemer, Alpers, Peperkorn, Shiban, & Mühlberger, 2015; Price, Mehta, Tone, & Anderson, 2011; Sanchez-Vives & Slater, 2005). For example, the most realistic VR program would project an entirely lifelike digital environment surrounding users; they would only need to reach out their hand to interact with the environment, feeling haptic feedback as they do so, and these feelings of presence could be heightened by also engaging the other senses (i.e., smell and hearing; Burdea & Coiffet, 2003; Steuer, 1992). Users would feel present in their environments when using the described technologies, and they would also experience similar psychological processes as their transfer environments. From these feelings of presence and similar cognitive processes, users may also gain better outcomes from their VR interventions, whether cognitive, emotional, or physical. While current technologies fall short of this scenario, new technologies are being continuously developed to create more immersive experiences, and these technologies can be categorized as input or output hardware.

Below, hypotheses are made about the impact of modern hardware and software on VR intervention outcomes. It should be noted, however, that the current meta-analysis does not explicitly test whether fidelity, presence, or any other mechanism may be the cause of these outcomes. Instead, the current meta-analysis only analyzes the effect of VR interventions, as well as certain aspects of hardware and software design on user outcomes.

1.1. Hardware

VR Input Hardware

When undergoing a VR intervention, all user interactions must be performed through an applied technology, such as a keyboard and mouse. Technologies that receive user inputs are categorized as input hardware. Due to their integral function to user experiences, input hardware are pivotal immersion and presence (Burdea, Richard, & Coiffet, 1996; Jayaram et al., 1997; Krijn et al., 2004). If the user must interact with their digital environment in an unnatural and non-immersive manner, then they would likely be unable to perceive their experiences as realistic, and would almost assuredly have difficulties in feeling present. Fortunately, developers have created input hardware that allows users to perform realistic and immersive actions to interact with their digital environments (Greffou et al., 2012; Krijn et al., 2004; Sanchez-Vives & Slater, 2005).

Currently, the traditional input hardware are the keyboard, mouse, and joystick (Rahman, 2010; Rizzo et al., 2013; Rostami et al., 2011). These three technologies are relatively cheap, providing accessibility to most practitioners or researchers. Also, almost anyone that has used a computer is familiar with a keyboard and mouse, and most are, at least, familiar with the functions of a joystick. Users can use these technologies without any prior training. While these three input hardware are commonplace and easy to apply, they provide an unrealistic and non-immersive user experience (Lull & Bushman, 2016; Martel et al., 2015; Mikropoulos & Strouboulis, 2004). Very few real-life experiences involve rapidly pressing buttons to perform activities, and even fewer involve interactions similar to a mouse or joystick. Likewise, these technologies do little to invoke feelings of presence, and likely reduce the effectiveness of any VR intervention that applies these technologies (Burdea et al., 1996; Kihara et al., 2012; Mon-Williams, Warm, & Rushton, 1993; O'Flynn et al., 2013). Nevertheless, due to the widespread use of these three input hardware, they serve as a comparison for all other input hardware.

Alternatively, specialized input hardware exist in an array of forms. The most popular specialized input hardware are sensor gloves, motion sensors, and trackpads, but many other specialized input hardware exist that allow users to perform realistic movements to interact with their digital environments (Krijn et al., 2004; Roy, Soni, & Dubey, 2013; Sanchez-Vives & Slater, 2005). Authors have proposed that, because these technologies allow realistic and immersive interactions, specialized input hardware may prompt feelings of presence, thereby improving subsequent VR intervention outcomes. These technologies have even been proposed to improve learning from a VR training program, emotional skills from a VR therapy program, and physical abilities from a VR rehabilitation program (Chang, Chen, & Huang, 2011; Greffou et al., 2012; Ip et al., 2013). The current study tests these suggestions through comparing traditional input hardware (keyboard, mouse, and joystick) against specialized input hardware for improving outcomes from VR intervention programs, with the belief that the latter is more effective.

Hypothesis 1: VR interventions that use specialized input hardware result in better outcomes than VR interventions using traditional input hardware (keyboard, mouse, and/or joystick).

VR Output Technology

While many developers have created innovative input hardware, the creation of displays that provide immersive experiences may be an even more popular research topic (Hou et al., 2012; Kihara et al., 2012; Rushton, Mon-Williams, & Wann, 1994). Displays can also be called output hardware. All visual information in a VR intervention, such as the digital environment and the results of users' actions, are presented via output hardware. Due to the central role of output hardware in VR interventions, this technology is extremely important for fidelity, immersion, and presence. Thereby, this technology effects more distal and important outcomes, such as learning, emotional skills, and physical abilities. Like input hardware, users may be unable to develop feelings of presence if the output hardware presents visual information in an unrealistic and non-immersive manner—and vice versa. Modern output hardware can be categorized as traditional or advanced.

Currently, the traditional output hardware is the standard computer monitor (Bohil et al., 2011; Brütsch et al., 2011; Rothbaum et al., 2014). Compared to other output hardware, computer monitors are inexpensive. Also, anyone who has used a computer is familiar with a computer monitor, allowing the technology to be used without any prior training. While these two factors lead to their widespread application, computer monitors provide an unrealistic and non-immersive experience. Computer monitors present visual information on a two-dimensional plane, inherently preventing a completely realistic presentation of a VR environment. Due to this artificiality, computer monitors do little to invoke feelings of presence, possibly reducing VR intervention outcomes (Bouchard et al., 2012; Bulu, 2012; Burdea et al., 1996; Chang et al., 2011; Mon-Williams et al., 1993). Due to their widespread application, however, computer monitors serve as a frame of reference for new and innovative output technology.

Many types of advanced output hardware exist. The most common advanced output hardware are surround-screen displays, HMDs, and three-dimensional monitors (Howard, 2017a; Maiberg, 2014; Robinett & Rolland, 1992; Shibata, 2002). To create a surround-screen display, multiple two-dimensional screens are placed or projected around the user, and each screen presents visual information in sync with the others. Alternatively, HMDs are worn over the users' head, appearing similar to night-vision goggles, and completely cover users' field of vision. HMDs are programmed to track users' head movements, and the displayed visual information changes in alignment with these head movements. Lastly, a three-dimensional monitor creates an artificial depth perception, often in conjunction with specialized glasses. Of these, the most acclaimed is likely the HMD, especially after the purchase of Oculus, the creator of the Rift (a HMD), for two billion dollars by Facebook (Maiberg, 2014).

These specialized displays are able to present visual information in a more realistic manner, whether through providing a truly immersive experience (surround-screen display and HMD) or presenting environments as they more naturally appear (three-dimensional monitor). Authors regularly propose that these advanced displays can increase feelings of presence, which thereby causes users to become more emotionally and cognitively engaged in their VR experiences (Bouchard et al., 2012; Bulu, 2012; Mon-Williams et al., 1993; Ward, Azuma, Bennett, Gottschalk, & Fuchs, 1992). Further, authors have likewise suggested that advanced output hardware can improve participant enjoyment and motivation, which may influence developmental outcomes (Grossberg, 1999; Mayes & Calhoun, 2007). Lastly, HMDs have been suggested to narrow the attention of participants to the developmental activities, as these devices restrict the field-of-vision (Maiberg, 2014; Robinett & Rolland, 1992; Shibata, 2002). Being unable to view the "real-world," participants may be therefore forced to only engage with their digital environments.

With these assumptions, more advanced displays should result in better VR intervention outcomes. Although this sentiment is regularly endorsed, a wide-scale comparison between traditional and advanced output hardware has yet to be performed. Researchers and practitioners may be spending resources on hardware which provides few benefits. Therefore, a comparison between traditional, HMD, and other advanced output hardware is presented in the current article, with the belief that HMDs and advanced displays produce better outcomes than monitors.

Hypothesis 2: VR interventions that use HMDs result in better outcomes than VR interventions using computer monitors.

Hypothesis 3: VR interventions that use advanced display hardware, other than HMDs, result in better outcomes than VR interventions using computer monitors.

1.2. Software

In addition to hardware, software is also pivotal in predicting VR intervention success (Burdea & Coiffet, 2003; Mirelman et al., 2010; Seth et al., 2011). Developers often create new software to provide more realistic experiences through making programs with better graphical capabilities. Like hardware, researchers and practitioners propose that more realistic and immersive programs may prompt presence, thereby improving other outcomes. While the creation of more realistic software programs is extremely popular, authors have also focused on another important attribute of VR intervention software.

Possibly, the most popular new direction in software development for VR interventions is the serious game and gamification movement (Deterding et al., 2011; Kapp, 2012; Zichermann & Cunningham, 2011), which has spread to almost every type of VR application. Game elements have been incorporated into rehabilitation programs to improve physical capabilities, organizational training programs to improve employee performance, and even therapy programs to improve emotional capabilities (Burdea & Coiffet, 2003; Mirelman et al., 2010; Seth et al., 2011). This area of research poses that game elements can improve computer-based interventions, as users often find these game elements naturally exciting. For instance, many students find educational math programs boring, but they may be more motivated if these programs include interactive elements (i.e., story progression for correct answers, customizable characters) or rewards (i.e., unlockable features for correct answers, high-score list). Further, adding game elements can improve the cognitive fidelity of a VR intervention (Cairns et al., 2013; Hou et al., 2012; Slater, 2003). Certain game elements, such as score, add consequences to developmental activities, causing the user to possibly feel stress to perform well. If stress is involved in the transfer activity, then the added game element may indeed improve the cognitive fidelity of the VR intervention, which would assumingly improve VR intervention outcomes. Thus, game elements may improve outcomes via a variety of mechanisms that may be relevant to a wide array of contexts, which may partially explain their widespread application.

To understand the effects of serious games, certain authors have created game element taxonomies (Bedwell et al., 2012; King, Delfabbro, & Griffiths, 2010; Rego, Moreira, & Reis, 2010). These taxonomies separate and describe individual game elements that may be included in any serious game. Among these taxonomies, the most popular and empirically supported is Bedwell and colleagues' (2012) taxonomy. This taxonomy was created to identify the effect of game elements on learning, but the authors note that the elements may extend to other contexts, such as physical rehabilitation and therapy programs (Bedwell et al., 2012). Further, Bedwell and colleagues' (2012) taxonomy identifies 19 different game elements, with some of the most commonly occurring being challenge, conflict, and rules/goals. Using this taxonomy, researchers and practitioners can identify the game elements present within their VR interventions, and design VR interventions with certain combinations of elements. Also, this taxonomy can help identify whether a VR intervention may be considered a serious game—a distinction that is not always clear.

Due to the direct relationship with motivation and cognitive fidelity, authors have repeatedly proposed that serious games and serious game elements lead to improved VR intervention outcomes. While this notion has been tested in individual studies, studies on serious game elements often suffer the same concerns as research on hardware. That is, study results can provide important information about the overall VR intervention, but little can be said about the impact of game elements themselves. For this reason, the current study provides an overall investigation into the impact of game elements on VR intervention success. It is expected that VR interventions that include game elements are more effective than those that do not.

Hypothesis 4: VR interventions which include game elements result in better outcomes than VR interventions that do not use game elements.

Relatedly, many authors have noted that serious games are rapidly growing in popularity. So much, in fact, that some authors consider VR programs and serious games to be the same (Sitzmann, 2011). To test this notion, the current study analyzes the number of VR interventions which do and do not specifically mention the inclusion of popular game elements.

Research Question 1: Do Most VR Interventions Include Game Elements?

A final note should be made about the study of game elements within the current article. This sole aspect of software design was chosen due to its relative popularity in research, as well as practice across several domains (Burdea & Coiffet, 2003; Mirelman et al., 2010; Seth et al., 2011), and the existence of several typologies that can be used to identify whether a VR intervention included game elements (Bedwell et al., 2012; King et al., 2010; Rego et al., 2010). Other aspects of software design have also seen popularity in research and practice, but many do not currently have associated taxonomies and lack clear guidelines for categorization. For instance, many authors have shown interest in the effects of VR intervention realism (Bouchard et al., 2012; Kwon, Powell, & Chalmers, 2013). It would be difficult to accurately categorize the realism of a VR intervention presenting a real-world environment but poor graphics relative to a VR intervention presenting a foreign environment but exceptional graphics. For these reasons, we believe that game elements are prime for inclusion in the present meta-analysis, whereas other elements of software design are less ideal and therefore not included.

1.3. Application Differences

As mentioned, VR has been applied for an array of developmental purposes, ranging from learning and training, to psychological therapy, to physical rehabilitation (Burdea & Coiffet, 2003; Mirelman et al., 2010; Seth et al., 2011). When applied for learning and training, VR interventions are often used in the workplace or classroom with the intentions of improving declarative knowledge and/or skills, such as a surgery simulation to train surgeons (Huang, Rauch, & Liaw, 2010; Monahan et al., 2008). When applied for psychological therapy, VR interventions are often used in conjunction with a therapist to improve emotional states and/or conquer certain phobias, such as a military simulation to address Post Traumatic Stress Disorder (PTSD) (Anderson et al., 2013; Malbos, Rapee, & Kavakli, 2013). When applied for physical rehabilitation, VR interventions are often used with a physical therapist or trainer to develop muscle and/or regain physical abilities, such as a virtual running simulation to develop gait abilities (Howard, 2017b; Meldrum et al., 2012; Tao et al., 2013). Each of these applications is extremely important, and their differences likely cause attributes of the VR intervention to have differing effects based on the application.

Primarily, three mechanisms were proposed above: fidelity/presence, attention, and motivation. Prior research has argued that presence may improve all developmental outcomes (Cairns et al., 2013; Mirelman et al., 2010; Schmierbach et al., 2012), and a VR intervention characteristic may have an influence on presence if it positively effects all outcomes. Authors have often suggested that attention positively influences cognitive outcomes, and some have suggested the same for psychological outcomes (Maiberg, 2014; Robinett & Rolland, 1992; Shibata, 2002); however, attention has generally not been attributed to physical outcome success. Some authors have even suggested that physical outcomes may be best developed when participants are not mentally engaged in the task at hand (Brütsch et al., 2011; Yen et al., 2011). Lastly, participants often find the development of cognitive outcomes, especially for school or work purposes, relatively boring and unexciting (Grossberg, 1999; Mayes & Calhoun, 2007). On the other hand, participants are often intrinsically motivated to develop their psychological or physical outcomes (Brütsch et al., 2011; Yen et al., 2011). For this reason, improvements to motivation would be expected to improve cognitive outcomes, but have little effect on psychological or physical outcomes—as users are already motivated to achieve these outcomes.

The hypotheses presented in the current article are studied across all VR interventions, as well as within the three noted applications: learning and training, psychological therapy, and physical rehabilitation. The mechanisms that cause VR intervention features to influence outcomes are not directly studied in the current article; however, the pattern of results could provide indirect inferences regarding the relevant mechanisms. For this reason, results are analyzed together, as well as separated by outcome, but no *a priori* hypotheses are made.

Research Question 2: What is the differential impact of VR intervention hardware and software in the applications of learning and training, psychological therapy, and physical rehabilitation?

2. METHODS

2.1. Literature Search[2]

To perform a meta-analysis, relevant studies investigating the content of interest need to be discovered, and these studies can be considered as the "data points." Multiple strategies were used to discover all studies, both published and unpublished, that empirically analyzed VRT programs. Searches were conducted in December 2014 within the following scholarly databases: PsycInfo, EBSCOhost, and Dissertation Abstracts International. For these separate searches, relevant keywords were "virtual reality," "digital simulation," and "computer simulation" (quotations included) followed by the words training, intervention, therapy, enhance, promote, and support. Second, reference sections of review articles and meta-analyses were cross-referenced. Third, e-mails were sent to authors in VR intervention research inquiring

[2] The current literature search coincided with Howard (2017b). The reported meta-analytic effect sizes do not overlap with those reported in Howard (2017b).

about unpublished data. Through these efforts, a large set of initial studies was identified.

2.2. Inclusion Criteria

Initially, 4,488 articles were recorded, but they had to meet the several criteria to be included within the meta-analysis. For each criterion, listed below, trained coders independently coded the same set of articles, and interrater agreement statistics were calculated. If the ICC(2, k) value was above .8, then the coders continued coding the articles independently. If the ICC(2, k) value was below .8, the coders discussed their coding decisions until a consensus was reached, and the process was repeated until an ICC(2, k) value of .8 was reached.

First, the article must have included an empirical study. This resulted in 2,687 articles from the initial search. Second, article must include (1) quantitative statistics, (2) a total sample size larger than nine, and (3) human participants. After removing these articles, 1,050 remained. Articles were further screened for whether they analyzed the effectiveness of a VR intervention to alter participant characteristics (i.e., knowledge, skills and abilities, emotions, etc.). After removing these articles, 662 remained. Next, articles were screened for whether they included an outcome measure administered after the VR intervention and whether the article included a control group. Once removing these articles, 264 remained. Lastly, many of these articles did not report sufficient statistical information to be included within analyses. If an article did not report sufficient statistical information, such as including effect sizes or information to calculate effect sizes, the corresponding author was contacted and two follow-up e-mails were sent. If the author never replied, any outcome which could be included was included, but many articles were forced to be excluded. Therefore, 184 articles, 7 dissertations, and 1 book chapter were included within the current analyses.

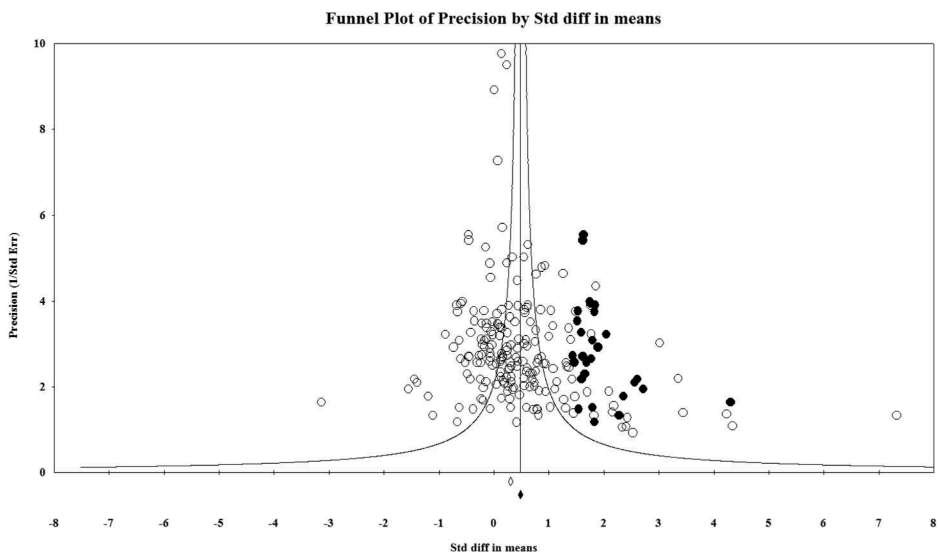
Further, justifications for certain inclusion criteria should be noted. Prior authors have suggested that empirical tests of interventions with sample sizes smaller than 10 are often pilot studies. These studies are not meant to test the impact of an intervention, but rather the possibility of applying a subsequent, more effective intervention. While studies could have been coded for being a pilot, using a sample size cutoff provides a more objective measure, and it also follows the suggestions and practices of prior authors (Hedges & Olkin, 2014; Schmidt & Hunter, 2014). Also, only studies with a control groups were included due to their benefits to internal validity. Intervention studies without a control group cannot reliably ascertain that any observed effects are due to the intervention, rather than the general malleability of the outcome and/or placebo effects (Keppel, 1991; Kirk, 1982; Winer, Brown, & Michels, 1971). Through only including controlled studies, the current meta-analysis can establish that observed effects are likely due to the true impact of the VR intervention. Similarly, analyses using studies with alternative treatment groups can identify the effect of VR interventions beyond traditional interventions—an important research question for the current study. Thus, the results include between-subject designs, as well as between-subject designs that made within-subject comparisons. No studies were included that only made within-subject comparisons. For this reason, we did not make any corrections to account for variance across studies due to individual differences as a random effect, as all studies included some type of between-subject comparison.

2.3. Coding

After the selection process was completed, two trained coders coded the articles on the attributes detailed in Figure 1 (Output Hardware, Input Hardware, Game Elements, Application, Type of Control Group, and Year). For every article, the two coders independently coded all attributes and discussed their coding decisions, even after their ICC(2,2) value for coding agreement exceeded .80. Each disagreement was resolved as a team.

Also, two coded aspects were not relevant to the proposed hypotheses. The first was the year of publication. The second was the type of control group (no-treatment or alternative treatment group). These two aspects could alter the results of a VR intervention, and should be controlled in analyses testing the effects of interest. For this reason, the publication year and the choice of control group were recorded and controlled in the meta-regressions detailed below.

FIGURE 1. Visual Representation of Studies Included in the Overall Meta-Analysis (Funnel Plot).



Note: White circles are observed study effect sizes. Black circles are implied missing study effect sizes.

2.4. Statistical Procedures

All analyses were calculated with the Comprehensive Meta-Analysis (Borenstein, Higgins, & Rothstein, 2005) software using a random effects model, guided by the suggestions of prior authors (Aguinis, Pierce, Bosco, Dalton, & Dalton, 2011; Aguinis, Sturman, & Pierce, 2008; Hedges & Olkin, 2014; Lipsey & Wilson, 2001; Schmidt & Hunter, 2014). Multiple effect sizes for the same category of outcome within a single study were averaged together. For example, if a study tested a VR intervention to develop cognitive outcomes and three cognitive outcome measures were used to gauge its effectiveness, these three effect sizes would be averaged together in analyses.

No corrections were made when calculating the results and this decision was made for four reasons. First, the outcomes represented an extremely wide array of constructs, and applying corrections to those without reported reliabilities through assumptions based on observed reliabilities for entirely different constructs (i.e., artifact distribution method) would bias results. Second, recent authors have placed great concern toward correcting for unreliability (LeBreton, Scherer, & James, 2014). These authors argue that correcting for unreliability may grossly overinflates relationships under certain conditions, resulting in an inaccurate view of constructs. This view is shared in the current study. Third, authors have suggested that uncorrected meta-analysis results should be interpreted when the outcome serves a practical or operational purpose, such as in selection contexts (Griffeth, Hom, & Gaertner, 2000; Harrison, Newman, & Roth, 2006; Van Iddekinge, Roth, Raymark, & Odle-Dusseau, 2012; Zimmerman, 2008). In the articles used, authors regularly applied the results from a comparison of two intervention programs to determine the superiority of an intervention beyond another. Given that these results were used for practical purposes, it is not appropriate to correct for unreliability. Fourth, none of the samples posed concerns toward range restriction, circumventing the need to correct for range restriction.

3. RESULTS

Several methods were used to observe publication bias (Table 2; Hedges & Olkin, 2014; Lipsey & Wilson, 2001; Schmidt & Hunter, 2014). First, the funnel plot indicated that many studies fell outside the expected range of standard error and effect size; however, the outliers fall on both sides of the mean effect, indicating that outliers may not be a concern. Second, the fail-safe *n* was 10,747 when analyzing all effects together, indicating that the effect would require an extreme amount of missing studies to alter the overall interpretation of results. The fail-safe *n* for the individual outcomes (cognitive, emotional, and physical) ranged from 200 to 3688. Third, Egger's test was significant when analyzing all outcomes together and cognitive outcomes, indicating that some biases may be present in the analyses for these two outcomes. Fourth, when using the trim-and-fill method with a random effects

TABLE 1. Variables coded within the current study.

Variable	Definition
Output hardware	The type of display could be categorized into one-of-three categories. The first was a standard computer monitor. The second was a head-mounted display. The third was any other display hardware, such as surround-screen displays or three-dimensional displays. No study used multiple types of display hardware.
Input hardware	The type of input could be categorized into one-of-two categories. The first was standard, which meant that the VR intervention utilized only a keyboard, mouse, and/or joystick. The second was specialized, which included any other input hardware (e.g., motion sensor, floor pads, etc.). It should be noted, however, that a HMD itself was not considered a type of input hardware. While HMDs indeed record user input, they often do so for the purposes of accurate output (e.g., aligning visual presentation with user point-of-view) rather than manipulating the digital environment.
Game elements	The current study coded whether the VR program included common game elements, such as score or adventure, when describing the details of the VR program.
Application	Each article was coded by the application of the VR intervention. The vast majority of studies were coded into the three categories below; however, if a study could not be coded into these categories, it was coded as "other" and only included in the overall analyses.
Cognitive	A study was coded as, "cognitive outcome development," if it taught new information or trained new skills. In general, the studies that were coded as cognitive outcomes involved the development of declarative knowledge, wayfinding, mental models, skills, spatial reasoning, and/or cognitive abilities.
Emotional	A study was coded as, "emotional outcome development," if it altered emotional and/or therapeutic outcomes. In general, the studies that were coded as emotional outcomes involved the altering of mood, emotions, craving, self-perceptions, attitudes, motivation, phobias, and/or post-traumatic stress disorder.
Physical	A study was coded as "physical outcome development, if it improved physical abilities. Most often, studies coded as physical outcomes were assessments of physical rehabilitation programs. In general, the studies that were coded as physical outcomes involved the improvement of motor control, balance, gait, and/or strength.
Type of control group	The control group for each article was coded into one-of-two categories: whether it was a no-treatment (i.e., wait-list design) or alternative treatment control group. Alternative treatment control groups included those that received any type of intervention that did not include VR. For instance, some physical rehabilitation studies compare a group that does physical exercise (control group) against a group that does the same physical exercise while wearing a HMD (treatment group). In this case, we considered the control group as an alternative treatment group.
Year	The year that the article was published was coded.

model, the results indicated that a large number of studies may be missing when analyzing all outcomes together (29; Table 1) and cognitive outcomes (14). When investigating the implied missing effect sizes, all were to the right of the mean, suggesting that any observed effects for these outcomes may actually be under-estimated. Thus, in agreement with previous meta-analyses, only effects that were calculated with more than three studies were considered reliable results.

TABLE 2. Publication bias analyses results and I^2.

	Number of studies	Total sample size	I^2	Q	Fail-Safe n	Egger's test β_0	Egger's test t	Trim-and-fill Left of mean	Trim-and-fill Right of mean
All outcomes	192	8508	79.880	949.322	10747	1.096	2.922**	0	29
Cognitive outcomes	84	4041	86.095	596.911	3688	1.538	2.720**	0	14
Emotional outcomes	80	3288	71.725	279.400	688	.528	.833	1	0
Physical outcomes	27	1002	55.035	57.823	200	.297	.337	4	0

*$p < .05$**$p < .01$

In regard to the hypotheses, the current study proposed several effects related to the software, hardware, and participant population. Effect sizes were calculated for each category of the variables, and the confidence intervals were compared to determine group differences. These results are presented in Table 3. Also, meta-regressions were performed to further test for the significance of each proposed effect simultaneously, which allows for analyses of each variable while controlling for the others. For the meta-regressions, each variable was dummy-coded. Table 4 presents a meta-regression with all studies. Table 5 presents a meta-regression of studies investigating cognitive outcome development (learning and training). Table 6 presents a meta-regression of studies investigating emotional outcome development (psychological therapy). Table 7 presents a meta-regression of studies investigating physical outcome development (physical rehabilitation).

TABLE 3. Effect of moderator variables on VR effectiveness (random effects).

Variable	Effect size (confidence interval) Q-value		
	Standard	Specialized	
1.) Input hardware	.345 (.210–.480) 705.532	.474 (.309–.640) 236.591	
	Monitor	HMD	Other
2.) Output hardware	.347 (.223–.472) 441.009	.407 (.208–.605) 401.617	.563 (.041–1.084) 126.408
	Not Included	Included	
3.) Game elements	.322 (.206–.437) 706.679	.685 (.445–.925) 191.246	
	No-treatment	Alternative treatment	
4.) Type of control group	.708 (.465–.950) 146.901	.326 (.211–.440) 772.655	

Note: All Q-values are statistically significant at $p < .05$ level.

TABLE 4. Meta-regression of each moderator variable tested together (using all outcomes).

Variable	Coefficient	*SE*	95% CI	*Z*	*p*
Intercept	−9.261	24.468	−57.218–38.696	−.378	.705
1.) Input hardware	.180	.121	−.057–.417	1.488	.137
2a.) Output hardware (Monitor–HMD)	.157	.126	−.090–.403	1.245	.213
2b.) Output hardware (Monitor–Other)	.228	.210	−.185–.640	1.083	.279
3.) Game elements	.312	.139	.039–.585	2.238	**.025**
4.) Year	.005	.012	−.019–.029	.401	.688
5.) Type of control group	−.429	.149	−.722–−.137	−2.875	**.004**
Q				929.532	**<.001**

k = 186; *n* = 8241 Note: The bold numbers represent statistical significance at a .05 level

TABLE 5. Meta-regression of each moderator variable tested together (only using cognitive outcomes).

Variable	Coefficient	*SE*	95% CI	*Z*	*p*
Intercept	−20.224	36.413	−91.593–51.145	−.555	.579
1.) Input hardware	.223	.193	−.156–.602	1.155	.248
2a.) Output hardware (Monitor–HMD)	.675	.275	.136–1.214	2.455	**.014**
2b.) Output hardware (Monitor–Other)	.973	.412	.165–1.782	2.360	**.018**
3.) Game elements	.732	.295	.155–1.310	2.486	**.013**
4.) Year	.011	.018	−.025–.046	.588	.556
5.) Type of control group	−1.067	.317	−1.688–−.446	−3.367	**.001**
Q				456.106	**<.000**

k = 82; *n* = 3916 Note: The bold numbers represent statistical significance at a .05 level

TABLE 6. Meta-regression of each moderator variable tested together (only using emotional outcomes).

Variable	Coefficient	*SE*	95% CI	*Z*	*p*
Intercept	−37.565	35.005	−106.174–31.043	−1.073	.283
1.) Input hardware	.324	.182	−.033–.681	1.779	.075
2a.) Output hardware (Monitor–HMD)	.349	.177	.002–.695	1.970	**.049**
3b.) Output hardware (Monitor–Other)	−.013	.289	−.580–.554	−.045	.964
3.) Game elements	.287	.223	−.149–.724	1.290	.197
4.) Year	.019	.018	−.015–.053	1.076	.282
5.) Type of control group	−.317	.203	−.714–.080	−1.567	.117
Q				225.939	<.000

k = 78; *n* = 3236 Note: The bold numbers represent statistical significance at a .05 level

TABLE 7. Meta-regression of each moderator variable tested together (only using physical outcomes).

Variable	Coefficient	*SE*	95% CI	*Z*	*p*
Intercept	−114.894	105.848	−322.352 – 92.564	−1.085	.278
1.) Input hardware	Not enough studies tested standard inputs to include in analyses				
2a.) Output hardware (Monitor–HMD)	−.698	.714	−2.097 –.701	−.143	.328
2b.) Output hardware (Monitor–Other)	−.097	.680	−1.430–1.236	−.759	.887
3.) Game elements	−.095	.326	−.734–.544	−.292	.771
4.) Year	.058	.053	−.046–.162	1.091	.275
5.) Type of control group	−.068	.340	−.735–.599	−.201	.841
Q				47.475	**.001**

k = 27; n = 1002 Note: The bold numbers represent statistical significance at a .05 level

Three analyses were used to investigate the hypotheses: standardized effect sizes, confidence intervals, and meta-regressions. In regard to input hardware, studies that applied standard hardware (i.e., keyboard, mouse, and/or joystick) had an average standard effect of .345 (95% C.I. = .210–.480), whereas studies that applied specialized input hardware (i.e., motion sensors, floor pads, etc.) had an average standard effect of .474 (95% C.I. = .309 –.640). When analyzing the meta-regression with all studies together, the effect of input hardware was not significant (β = .180, Z = 1.488, p > .05). When analyzing the meta-regression with studies that developed cognitive outcomes alone (β = .223, Z = 1.155, p > .05) and emotional outcomes alone (β = .324, Z = 1.779, p > .05), the effect of input hardware was again not significant. The effect of input hardware could not be tested for physical outcome development, as only one study did not use specialized input hardware to develop this outcome. Together, Hypotheses 1 was not supported, as specialized input hardware was not more effective across analyses than standard input hardware; however, in the case of physical outcome development, specialized input hardware appear to be mandatory to develop muscle and/or physical abilities.

In regard to output hardware, studies that applied computer monitors had an average standard effect of .347 (95% C.I. = .223–.472); studies that applied HMD had an average standard effect of .407 (95% C.I. = .407–.605); and studies that applied other output hardware (3D Monitors, surround-screen displays, etc.) had an average standard effect of .563 (95% C.I. = .041–1.084). When analyzing the meta-regression with all studies together, the coefficient representing the effect of HMDs in comparison to monitors was not significant (β = .157, Z = 1.245, p > .05) as was the coefficient representing the effect of other output hardware in comparison to monitors (β = .228, Z = 1.083, p > .05). When analyzing the meta-regression with studies that developed cognitive outcomes alone (Monitor–HMD, β = .675, Z = 2.455, p < .05; Monitor–Other, β = .973, Z = 2.360, p < .05) and emotional outcomes alone (Monitor–HMD, β = .349, Z = 1.970, p < .05; Monitor–Other, β = −.013, Z = −.045, p > .05), HMDs were more effective than monitors in both analyses, and other output hardware was more effective than monitors for cognitive outcome development alone. Lastly, output hardware had no notable effects for the meta-regression with physical outcomes alone (Monitor–HMD, $\beta = -.698$, $Z = -.143$, $p > .05$; Monitor–Other, $\beta = -.097$, $Z = -.759$, $p > .05$). Thus, HMDs produce greater effects than monitors when developing cognitive and emotional outcomes, and other output hardware produce greater effects than monitors when developing cognitive outcomes; however, advanced output hardware did not have a significant impact on physical outcome development or when analyzing all outcomes together. Hypotheses 2 and 3 were partially supported.

In regard to game elements, studies that did not include game elements in their VR interventions had an average standard effect of .322 (95% C. I. = .206–.437), whereas those that included game elements had an average standard effect of .685 (95% C.I. = .445–.925). When analyzing the meta-regression with all studies, the effect of game elements was significant ($\beta = .312$, $Z = 2.238$, $p < .05$). The effect of game elements was also significant in the meta-regression only using studies that analyzed cognitive outcome development ($\beta = .732$, $Z = 2.486$, $p < .05$). Alternatively, the meta-regressions for studies that developed emotional outcomes alone ($\beta = .287$, $Z = 1.290$, $p > .05$) and physical outcomes alone ($\beta = -.095$, $Z = -.292$, $p > .05$) indicated that the effect of game elements was not significant. Together, game elements had a significant effect when analyzing all studies and those that developed cognitive outcomes, but they did not have a significant effect when restricting analyses to studies that developed emotional outcomes alone and physical outcomes alone. Hypothesis 4 was partially supported.

Also, Research Question 1 questioned whether game elements were presented in most VR interventions. Of the 192 studies included in the meta-analysis, only 38 studies described game-like elements of their VR intervention. Therefore, VR and serious games are distinct concepts in the current literature, and most VR interventions do not contain game elements.

Lastly, two additional aspects of the studies were investigated: the year that the article was published and the choice of control group (no-treatment vs. alternative treatment). The effect of year was not significant in any meta-regression (All Outcomes. $\beta = .005$, $Z = .401$, $p > .05$; Cognitive Outcomes; $\beta = .011$, $Z = .588$, $p > .05$; Emotional Outcomes; $\beta = .019$, $Z = 1.076$, $p > .05$; Physical Outcomes, $\beta = .058$, $Z = 1.091$, $p > .05$). Alternatively, studies that used no-treatment control groups demonstrated an average standard effect of .708 (95% C.I. = .465–950), whereas studies that used an alternative treatment group had an average standard effect of .326 (95% C.I. = .211–.440). The coefficients representing the type of control group in the meta-regressions analyzing all outcomes together ($\beta = -.429$, $Z = -2.875$, $p < .001$) as well as cognitive outcomes alone ($\beta = -1.067$, $Z = -3.367$, $p < .001$) were very strong and significant, whereas the same coefficients in the meta-regressions analyzing emotional outcomes alone ($\beta = -.317$, $Z = -1.567$, $p > .05$) and physical outcomes alone ($\beta = -.068$, $Z = -.201$, $p > .05$) were not significant.

TABLE 8. Visual representation of meta-analytic results.

	All outcomes together	Cognitive outcomes	Emotional outcomes	Physical outcomes
1.) Input hardware				Could not be tested
2a.) Output hardware (Monitor–HMD)		✓	✓	
2b.) Output hardware (Monitor–Other)		✓		
3.) Game elements	✓	✓		
4.) Year				
5.) Type of control group	✓	✓		

Note: Highlighted boxes with a □ indicate that the variable (row) was statistically significant in the respective application (column).

Therefore, the year that the study was completed did not have a large effect on the resultant effect sizes, but the type of control group had a notable impact.

A visual representation of all significant meta-regression effects is presented in Table 8, which can be used to answer Research Question 2.

4. DISCUSSION

VR has been repeatedly applied for personal development purposes, ranging from learning and training (cognitive outcomes), to psychological therapies (emotional outcomes), to physical rehabilitation (physical outcomes). Across these applications, authors have applied various input hardware, output hardware, and software. In regard to input hardware, authors often discuss the relative impact of specialized inputs (motion sensors, floor pads, etc.) beyond traditional inputs (keyboard, mouse, joystick). In regard to output hardware, authors often discuss the relative impact of advanced outputs (HMDs, surround-screen displays, etc.) beyond traditional outputs (monitor). Lastly, in regard to software, authors often discuss the relative impact of serious games beyond traditional programs. The goal of the current article was to analyze the actual relative impact of these advanced hardware and software across all applications of VR interventions, as well as within the three specific applications of note.

The pattern of findings for each of these VR intervention features are reviewed below. In doing so, suggestions are provided for the causes of these findings based on prior research. Particularly, these findings are related to fidelity/presence, attention, and motivation—of which all studied VR intervention features could influence. Further, as noted in the Background section, it is also believed that certain patterns of results may help identify the mechanism that a particular feature effects. Improvements to presence would be expected to influence all outcomes; improvements to attention would benefit cognitive and perhaps emotional outcomes; and improvements to motivation would benefit cognitive outcomes alone. Thus, if these patterns are found for a VR intervention feature, then the feature may influence the relevant mechanism.

4.1. Input Hardware

The meta-analysis demonstrated that specialized input hardware did not have a significant effect when analyzing all outcomes together, cognitive outcomes alone, and emotional outcomes alone. The effect of specialized input hardware could not be tested for physical outcome development, however, as only one study did not use specialized input hardware to develop physical outcomes. These results suggest that the input hardware has little impact on user outcomes, unless the application requires a specialized input. For example, gait abilities are unable to be developed through a keyboard and mouse, and certain skills may only be developed through using realistic input devices that mimic the application of interest (i.e., treadmill). This widespread use of specialized input hardware can be seen throughout the study of VR interventions for physical development (O'Flynn et al., 2013; Regterschot et al., 2014).

The nonsignificant impact of input hardware may be due to several reasons, four of which are discussed. First, input hardware may have little impact on mechanisms that may subsequently influence outcomes, such as presence; however, this possibility is unlikely given that Cummings and Bailenson (2016) meta-analytically showed that level of tracking, an aspect of input hardware, does have a large influence on presence. Second, it is possible that current input hardware influences important mechanisms, but input hardware may not influence these mechanisms enough to subsequently influence more distal outcomes. If this is the case, then input hardware may still have an influence on VR intervention outcomes, but these effects may not be seen until several more years of development. Third, input hardware may influence mechanisms that have little effect on VR intervention outcomes, whereas other more important mechanisms may be relatively unaffected by input hardware. For instance, input hardware may influence presence, but presence may have little influence on outcomes. Fourth, specialized input hardware may not influence the nature of tasks enough to incur a substantive effect on outcome. In other words, the interventions may prompt similar cognitive and physical processes, regardless of whether a specialized input is used or whether a keyboard and mouse are used. Thus, the current article demonstrated that input hardware may not have a noteworthy effect for developing most outcomes, but only suggestions can be provided for the cause for this result.

It should be noted that some variation was seen in the type of specialized input applied, ranging from replications of surgical tools, to motion sensors, to pressure pads. Analyses that compare the different specialized input devices are provided in the supplemental material. Of the different types, replications of surgical tools and motion sensors had the largest effects, whereas pressure pads and other types of specialized input had smaller effects. The confidence intervals of these specialized input types, except pressure pads, excluded zero; each of their confidence intervals overlapped with each other; and each of their confidence intervals overlapped with the standard input hardware confidence interval. Thus, specialized input devices again saw little effect beyond standard input devices, even when analyzed separately.

4.2. Output Hardware

The meta-analysis showed that advanced output hardware had a significant effect when analyzing cognitive outcomes alone and emotional outcomes alone. Both HMDs and other display devices were more effective than monitors for the former, whereas only HMDs were more effective than monitors for the latter. Output hardware did not demonstrate significant effects with analyzing all outcomes together and physical outcomes alone. These results indicate that current advanced output devices, especially HMDs, may be sufficiently developed to trigger improved VR intervention outcomes. Likewise, these results suggest that the further development of these devices may be a worthwhile endeavor, as possibly even more improved outcomes may be obtained in future research and practice.

The cause of these improvements cannot be definitively identified in the current article, but suggestions can be provided based on the pattern of results and prior theory. Output devices may improve outcomes that involve certain psychological processes, such as cognitive and emotional development, but they may have less of an impact on outcomes that do not involve these psychological processes, such as physical development. Specialized input devices may improve outcomes through prompting feelings of presence (Bouchard et al., 2012; Ling et al., 2013; Wirth et al., 2007). Presence may be important to cognitive of emotional intervention outcomes, but the same may not be true for physical intervention outcomes. Likewise, these results may be due to the ability of HMDs to reduce outside distractions. When using a HMD, the user's entire field of vision becomes the display, which may provide benefits for cognitive and emotional outcomes but not for physical outcomes (Maiberg, 2014; Robinett & Rolland, 1992).

Further, it is not believed that the success of HMDs is due to increased enjoyment and motivation. If this were the case, HMDs would be expected to provide better results for outcomes that are not naturally motivating. Cognitive outcomes may not be naturally motivating, as these are often developed for required work purposes. A long line of training research has supported that trainees may require extra incentives to be properly motivated during a training program (Baldwin, Magjuka, & Loher, 1991; Weissbein, Huang, Ford, & Schmidt, 2011), thereby suggesting that trainees often have deficient motivation to improve cognitive outcomes. On the other hand, participants may already be motivated to develop emotional and physical outcomes, as these are often developed to overcome personal difficulties. For instance, VR interventions have been applied to address depression and phobias (emotional outcomes; Bohil et al., 2011; Rothbaum et al., 2014) as well as gait and balance (physical outcomes; Meldrum et al., 2012; Tao et al., 2013). In each of these cases, improvements to the outcome would likely result in a better quality of life, and trainees may therefore be more naturally motivated to improve these outcomes. If HMDs improved outcomes due to increased motivation, it would be expected to improve cognitive outcomes but have a smaller impact on both emotional and physical outcomes. This was not the case, and the pattern of results for input hardware appear to be more likely due to increased presence or decreased distraction.

Some authors have suggested, however, that participant dropout in interventions for emotional and physical outcomes may be because their motivation is not sufficient for the intensive efforts that they must allocate to the intervention (Crane & Williams, 2010; Postel, De Haan, Ter Huurne, Becker, & De Jong, 2010). Given this suggestion, a significant effect of HMDs on all three types of outcomes (cognitive, emotional, and physical) could support that improvements to motivation was the driving force behind the observed outcomes. While this was not the case, the influence of increased motivation on outcomes cannot be ruled out without additional research.

Lastly, further analyses are reported in the supplemental material that analyzes each type of advanced output device, which provides analyses that differentiate three-dimensional displays, surround-screen displays, and other types of advanced output device. The magnitude of effects, ordered by largest to smallest, were as follows: three-dimensional displays, other types of displays, and surround-screen displays. The confidence interval of other types excluded zero, but the confidence intervals of three-dimensional and surround-screen displays included zero. Each of their confidence intervals overlapped with each other. Thus, while some variation was seen among the other types of advanced output hardware, they did demonstrate similar effects.

4.3. Game Elements

Game elements had a significant impact when analyzing all outcomes together and cognitive outcomes alone, but they had no effect when analyses were restricted to emotional outcomes alone and physical outcomes alone. This suggests that game elements are able to improve outcomes for VR interventions that are not naturally motivating, but they may have less of an effect for interventions to develop outcomes with an intrinsic purpose. Thus, the differential impact of game elements may be due to their impact on user motivation, rather than other psychological processes.

Two further notes should be made about game elements. Most VR interventions did not include game elements, and VR and serious games are two distinct concepts in the modern literature. Also, the current article demonstrated that game elements had a notable effect, but it did not specifically analyze the impact of particular game elements. For instance, the results of the current article cannot be used to explicitly support control, score, or any other specific game element (Bedwell et al., 2012). Thus, while game elements may have a general positive impact on cognitive outcomes, it remains still unclear which specific game elements are effective.

4.4. Outcomes and Other Attributes

A note should be made about studying predictors by the application. When analyzing all outcomes together, game elements and the type of control group were significant. When analyzing cognitive outcomes alone, both types of output devices, game elements, and the type of control group were significant. When analyzing emotional outcomes alone, only the comparison of HMDs and monitors was significant. When analyzing physical outcomes alone, no predictor was statistically significant. While certain predictors may be more predictive across applications, it is also possible that certain applications are more susceptible to effects. If this is true, then cognitive outcomes may be most influenced by attributes of the VR intervention, whereas emotional outcomes and physical outcomes are less influenced. Again, it is possible that users developing these latter outcomes are already motivated, and they are already receiving the best outcomes as possible from their interventions—whether they are technologically advanced or not. While the results of the current article cannot support this notion, future research should certainly investigate this possibility, as discussed below.

Further, two analyses focused on aspects of the studies: the year that the article was published and the choice of control group (no-treatment vs. alternative treatment). Year was not significant in any analysis. The choice of the control group was significant when analyzing all outcomes together and cognitive outcomes alone. When analyzing the results of studies, authors should be cognizant of the choice of control group. As many prior authors have suggested (Keppel, 1991; Kirk, 1982; Winer et al., 1971) and the current article demonstrated, studies using no-treatment control groups (i.e., wait-list designs) often result in larger effects than studies using alternative treatment groups. Thus, although a VR intervention may appear to be more effective than alternatives, its effects may actually be driven by the choice of control group.

4.5. Implications and Future Directions

With these results taken into consideration, implications and future directions should be noted. First, authors should continue the study of specialized input hardware with a particular focus on two directions. As mentioned, specialized input hardware are required to develop certain personal characteristics, such as physical abilities and particular skills (Burdea et al., 1996; Kihara et al., 2012; Mon-Williams et al., 1993; O'Flynn et al., 2013), and these specialized input hardware often provide tactile resistance to certain physical movements. Without a doubt, specialized input hardware should still be applied and studied for this purpose; muscle development, for example, would be difficult to encourage without manipulating the muscles of interest. Alternatively, specialized input hardware are sometimes applied in situations in which it is not required, often with the intentions to provide a more realistic experience. Given that specialized input hardware were not more effective than traditional input hardware in any analyses, these applications may be relatively ineffective. A possible future direction, however, could be the creation and application of specialized input devices meant to improve other aspects of VR interventions, such as user enjoyment and subsequent motivation. Indeed, some authors have already created and applied specialized input devices for this purpose, providing some initial insights for future research and practice (Lange et al., 2011; Li, Lou, Tsai, & Shih, 2012; Shafer, Carbonara, & Popova, 2011), and this research stream may continue to be a fruitful endeavor.

Second, authors should further investigate advanced output hardware, particularly HMDs. As mentioned, Facebook purchased Oculus for two billion dollars to own the production of the Rift, and the application of HMDs is expected to grow in the future. Further, the current article demonstrated that HMDs had a significant and positive influence on VR interventions to develop cognitive and emotional outcomes. Prior authors have proposed that HMDs influence VR intervention outcomes by heightening perceptions of presence (Bouchard et al., 2012; Bulu, 2012; Ling et al., 2013; Wirth et al., 2007). The current findings support this notion. VR interventions for cognitive and emotional outcomes, specifically, may benefit from HMDs because users experience heightened immersion and presence which may result in heightened cognitive and/or emotional engagement. From this heightened engagement, users may be more attuned to the information and experiences provided through the HMD. This notion is also in agreement with the null effect of HMDs on physical outcome development, as users may not need to be engaged with the VR intervention to develop physical abilities. Instead, they may only need to be motivated. Thus, while the current article supported that immersion, presence, and engagement may explain the link between HMDs and outcomes, future research is needed to explicitly demonstrate these mediators as the cause of beneficial VR intervention results.

In performing these subsequent investigations, it would also be important to perform a finer grained analysis of VR output hardware. The current meta-analysis only intended to provide broad strokes of the investigated hardware and software together, but many other characteristics may be important in predicting the effectiveness of VR output hardware. Particularly, output devices may have a differing resolution or field-of-vision or resolution, which should be considered in future investigations on VR interventions.

Relatedly, input and output hardware have largely been studied in isolation. As these hardware are often developed from a common theoretical perspective, more research should study their effects together. It is possible, if not probable, that interaction effects may be present between many specialized input hardware and advanced output hardware, and these interactions are undiscovered in current research. Further, authors should apply novel experimental designs when testing for these interactions to simultaneously study the direct effects and interactions of the technologies with the same methodological rigor as studying the direct effects alone (Collins, Murphy, & Strecher, 2007; Shadish, Cook, & Campbell, 2002). Through these efforts, a more complete understanding can be obtained about input hardware, output hardware, and their combined effects.

Third, research should further study the cause of game element benefits on VR interventions. Several prior authors have suggested that game elements are naturally entertaining (Bedwell et al., 2012; Deterding et al., 2011; Sitzmann, 2011). Also, more recent research has integrated popular motivational theory, such as self-determination theory, into serious game scholarship to suggest that certain game elements may provide motivational benefits (Deci & Ryan, 2011; Ryan & Deci, 2000; Thompson et al., 2010; Wouters et al., 2009). The findings of the current article support these notions. Serious game only had a significant impact on cognitive outcome development. When developing cognitive outcomes, users may be generally unmotivated. For example, a host of research has studied methods to improve employee and student motivations to learn, both being primary targets of cognitive outcome development (Baldwin et al., 1991; Orpen, 1999; Weissbein et al., 2011). These studies repeatedly note that employees and students may not naturally value the outcomes from their intervention programs, and they may see these outcomes as unrewarding and/or distal. Game elements may motivate these otherwise unmotivated individuals, resulting in improved outcomes.

When developing emotional and physical outcomes, however, users may be more intrinsically motivated (Gold et al., 2013; Miller, 1985; Roth & Fonagy, 2013). While not always, users often voluntarily choose to undergo a therapy or rehabilitation program to improve certain personal attributes. When undergoing these VR interventions, the outcomes have a direct impact on the well-being of the user, such as regaining the ability to walk, and users may be more conscious of these immediate and proximal outcomes. In these cases, game elements may have less of an impact, as users are already motivated to achieve personally valuable outcomes. Future research is needed to firmly support this notion, but the results of the current study are in agreement with prior suggestions.

Relatedly, the current article only investigated the overall impact of including any game elements in a VR intervention. As mentioned, several game element taxonomies exist (Bedwell et al., 2012; King et al., 2010; Rego et al., 2010), and these taxonomies can be used to study individual game elements, such as challenge and control. Future research should perform a more honed analysis of game elements to determine their effect on a VR intervention.

Fourth, authors should consider another meta-analysis that tests the influence of VR interventions and certain design characteristics on the mechanisms that were proposed to influence outcomes. Increased motivation has been suggested as a mediating mechanism, which could be included in a future meta-analysis. Likewise, presence is often attributed to VR intervention success (Cairns et al., 2013; Mirelman et al., 2010; Schmierbach et al., 2012). Cummings and Bailenson's (2016) meta-analysis showed that immersion relates to presence, and the current meta-analysis showed that VR interventions may improve outcomes. A meta-analysis analyzing the effect of presence on outcomes could provide the final piece of the causal puzzle, and a full understanding of VR interventions, presence, and outcomes could be achieved.

Fifth, authors should perform meta-analyses that have more focused investigations into a single outcome category. For instance, it is likely that the effect of VR is not the same for the development of all cognitive outcomes, and a more focused evaluation could determine the conditions that VR is more effective. Particularly, because VR is able to present a complete visual environment, the technology may be able to better train complex skills and abilities that require the interpretation of multiple cues rather than more basic skills (Burdea & Coiffet, 2003; Mirelman et al., 2010; Seth et al., 2011). It is also expected that similar results could be found across the other outcomes studied in the current article.

4.6. Limitations

With the implications and future directions noted, certain limitations of the current article should also be considered. As with all meta-analyses, the current article was able to provide broad, overall findings, but it was unable to consider certain contextual factors of VR interventions. For instance, while serious games were shown to have a notable impact on cognitive outcome development, it is possible that cognitive outcomes are relatively unaffected by serious game elements when user motivation is high. For this reason, meta-analyses are unable to provide completely conclusive inferences about a phenomenon of interest, but the results can nevertheless guide future research and practice on VR interventions.

Additionally, several efforts were made to discover unpublished studies that investigated a VR intervention, but it is impossible to conclusively determine that all studies were discovered. For this reason, the results of the publication bias analyses should be carefully considered, and certain effects should only be interpreted while considering that these biases may affect any observed results. It should be noted, however, that controlled studies of VR interventions are relatively difficult to perform, and many journals will consider publishing such studies even if results are not significant. While publication biases may be a notable concern for many meta-analyses, this factor may reduce these concerns in the current article.

Meta-analytic results are only as valid as the studies analyzed. If certain methodological concerns are persistent in the literature on VR interventions, then a meta-analysis would be biased by these concerns. Given that the current meta-analysis contained a large number of articles (192), biases of individual or even several studies would not greatly impact the results.

The current article also assumed that the technologies included within each category were relatively homogeneous. For example, HMDs were assumed to be overall more immersive than monitors. Not all devices within these categories are identical, however. Some HMDs may have smaller fields of view and/or worse resolutions than some monitors, and similar sentiments can be expressed for the other comparison categories. For this reason, the current results should be seen as general, broad inferences about the differences between these technology categories, and they should not be seen as absolute truths across all possible technology comparisons.

Lastly, when coding for game elements, the current article included commercial games, serious games, as well as traditional simulations with game elements included. While each of these types of simulations include game elements, prior research has argued that distinctions between these experiences have important implications for user motivation and possibly presence (Bedwell et al., 2012; Deterding et al., 2011; Sitzmann, 2011). While the current article did not provide nuanced analyses that separated these three types of simulations, a future meta-analysis on game elements should consider such analyses. Doing so could provide important inferences on their relative influence on user development and ultimately performance.

5. CONCLUSION

The current article investigated the impact of input hardware, output hardware, and game elements on VR interventions for cognitive, emotional, and physical outcome development. The results suggested that input hardware did not have a notable effect across outcomes, but certain applications, such as VR physical rehabilitation programs, require specialized input hardware to develop relevant outcomes. Output hardware had a notable impact on cognitive and emotional outcome development, suggesting that output hardware influences presence and engagement. Lastly, game elements had a significant impact on cognitive outcome development, suggesting that game elements may motivate otherwise unmotivated users. Together, these results provide several inferences about the benefits (and lack of benefits) of popular hardware and software applied in VR interventions for personal development.

NOTES

Background. The current article is based on the dissertation of the author.

Acknowledgments. I would like to thank Rick Jacobs, James LeBreton, and Sam Hunter for their comments on previous versions of the ideas presented in the current article. I would also like to thank the Business and CyberPsychology Lab, especially Juseob Lee, for their work that contributed to the current article.

Supplemental Data. Supplemental data for this article can be accessed at www.tandfonline.com/hhci.

HCI Editorial Record. First received February 17, 2016. Revisions received April 12, 2018. Accepted by Susan Fussell. Final manuscript received on April 12, 2018. — *Editor*

REFERENCES

Adkins, S. S. (2013). The 2012–2017 worldwide game-based learning and simulation-based markets: Key findings from recent Ambient Insight research. Serious Play Conference. Retrieved from http://www.ambientinsight.com/Resources/Documents/AmbientInsight_SeriousPlay2013_WW_GameBasedLearning_Market.pdf

Aguinis, H., Pierce, C. A., Bosco, F. A., Dalton, D. R., & Dalton, C. M. (2011). Debunking myths and urban legends about meta-analysis. *Organizational Research Methods*, *14*(2), 306–331. doi:10.1177/1094428110375720

Aguinis, H., Sturman, M. C., & Pierce, C. A. (2008). Comparison of three meta-analytic procedures for estimating moderating effects of categorical variables. *Organizational Research Methods*, *11*(1), 9–34.

Anderson, P. L., Price, M., Edwards, S. M., Obasaju, M. A., Schmertz, S. K., Zimand, E., & Calamaras, M. R. (2013). Virtual reality exposure therapy for social anxiety disorder: A randomized controlled trial. *Journal of Consulting and Clinical Psychology*, *81*(5), 751. doi:10.1037/a0033559

Baldwin, T. T., Magjuka, R. J., & Loher, B. T. (1991). The perils of participation: Effects of choice of training on trainee motivation and learning. *Personnel Psychology*, *44*(1), 51. doi:10.1111/j.1744-6570.1991.tb00690.x

Bohil, C. J., Alicea, B., & Biocca, F. A. (2011). Virtual reality in neuroscience research and therapy. *Nature reviews neuroscience*, 12(12), 752.

Borenstein, M., Hedges, L., Higgins, J., & Rothstein, H. (2005). *Comprehensive meta-analysis version 2*. Englewood, NJ: Biostat.

Bouchard, S., Dumoulin, S., Talbot, J., Ledoux, A. A., Phillips, J., Monthuy-Blanc, J., … Renaud, P. (2012). Manipulating subjective realism and its impact on presence: Preliminary results on feasibility and neuroanatomical correlates. *Interacting with Computers*, *24*(4), 227–236. doi:10.1016/j.intcom.2012.04.011

Brütsch, K., Koenig, A., Zimmerli, L., Mérillat-Koeneke, S., Riener, R., Jäncke, L., … & Meyer-Heim, A. (2011). Virtual reality for enhancement of robot-assisted gait training in children with neurological gait disorders. *Journal of Rehabilitation Medicine*, 43(6),493–499.

Bulu, S. T. (2012). Place presence, social presence, co-presence, and satisfaction in virtual worlds. *Computers & Education*, *58*(1), 154–161. doi:10.1016/j.compedu.2011.08.024

Burdea, G., & Coiffet, P. (2003). Virtual reality technology. *Presence: Teleoperators and Virtual Environments*, *12*(6), 663–664. doi:10.1162/105474603322955950

Burdea, G., Richard, P., & Coiffet, P. (1996). Multimodal virtual reality: Input-output devices, system integration, and human factors. *International Journal of Human-Computer Interaction*, *8*(1), 5–24. doi:10.1080/10447319609526138

Cairns, P., Cox, A. L., Day, M., Martin, H., & Perryman, T. (2013). Who but not where: The effect of social play on immersion in digital games. *International Journal of Human-Computer Studies*, *71*(11), 1069–1077. doi:10.1016/j.ijhcs.2013.08.015

Chang, Y. J., Chen, S. F., & Huang, J. D. (2011). A Kinect-based system for physical rehabilitation: A pilot study for young adults with motor disabilities. *Research in Developmental Disabilities*, *32*(6), 2566–2570. doi:10.1016/j.ridd.2011.07.002

Collins, L. M., Murphy, S. A., & Strecher, V. (2007). The multiphase optimization strategy (MOST) and the sequential multiple assignment randomized trial (SMART): New methods for more potent eHealth interventions. *American Journal of Preventive Medicine*, *32*(5), S112–S118. doi:10.1016/j.amepre.2007.01.022

Crane, C., & Williams, J. M. G. (2010). Factors associated with attrition from mindfulness-based cognitive therapy in patients with a history of suicidal depression. *Mindfulness*, *1*(1), 10–20. doi:10.1007/s12671-010-0003-8

Cummings, J. J., & Bailenson, J. N. (2016). How immersive is enough? A meta-analysis of the effect of immersive technology on user presence. *Media Psychology*, *19*(2), 272–309. doi:10.1080/15213269.2015.1015740

Deci, E. L., & Ryan, R. M. (2011). Self-determination theory. *Handbook of Theories of Social Psychology*, *1*, 416–433.

Deterding, S., Sicart, M., Nacke, L., O'Hara, K., & Dixon, D. (2011, May). Gamification. Using game-design elements in non-gaming contexts. In *CHI'11 Extended Abstracts on Human Factors in Computing Systems* (pp. 2425–2428) Vancouver, BC: Canada. ACM, May 7-12, 2011.

Diemer, J., Alpers, G. W., Peperkorn, H. M., Shiban, Y., & Mühlberger, A. (2015). The impact of perception and presence on emotional reactions: A review of research in virtual reality. *Frontiers in Psychology*, *6*, 26. doi:10.3389/fpsyg.2015.00026

Dow, S., Mehta, M., Harmon, E., MacIntyre, B., & Mateas, M. (2007, April). Presence and engagement in an interactive drama. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, April 28-May 3, 2007 (pp. 1475–1484), San Jose, CA: USA. ACM.

Fox, J., Bailenson, J., & Binney, J. (2009). Virtual experiences, physical behaviors: The effect of presence on imitation of an eating avatar. *Presence: Teleoperators and Virtual Environments*, *18*(4), 294–303. doi:10.1162/pres.18.4.294

Gold, C., Mössler, K., Grocke, D., Heldal, T. O., Tjemsland, L., Aarre, T., … Rolvsjord, R. (2013). Individual music therapy for mental health care clients with low therapy motivation: Multicentre randomised controlled trial. *Psychotherapy and Psychosomatics*, *82*(5), 319–331. doi:10.1159/000348452

Greffou, S., Bertone, A., Hahler, E. M., Hanssens, J. M., Mottron, L., & Faubert, J. (2012). Postural hypo-reactivity in autism is contingent on development and visual environment: A fully immersive virtual reality study. *Journal of Autism and Developmental Disorders*, *42*(6), 961–970. doi:10.1007/s10803-011-1326-6

Griffeth, R. W., Hom, P. W., & Gaertner, S. (2000). A meta-analysis of antecedents and correlates of employee turnover: Update, moderator tests, and research implications for the next millennium. *Journal of Management*, *26*, 463–488. doi:10.1177/014920630002600305

Grossberg, S. (1999). The link between brain learning, attention, and consciousness. *Consciousness and Cognition*, *8*(1), 1–44. doi:10.1006/ccog.1998.0372

Harrison, D. A., Newman, D. A., & Roth, P. L. (2006). How important are job attitudes? Meta-analytic comparisons of integrative behavioral out-comes and time sequences. *Academy of Management Journal*, *49*, 305–325. doi:10.5465/AMJ.2006.20786077

Hedges, L. V., & Olkin, I. (2014). *Statistical methods for meta-analysis.*, Massachusetts: Cambridge, Academic press.

Hou, J., Nam, Y., Peng, W., & Lee, K. M. (2012). Effects of screen size, viewing angle, and players' immersion tendencies on game experience. *Computers in Human Behavior*, *28*(2), 617–623. doi:10.1016/j.chb.2011.11.007

Howard, M. C. (2017a). Investigating the simulation elements of environment and control: Extending the uncanny valley theory to simulations. *Computers & Education*, *109*, 216–232. doi:10.1016/j.compedu.2017.03.005

Howard, M. C. (2017b). A meta-analysis and systematic literature review of virtual reality rehabilitation programs. *Computers in Human Behavior*, *70*, 317–327. doi:10.1016/j.chb.2017.01.013

Huang, H. M., Rauch, U., & Liaw, S. S. (2010). Investigating learners' attitudes toward virtual reality learning environments: Based on a constructivist approach. *Computers & Education*, *55*(3), 1171–1182. doi:10.1016/j.compedu.2010.05.014

Ip, H. H. S., Byrne, J., Lau, K. S. Y., Li, R. C., Tso, A., & Choi, C. (2013). Interactive Sensory Program for Affective Learning (InSPAL): An innovative learning program combining interactive media and virtual reality for severely intellectually disabled students. In Fong, S. K. C. J., Cheung, S. K., Fong, W., & Wang, F. L. (Eds.), *Hybrid learning and continuing education* (pp. 199–207), Berlin. Springer Berlin Heidelberg.

Jayaram, S., Connacher, H. I., & Lyons, K. W. (1997). Virtual assembly using virtual reality techniques. *Computer-Aided Design*, *29*(8), 575–584. doi:10.1016/S0010-4485(96)00094-2

Kapp, K. M. (2012). *The gamification of learning and instruction: Game-based methods and strategies for training and education*, Hoboken: NJ. John Wiley & Sons.

Keppel, G. (1991). *Design and analysis: A researcher's handbook*, Upper Saddle River: New Jersey. Prentice-Hall, Inc.

Kihara, K., Fujii, Y., Masuda, H., Saito, K., Koga, F., Matsuoka, Y., … Kojima, K. (2012). New three-dimensional head-mounted display system, TMDU-S-3D system, for minimally invasive surgery application: Procedures for gasless single-port radical nephrectomy. *International Journal of Urology*, *19*(9), 886–889. doi:10.1111/j.1442-2042.2012.03044.x

King, D., Delfabbro, P., & Griffiths, M. (2010). Video game structural characteristics: A new psychological taxonomy. *International Journal of Mental Health and Addiction*, *8*(1), 90–106. doi:10.1007/s11469-009-9206-4

Kirk, R. E. (1982). *Experimental design*, Hoboken: NJ. John Wiley & Sons, Inc.

Krijn, M., Emmelkamp, P. M. G., Biemond, R., De Wilde, D. L., Schuemie, M. J., & Van, D. M. (2004). Treatment of acrophobia in virtual reality: The role of immersion and presence. *Behaviour Research and Therapy*, *42*(2), 229–239. doi:10.1016/S0005-7967(03)00139-6

Kwon, J. H., Powell, J., & Chalmers, A. (2013). How level of realism influences anxiety in virtual reality environments for a job interview. *International Journal of Human-Computer Studies*, *71*(10), 978–987. doi:10.1016/j.ijhcs.2013.07.003

Lange, B., Chang, C. Y., Suma, E., Newman, B., Rizzo, A. S., & Bolas, M. (2011, August). Development and evaluation of low cost game-based balance rehabilitation tool using the Microsoft Kinect sensor. In *Engineering in Medicine and Biology Society, EMBC, 2011 Annual International Conference of the IEEE*, August 30 - September 3, 2011 (pp. 1831–1834). IEEE, Boston, Massachusetts: USA.

LeBreton, J. M., Scherer, K. T., & James, L. R. (2014). Corrections for criterion reliability in validity generalization: A false prophet in a land of suspended judgment. *Industrial and Organizational Psychology*, *7*(4), 478–500.

Leong, P. (2011). Role of social presence and cognitive absorption in online learning environments. *Distance Education*, *32*(1), 5–28. doi:10.1080/01587919.2011.565495

Li, K. H., Lou, S. J., Tsai, H. Y., & Shih, R. C. (2012). The effects of applying game-based learning to webcam motion sensor games for autistic students' sensory integration training. *Turkish Online Journal of Educational Technology—TOJET*, *11*(4), 451–459.

Ling, Y., Nefs, H. T., Brinkman, W. P., Qu, C., & Heynderickx, I. (2013). The relationship between individual characteristics and experienced presence. *Computers in Human Behavior*, *29*(4), 1519–1530. doi:10.1016/j.chb.2012.12.010

Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis.*, Thousand Oaks: California, Sage Publications, Inc.

Lombard, M. (2000). *Resources for the study of presence: Presence explication*. Retrieved September 2000, from http://nimbus.temple.edu/~mlombard/Presence/explicat.htm

Lull, R. B., & Bushman, B. J. (2016). Immersed in violence: Presence mediates the effect of 3D violent video gameplay on angry feelings. *Psychology of Popular Media Culture*, *5*(2), 133. doi:10.1037/ppm0000062

Maiberg, E. (2014). Minecraft's Notch now okay with Facebook buying Oculus, John Carmack offers help. Gamespot. Gamespot.com. Retrieved from http://www.gamespot.com/articles/minecrafts-notch-now-okay-with-facebook-buying-ocu/1100-6421777/

Malbos, E., Rapee, R. M., & Kavakli, M. (2013). A controlled study of agoraphobia and the independent effect of virtual reality exposure therapy. *Australian and New Zealand Journal of Psychiatry*, *47*(2), 160–168. doi:10.1177/0004867412453626

Martel, E., Su, F., Gerroir, J., Hassan, A., Girouard, A., & Muldner, K. (2015). Diving head-first into virtual reality: Evaluating HMD control schemes for VR games. Proceedings of the 10th International Conference on the Foundations of Digital Games (FDG 2015), FDG, June 22-25, 2015, Pacific Grove, CA, USA.

Mayes, S. D., & Calhoun, S. L. (2007). Learning, attention, writing, and processing speed in typical children and children with ADHD, autism, anxiety, depression, and oppositional-defiant disorder. *Child Neuropsychology*, *13*(6), 469–493. doi:10.1080/09297040601112773

Mcquiggan, S. W., Rowe, J. P., Lee, S., & Lester, J. C. (2008, June). Story-based learning: The impact of narrative on learning experiences and outcomes. In International Conference on *Intelligent Tutoring Systems* (pp. 530–539). Springer: Berlin.

Meldrum, D., Glennon, A., Herdman, S., Murray, D., & McConn-Walsh, R. (2012). Virtual reality rehabilitation of balance: assessment of the usability of the Nintendo Wii® Fit Plus. *Disability and rehabilitation: assistive technology*, 7(3),205–210.

Mikropoulos, T. A., & Strouboulis, V. (2004). Factors that influence presence in educational virtual environments. *CyberPsychology & Behavior*, *7*(5), 582–591. doi:10.1089/cpb.2004.7.582

Miller, W. R. (1985). Motivation for treatment: A review with special emphasis on alcoholism. *Psychological Bulletin*, *98*(1), 84. doi:10.1037/0033-2909.98.1.84

Mirelman, A., Patritti, B. L., Bonato, P., & Deutsch, J. E. (2010). Effects of virtual reality training on gait biomechanics of individuals post-stroke. *Gait & posture*, 31(4),433–437.

Monahan, T., McArdle, G., & Bertolotto, M. (2008). Virtual reality for collaborative e-learning. *Computers & Education*, *50*(4), 1339–1353. doi:10.1016/j.compedu.2006.12.008

Mon-Williams, M., Warm, J. P., & Rushton, S. (1993). Binocular vision in a virtual world: Visual deficits following the wearing of a head-mounted display. *Ophthalmic and Physiological Optics*, *13*(4), 387–391. doi:10.1111/opo.1993.13.issue-4

O'Flynn, B., Sanchez, J. T., Angove, P., Connolly, J., Condell, J., Curran, K., & Gardiner, P. (2013, May). Novel smart sensor glove for arthritis rehabilitation. In *Body Sensor Networks (BSN), 2013 IEEE International Conference on* 6-9 May 2013 (pp. 1–6). IEEE, Cambridge: MA.

Orpen, C. (1999). The influence of the training environment on trainee motivation and perceived training quality. *International Journal of Training and Development*, *3*(1), 34–43. doi:10.1111/ijtd.1999.3.issue-1

Postel, M. G., de Haan, H. A., Ter Huurne, E. D., Becker, E. S., & de Jong, C. A. (2010). Effectiveness of a web-based intervention for problem drinkers and reasons for dropout: Randomized controlled trial. *Journal of Medical Internet Research*, *12*(4), e68. doi:10.2196/jmir.1642

Price, M., Mehta, N., Tone, E. B., & Anderson, P. L. (2011). Does engagement with exposure yield better outcomes? Components of presence as a predictor of treatment response for virtual reality exposure therapy for social phobia. *Journal of Anxiety Disorders*, *25*(6), 763–770. doi:10.1016/j.janxdis.2011.03.004

Rahman, S. A. (2010).Efficacy of virtual reality-based therapy on balance in children with Down syndrome. *World Applied Sciences Journal*, 10(3),254–261.

Rego, P., Moreira, P. M., & Reis, L. P. (2010, June). Serious games for rehabilitation: A survey and a classification towards a taxonomy. In *Information Systems and Technologies (CISTI), 2010 5th Iberian Conference on* 16-19 June 2010 (pp. 1–6). IEEE; Santiago de Compostela: Spain.

Rizzo, A., John, B., Newman, B., Williams, J., Hartholt, A., Lethin, C., & Buckwalter, J. G. (2013). Virtual reality as a tool for delivering PTSD exposure therapy and stress resilience training. *Military Behavioral Health*, 1(1),52–58.

Regterschot, G. R. H., Folkersma, M., Zhang, W., Baldus, H., Stevens, M., & Zijlstra, W. (2014). Sensitivity of sensor-based sit-to-stand peak power to the effects of training leg strength, leg power and balance in older adults. *Gait & Posture*, *39*(1), 303–307. doi:10.1016/j.gaitpost.2013.07.122

Robinett, W., & Rolland, J. P. (1992). A computational model for the stereoscopic optics of a head-mounted display. *Presence: Teleoperators and Virtual Environments*, *1*(1), 45–62. doi:10.1162/pres.1992.1.1.45

Rostami, H. R., Arastoo, A. A., Nejad, S. J., Mahany, M. K., Malamiri, R. A., & Goharpey, S. (2012). Effects of modified constraint-induced movement therapy in virtual environment on upper-limb function in children with spastic hemiparetic cerebral palsy: a randomised controlled trial. *NeuroRehabilitation*, 31(4),357–365.

Roth, A., & Fonagy, P. (2013). *What works for whom? A critical review of psychotherapy research*, New York City. Guilford Publications.

Rothbaum, B. O., Price, M., Jovanovic, T., Norrholm, S. D., Gerardi, M., Dunlop, B., & Ressler, K. J. (2014). A randomized, double-blind evaluation of D-cycloserine or alprazolam combined with virtual reality exposure therapy for posttraumatic stress disorder in Iraq and Afghanistan War veterans. *American Journal of Psychiatry*, 171(6),640–648.

Roy, A. K., Soni, Y., & Dubey, S. (2013, August). Enhancing effectiveness of motor rehabilitation using kinect motion sensing technology. In *Global Humanitarian Technology Conference: South Asia Satellite (GHTC-SAS)*, on 23-24 Aug. 2013 (pp. 298–304). IEEE Trivandrum: India.

Rushton, S., Mon-Williams, M., & Wann, J. P. (1994). Binocular vision in a bi-ocular world: New-generation head-mounted displays avoid causing visual deficit. *Displays*, *15*(4), 255–260. doi:10.1016/0141-9382(94)90073-6

Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist*, *55*(1), 68. doi:10.1037/0003-066X.55.1.68

Salmon, G. (2009). The future for (second) life and learning. *British Journal of Educational Technology*, *40*(3), 526–538. doi:10.1111/bjet.2009.40.issue-3

Sanchez-Vives, M. V., & Slater, M. (2005). From presence to consciousness through virtual reality. *Nature Reviews Neuroscience*, *6*(4), 332–339. doi:10.1038/nrn1651

Schmidt, F. L., & Hunter, J. E. (2014). *Methods of meta-analysis: Correcting error and bias in research findings.*, Thousand Oaks: California, Sage publications.

Schmierbach, M., Limperos, A. M., & Woolley, J. K. (2012). Feeling the need for (personalized) speed: How natural controls and customization contribute to enjoyment of a racing game through enhanced immersion. *Cyberpsychology, Behavior, and Social Networking*, *15*(7), 364–369. doi:10.1089/cyber.2012.0025

Seth, A., Vance, J. M., & Oliver, J. H. (2011). Virtual reality for assembly methods prototyping: A review. *Virtual Reality*, *15*(1), 5–20. doi:10.1007/s10055-009-0153-y

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*, Boston: MA. Wadsworth Cengage learning.

Shafer, D. M., Carbonara, C. P., & Popova, L. (2011). Spatial presence and perceived reality as predictors of motion-based video game enjoyment. *Presence: Teleoperators and Virtual Environments*, *20*(6), 591–619. doi:10.1162/PRES_a_00084

Shibata, T. (2002). Head mounted display. *Displays*, *23*(1), 57–64. doi:10.1016/S0141-9382(02)00010-0

Sitzmann, T. (2011). A meta-analytic examination of the instructional effectiveness of computer-based simulation games. *Personnel Psychology*, *64*(2), 489–528. doi:10.1111/peps.2011.64.issue-2

Slater, M. (2003). A note on presence terminology. *Presence Connect*, *3*(3), 1–5.

Slater, M., & Wilbur, S. (1997). A framework for immersive virtual environments (FIVE): Speculations on the role of presence in virtual environments. *Presence: Teleoperators and Virtual Environments*, *6*(6), 603–616. doi:10.1162/pres.1997.6.6.603

Steuer, J. (1992). Defining virtual reality: Dimensions determining telepresence. *Journal of Communication*, *42*(4), 73–93. doi:10.1111/jcom.1992.42.issue-4

Tao, G., Archambault, P. S., & Levin, M. F. (2013, August). Evaluation of Kinect skeletal tracking in a virtual reality rehabilitation system for upper limb hemiparesis. *In Virtual Rehabilitation (ICVR)*, 2013 International Conference on (pp. 164-165). IEEE.

Thompson, D., Baranowski, T., Buday, R., Baranowski, J., Thompson, V., Jago, R., & Griffith, M. J. (2010). Serious video games for health: How behavioral science guided the design of a game on diabetes and obesity. *Simulation & Gaming*, *41*(4), 587–606.

Van Iddekinge, C. H., Roth, P. L., Raymark, P. H., & Odle-Dusseau, H. (2012). The criterion-related validity of integrity tests: An updated meta-analysis. *Journal of Applied Psychology*, *97*(3), 499–530. doi:10.1037/a0021196

Ward, M., Azuma, R., Bennett, R., Gottschalk, S., & Fuchs, H. (1992, June). A demonstrated optical tracker with scalable work area for head-mounted display systems. In *Proceedings of the 1992 symposium on Interactive 3D graphics* on 29 March- 1 April 1992 (pp. 43–52), Massachusetts: Cambridge. ACM.

Weissbein, D. A., Huang, J. L., Ford, J. K., & Schmidt, A. M. (2011). Influencing learning states to enhance trainee motivation and improve training transfer. *Journal of Business and Psychology*, *26*(4), 423–435. doi:10.1007/s10869-010-9198-x

Winer, B. J., Brown, D. R., & Michels, K. M. (1971). *Statistical principles in experimental design* (Vol. 2, pp. 596). New York: USA, McGraw-Hill.

Wirth, W., Hartmann, T., Böcking, S., Vorderer, P., Klimmt, C., Schramm, H., … & Biocca, F. (2007). A process model of the formation of spatial presence experiences. *Media psychology*, 9(3),493–525.

Witmer, B. G., & Singer, M. J. (1998). Measuring presence in virtual environments: A presence questionnaire. *Presence: Teleoperators and Virtual Environments*, *7*(3), 225–240. doi:10.1162/105474698565686

Wouters, P., Van Der Spek, E. D., & Van Oostendorp, H. (2009). Current practices in serious game research: A review from a learning outcomes perspective. In Connolly, T. (Ed.). *Games-Based Learning Advancements for Multisensory Human Computer Interfaces: Techniques and Effective Practices* (pp. 232–255). IGI Global: Hershey, PA.

Yen, C. Y., Lin, K. H., Hu, M. H., Wu, R. M., Lu, T. W., & Lin, C. H. (2011). Effects of virtual reality–augmented balance training on sensory organization and attentional demand for postural control in people with parkinson disease: a randomized controlled trial. *Physical therapy*, 91(6),862–874.

Zichermann, G., & Cunningham, C. (2011). *Gamification by design: Implementing game mechanics in web and mobile apps*., Sebastopol, CAO'Reilly Media, Inc.

Zimmerman, R. A. (2008). Understanding the impact of personality traits on individuals' turnover decisions: A meta-analytic path model. *Personnel Psychology*, *61*, 309–348. doi:10.1111/j.1744-6570.2008.00115.x