Computers & Education 144 (2020) 103707

Contents lists available at ScienceDirect

Computers & Education

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

A meta-analysis of virtual reality training programs for social skill development

Matt C. Howard [a,*], Melissa B. Gutworth [b]

[a] Mitchell College of Business, University of South Alabama, 5811 USA Drive S., Rm. 337, Mobile, AL, 36688, United States

[b] Feliciano School of Business, Montclair State University, 1 E. Normal Ave., Rm. 455, Upper Montclair, NJ, 07043, United States

ARTICLE INFO

Keywords:
Virtual reality
Social skills
Learning
Training and development
Simulation

ABSTRACT

Recent years have seen an increase in the number of researchers and practitioners applying virtual reality (VR) to develop social skills, but varying levels of success have been observed when applying these VR training programs. Due to these disparities, review and summary work is needed to determine (a) whether these programs are effective and (b) the attributes of these programs that lead to success. In the current article, we perform a meta-analysis to determine the overall effectiveness of VR training programs for developing social skills, and we also study the effect of several moderating variables that may influence the effectiveness of these programs. We test whether certain aspects of the applied hardware (e.g. input devices, output devices), applied software (e.g. game elements, Second Life), participant population (e.g. general/specialized), and study design (e.g. type of control group, type of measure, others) influence the success of these programs. Sources were identified using EBSCO and Google Scholar, and, after our inclusion criteria were applied, 23 samples were included in our meta-analysis. Using a random-effects approach, we show that VR training programs, on average, perform better than alternative training programs for developing social skills, but almost all the other findings contradict current notions regarding these programs. Of note, gamified programs produced slightly worse outcomes than non-gamified programs, and programs utilizing immersive technologies (e.g. head-mounted displays) produced slightly worse outcomes than programs utilizing non-immersive displays (e.g. monitors). We provide suggestions regarding the effectiveness of VR training for social skills in comparison to alternative training approaches, the attributes of VR training programs that produce better outcomes, and directions for future research and practice.

Virtual reality (VR) is the presentation of a three-dimensional environment that emulates real or imagined spaces with interactive properties,[1] and it has been increasingly applied in recent years for the training and development of both employees and students (Fowler, 2015; Merchant, Goetz, Cifuentes, Keeney-Kennicutt, & Davis, 2014; Seth, Vance, & Oliver, 2011). While the technology is traditionally used to develop skills that cannot be easily or safely trained otherwise (e.g. flying, surgery), researchers and practitioners have begun applying the technology to develop a much wider range of knowledge, skills, abilities, and other characteristics (KSAOs; e.

* Corresponding author.
E-mail addresses: MHoward@SouthAlabama.edu (M.C. Howard), GutworthM@Montclair.edu (M.B. Gutworth).

[1] This definition includes VR presented via both non-immersive (e.g. monitor) and immersive (e.g. HMD) hardware. We recognize that some authors consider VR presented via immersive hardware to be “true VR” and VR presented by non-immersive hardware to be “VR 2.5D”. We opt for this broader interpretation due to its prevalence and the study of moderator variables allows us to provide separate results regarding all VR, “true VR”, and “VR 2.5D”.

https://doi.org/10.1016/j.compedu.2019.103707
Received 1 June 2019; Received in revised form 24 August 2019; Accepted 23 September 2019
Available online 26 September 2019
0360-1315/

g. creativity, elementary education) (Detrick, 2012; Hu, Wu, & Shieh, 2016). These researchers and practitioners claim that VR provides realistic and immersive experiences that allow the technology to produce a wider range of scenarios than equivalent real-life training and development programs, which is believed to result in better outcomes (Howard, 2017a, 2018; Manca, Brambilla, & Colombo, 2013). If VR can eliminate the need for other materials and/or trainers, then the technology may even be applied as a cost-saving measure.

In this broadening of training applications, VR has even been applied to develop social skills, such as public speaking (North, North, & Coble, 2015), interviewing (Tsang & Man, 2013), and interpersonal communication abilities (Morgan et al., 2014). Researchers using VR for this application regularly claim that the realistic nature of VR causes trainees to undergo similar psychological processes and perform similar behaviors as real-life scenarios (Coffey, Kamhawi, Fishwick, & Henderson, 2017; Wang, Calandra, Hibbard, & Lefaiver, 2012). For instance, participants have been shown to undergo similar psychological reactions (e.g. increased heartrate) when speaking in front of a virtual audience as they do when speaking in front of a real audience (Chen, Dey, Billinghurst, & Lindeman, 2017; Felnhofer et al., 2014; North et al., 2015). Other authors, however, have found lackluster results when applying VR to train social skills (Kampmann et al., 2016; Klaassen, de Vries, & Ceulemans, 2018; Real et al., 2017). These authors assert that VR may be able to develop these abilities one day, but present technological limitations prevent their current success. Developing social skills typically requires subtle social cues to be presented by a partner, and the trainee must properly identify and respond appropriately to these cues. Because VR cannot presently produce entirely lifelike digital environments, relevant behaviors and subsequent psychological processes may not be similar enough in VR to effectively generalize to real-world scenarios.

Due to these disparities, review and summary work is needed to further research and practice. In the current article, we perform a meta-analysis to determine the overall effectiveness of VR training programs for developing social skills, and we also study the effect of several moderating variables that may influence the effectiveness of these VR training programs. We test whether certain aspects of the utilized approach (e.g. knowledge-based, emotion-based, both, neither), applied hardware (e.g. input devices, output devices), applied software (e.g. game elements, Second Life), participant population (e.g. general/specialized), and study design (e.g. type of control group, type of measure, others) influence the success of these programs. By achieving these goals, we provide suggestions regarding the effectiveness of VR training for social skills in comparison to alternative training approaches, the attributes of VR training programs that produce better outcomes, and directions for future research and practice.

These results have many notable implications. Regarding practice, these results can identify whether VR training for social skills is presently ready for widescale applications. Many college public speaking classes already use VR for instructional purposes (Rogers, 2017; Sweigart & Hodson-Carlton, 2013; Wang et al., 2012), and the current article determines whether these applications, and others, are justified or whether they should be reconsidered until further technological improvements. Likewise, the current article identifies features of effective programs, and these features can subsequently be applied more widely and improve the effectiveness of present and future VR training programs for social skills development.

Regarding research, the current article resolves present disparities in the literature while uncovering research questions that require more investigation, and we highlight four such instances here. First, by identifying which training approach is most effective, we also identify theories that may be most effective for understanding VR training programs for social skill development. If emotion-based approaches are most effective, for instance, theories associated with emotion regulation may be key in improving VR training program outcomes (e.g. strength model, ego depletion; Baumeister, 2014; Hagger, Wood, Stiff, & Chatzisarantis, 2010). Second, we identify hardware that produces better VR social skills training outcomes, a contentious topic in the current literature (e.g. immersive vs. non-immersive displays; Cordeil et al., 2016; Cutolo et al., 2017; Qian et al., 2017); however we also provide suggestions for future research directions regarding why this hardware produces better outcomes (e.g. mediators). Third, we test whether certain types of VR software are more effective for social skills development, such as comparing gamified and non-gamified programs (Buckley & Doyle, 2016; Landers, Armstrong, & Collmus, 2017), and we use these results to identify whether future researchers should place a greater focus on certain types of programs. Fourth, we identify the aspects of study designs that notably influence outcomes. In doing so, researchers can view prior and future research through a new lens and determine whether certain results may have been driven primarily by research design. We also identify aspects of research designs that are presently underutilized, such as certain types of measures (e.g. behaviors, outcomes; Hughes et al., 2016; Landers & Armstrong, 2017), and may be detracting from our current understanding. Together, the current article answers many currently unknown questions, while highlighting new areas for investigation.

1. Background

Modern iterations of VR using computer technologies primarily emerged in the 1970s, but the greatest growth of the technology's popularity may be in 2012 with the dramatic rise of the company Oculus (Howard, 2018; Plunkett, 2014). Oculus created a crowd-funded project, the Rift, which was intended to be a cost-affordable head-mounted display (HMD) for personal VR use. Although their original crowdfunded goal was $250,000 USD, they raised over one million USD in only 24 h. Within 2 years, Oculus was then sold to Facebook for two billion USD. Today, Oculus still produces the Rift, but other companies, including Sony and HTC, have mimicked Oculus' success in producing their own cost-affordable HMDs.

Oculus ushered in a new era of VR in the 2010s, but all VR shares common core features. All VR presents a three-dimensional environment (Mirelman, Patritti, Bonato, & Deutsch, 2010; Seth et al., 2011). Although some authors discuss individuals' online life as their "virtual reality", most social media websites, discussion boards, and other similar online platforms are not VR. Instead, VR provides a three-dimensional perspective that mimics reality, such as the video games World of Warcraft or Fortnite. In these games, users can navigate around their digital world in a three-dimensional environment in a similar manner to the real world.

Further, all VR has interactive properties (Fowler, 2015; Merchant et al., 2014). Sometimes, these interactive properties are relatively basic, such as only having the capability to change the user's digital point of view (i.e. looking around). In other cases, the interactive properties are relatively advanced, such as moving in the digital environment, interacting with objects, and communicating with other people. Regardless of the level of sophistication, all VR must have some interactive properties. Without these interactive properties, the program would simply be a computer-generated video, and it would not satisfy the definition of VR.

Beyond these features, VR can be incredibly varied. Most applications of VR use avatars (Davis-Kendrick, 2015; Detrick, 2012; Hu et al., 2016). Avatars are digital beings, typically humans, which represent the user in their digital environments. The user typically takes a first-person perspective from the "eyes" of their avatar, but sometimes the user takes a third-person perspective from behind their avatar. In either case, the user "becomes" their avatar in their digital environment, which often causes the user to develop a sense of "connectedness" with their avatar. These feelings of connectedness can even predict user outcomes in VR rehabilitation, therapy, and training programs (Howard, 2017a; Vora et al., 2002).

Further, VR programs may incorporate game attributes, which are common features of games that promote user enjoyment and motivation. Many taxonomies and typologies of game elements exist, and each propose a varying number and nature of included game attributes. Among these, Bedwell, Pavlas, Heyne, Lazzara, and Salas (2012) is among the most popular, and their typology proposes 19 different types of game attributes, including conflict, progress, and surprise. Researchers can apply these typologies to determine the extent that each game element is present in a VR program as well as to determine whether a VR program is "gamified".

Beyond elements of the software, VR programs may also vary based on the hardware, which can be separated by output and input hardware (Howard, 2018). VR can be presented using the most common output hardware, computer monitors, but VR can also be presented using immersive output hardware, such as CAVEs or HMDs (Freina & Ott, 2015; Ip et al., 2018). CAVEs use multiple two-dimensional displays surrounding the user to present a pseudo three-dimensional experience, whereas HMDs are headsets that entirely cover the user's field of vision and track their head movements to present an immersive three-dimensional experience. Researchers and practitioners assert that immersive output hardware produces better outcomes than computer monitors due to increased fidelity that subsequently influences presence and motivation (Freina & Ott, 2015; Howard, 2018; Ip et al., 2018).

Alternatively, users can interact with VR environments using the most common input hardware, keyboard and mouse, but specialized hardware is also frequently used to provide user input (Bouchard et al., 2017; Zhang et al., 2018). In recent years, motion sensors have become increasingly common. These motion sensors may be video cameras with specialized software that track users' movement to register program input, such as the Microsoft Kinect, but the motion sensors may also take the form of hand-held controllers that register movements, such as the Wiimote or the HTC Vive's controller. Like immersive output hardware, researchers and practitioners assume that specialized input devices produce better outcomes than the keyboard and mouse due to increased fidelity that subsequently influences presence and motivation.

With the properties of VR reviewed, we now turn to describing VR training programs for social skill development. In doing so, we propose specific hypotheses to be tested.

1.1. VR training programs for social skills development

Social skills are a collection of KSAOs (attributes that include *k*nowledge, *s*kills, *a*bilities, and *o*ther characteristics) that facilitate interpersonal interactions (Matson, Rotatori, & Helsel, 1983; McFall 1982; Riggio, 1986). No list of social skills is comprehensive, but many social skills have been identified due to their particular importance to certain populations. For example, many authors have created interventions to develop the theory of mind and eye contact of individuals with autism, which are considered relatively basic social skills (Ozonoff & Miller, 1995; Slaughter & Gopnik, 1996). On the other hand, many authors have also created interventions to develop the negotiation abilities of managers in business, which is considered a complex social skill (Druckman & Robinson, 1998; ElShenawy, 2010). Thus, while not all social skills can be identified and labeled, many social skills with varying levels of complexity have been clearly defined and shown to be important to certain populations.

Most social skills are developed through socialization, typically from childhood to early adulthood (Matson et al., 1983; Slaughter & Gopnik, 1996). In this age range, people are expected to recognize and mimic the social abilities of others, learning the cultural social norms of their society, which leads to effective interpersonal interactions. Many complex social skills are difficult to develop, either due to their intricacies or the rarity of their occurrence (ElShenawy, 2010; McFall 1982). Most people do not have the regular opportunity to practice their negotiation skills, for example. When people have difficulties with their socialization or need to learn these complex social skills, it is common to use social skill training programs.

Social skills training programs are most often performed with real-life partners, but more authors have recently shown an interest in using VR for these purposes (Coffey et al., 2017; Klaassen et al., 2018; Rogers, 2017; Wang et al., 2012). These authors believe that VR can prompt the same psychological mechanisms and behaviors in trainees as real-life environments, and VR also allows for an increased standardization of procedures. While human partners may deviate from procedures, computer-controlled avatars can be programmed to express the same facial expressions, say the same cues, and perform the same behaviors during each session. This standardization may allow trainees to better identify social cues, especially when learning basic social skills. VR may also provide the safest social situation possible. An audience of computer-controlled avatars cannot judge the trainee, which may allow the trainee to remain calm and focus on their behaviors. While these benefits are believed to produce positive outcomes across all applications, VR training programs for social skills typically take one of four approaches.

First, these programs can provide no instruction but simply give the user a safe space to practice their social skills without embarrassment (Rogers, 2017; Xu, Park, & Baek, 2011). This approach assumes that trainees are already knowledgeable about the correct actions in social situations, but they need opportunities to perform these behaviors before becoming proficient. Because many people may not have access to role-play partners, a VR environment can provide an opportunity to practice social skills that may be unavailable to most people. For instance, some VR training programs present a virtual audience, and trainees can practice their presentation skills in front of digital people (Lane, Hays, Core, & Auerbach, 2013; Rogers, 2017; Xu et al., 2011). No information about public speaking may be provided, but practicing in front of the virtual audience may improve social skills that generalize to the real world.

Second, these programs can be knowledge-based, such that trainees are taught discrete information about certain social skills and then given a space to practice these newly-learned skills (Coffey et al., 2017; Klaassen et al., 2018; von Barnekow, Bonet-Codina, & Tost, 2017; Wang et al., 2012). This approach assumes that users are unknowledgeable about the correct actions in social situations, and they first need to be taught these actions. Then, trainees are provided a digital space to practice these newly-learned behaviors, which is believed to generalize to the real world. Sometimes when using this approach, trainees are given informational feedback based on their performance in the virtual environment, with the assumption that they are unknowledgeable of their personal behaviors in these situations (Strickland, Coles, & Southern, 2013; Tsang & Man, 2013; von Barnekow et al., 2017; Wang et al., 2012). For instance, a trainee may be asked to participate in a virtual job interview and then be given feedback afterwards based on their choice of answers and verbal tone. The trainee could then participate in subsequent virtual job interviews to improve their social skills, as they may have been previously unaware that their answers and verbal tone was inappropriate.

Third, these programs can be emotion-based, such that trainees are provided strategies to handle social anxiety and given a space to practice these strategies (Kampmann et al., 2016; Morgan et al., 2014; Pot-Kolder et al., 2018). This approach assumes that users may experience anxiety when interacting with others and teaching emotional regulation strategies can improve social skills. For instance, a trainee may be taught how to develop positive outcome expectancies and shift their focus away from negative thoughts (Bouchard et al., 2017; Yuan & Ip, 2018). The trainee could then interact with a digital person while engaging in appropriate emotional regulation strategies, which is believed to generalize to real-world situations.

Lastly, these programs can be a mix of knowledge- and emotion-based, such that trainees are provided discrete information as well as emotional-regulation strategies and given a space to practice these strategies. This approach assumes that both types of skills are required to appropriately behave in social situations.

All four approaches to VR training programs are suggested to be effective in improving social skills. In the current article, we meta-analytically test whether this suggestion is true by comparing the effectiveness of VR training programs to control groups. VR training programs almost assuredly produce some amount of improvement in social skills; however, these programs cannot be deemed effective unless they improve social skills better than relative comparisons.

Hypothesis 1. **VR training programs are more effective than control groups at improving social skills.**

We also test whether the four approaches have differing effects on social skill improvement, but we do not propose an a priori hypothesis regarding their relative effectiveness.

Research Question 1: Which VR training approach for social skills development is most effective?

1.2. VR software and hardware

Many aspects of the VR software can influence training outcomes, and perhaps the most influential aspect is whether the software includes game attributes (Bedwell et al., 2012; Nacke & Deterding, 2017). VR training programs with game attributes are believed to elicit greater trainee motivation and better align the trainees with the training program. For example, trainees may be otherwise uninterested in completing training objectives, but they are often enthusiastic in obtaining points, receiving badges, and topping leaderboards – each of which are game elements (Armstrong, Landers, & Collmus, 2015; Bedwell et al., 2012; Nacke & Deterding, 2017). While a training program without game elements may be boring and fail to capture attention, a gamified training program can often sustain trainees' attention during the entire course of the training (Burke et al., 2006.). Further, gamification elements can prompt beneficial states that have shown promise for training and learning outcomes, such as presence and moderate amounts of challenge (Landers et al., 2017). To investigate the effectiveness of gamification, Dicheva, Dichev, Agre, and Angelova (2015) qualitatively reviewed 34 studies on gamification in learning and found that game elements had a positive effect on outcomes such as engagement, attendance and participation, and performance scores. We assert that the same may be true regarding VR training for social skills, and we therefore posit the following hypothesis:

Hypothesis 2. **Gamified VR training programs are more effective than non-gamified VR training programs.**

Many software programs have been used to create and administer VR trainings, and the open-source program, Second Life, is often claimed to be the most popular (Papadopoulos, Pentzou, Louloudiadis, & Tsiatsos, 2013; Sullivan et al., 2016; Sweigart & Hodson-Carlton, 2013; Xu et al., 2011). Second Life provides a three-dimensional representation of a digital world, in which users can create almost any setting or object. Some users recreate real-world locations, such as London or Paris, whereas others create representations of fictional places, such as Middle-Earth or the U.S.S. Enterprise. Users can freely interact with these settings via their avatars, if granted permission, and thereby organizations can create environments for the benefit of others. For instance, Harvard developed a virtual campus in Second Life, where users can enroll in courses taught by Harvard professors (Kaplan & Haenlein, 2009; Zhu, Wang, & Jia, 2007). Researchers have also expressed that Second Life may be the most user-friendly program to develop three-dimensional digital environments, thereby removing certain barriers to entry for VR (Kaplan & Haenlein, 2009; Sweigart and Hodson-Carlton, 2013; Xu et al., 2011; Zhu et al., 2007).

Given these considerations, it is unsurprising that many researchers have turned to Second Life to develop VR training programs to develop social skills. Second Life provides the freedom to program detailed events, and it allows an easy approach for users to enroll in these programs. Participants with little computer experience have expressed positive reactions to installing Second Life, locating the proper locations within the digital world, and enrolling in the proper training program (Papadopoulos et al., 2013; Xu et al., 2011). Given the popularity and apparent benefits of Second Life, it is necessary to quantitatively determine whether VR training programs for social skills development are more effective when developed and administered in Second Life or an alternative software program.

Hypothesis 3. VR training programs developed and administered in Second Life are more effective than those developed and administered via other software programs.

Likewise, many aspects of the VR hardware can influence training outcomes, and we consider both the output and input hardware. Regarding output hardware, most researchers have differentiated the effects of monitors from immersive displays, including CAVEs and HMDs (Cordeil et al., 2017; Freina & Ott, 2015; Howard, 2018; Ip et al., 2018), and proposed that immersive displays produce better outcomes due to their effect on presence and motivation. Immersive displays can enable users to become deeply involved in their experiences, and users may even lose track of reality (Oh, Fiore, & Jeoung, 2007; Pine & Gilmore, 1999). That is, the user may begin to perceive their digital experiences as the "real world" and enter a state of maximum presence. When this occurs, the user is typically more motivated to complete their objectives within their digital environments, such as completing a VR training for social skills. Further, Jensen and Konradsen (2018) as well as Howard and Rose (2019) found that immersive displays are useful for certain cognitive, psychomotor, and affective skills trainings, but they also identified that more research is needed to determine exactly which trainings benefit from immersive displays. For instance, an immersive display may benefit a VR training program to develop procedural skills in engineering (e.g. building an engine), but may not benefit a VR training program to develop basic arithmetic (e.g. addition) (Howard and Rose, 2019; Jensen & Konradsen, 2018). Thus, it is important to test whether immersive displays have a positive impact on VR training programs for social skills, as it cannot be guaranteed that they produce – or do not produce – beneficial effects for any application.

Hypothesis 4. VR training programs using immersive displays are more effective than VR training programs using computer monitors.

Regarding input hardware, most researchers have differentiated the effects of the keyboard and mouse from specialized input hardware, including motion sensors (Kingston, Riddell, McKinnon, Gallagher, & Callaghan, 2016; Luu, He, Brown, Nakagome, & Contreras-Vidal, 2016; Zhang et al., 2018), and proposed that specialized input hardware produce better outcomes due to their effect on presence and motivation. In their review of immersive virtual reality, Freina and Ott (2015) found that immersive components can allow individuals to feel objects and events that are not physically present, which may increase learner motivation compared to non-immersive virtual reality. We again make the same comparison in the current article.

Hypothesis 5. VR training programs using specialized input devices are more effective than VR training programs using a keyboard and mouse.

1.3. Study design

Many aspects of the study design can also influence the observed effectiveness of VR training programs to develop social skills. We investigate six aspects in the current article.

Several types of social skills have been differentiated in prior research, and VR may be more effective at developing certain types of social skills more so than others. We broadly define three categories of social skills for comparisons in the current article, as similarly done in prior research (Kass, Burke, Blevis, & Williamson, 1994; Krasny, Williams, Provencal, & Ozonoff, 2003; Matson et al., 1983; McFall 1982; Riggio, 1986). The first is complex social skills, which involve the simultaneous application of most social competencies. For instance, holding a conversation involves the regulation of body language, verbal tone, dialogue content, and responsiveness. Alternatively, basic social skills involve the application of a single social competency. For instance, many VR training programs for individuals with autism educate users on the proper methods to select a seat on the bus, which primarily involves the application of a single social competency (awareness of social space) (Irish, 2013; Odom et al., 2015). The middle-ground between these two groups is moderate-complexity social skills, which require a subset of social competencies to perform. For instance, giving a monologue requires the regulation of body language and verbal tone, but it does not necessitate a responsiveness to others. We do not propose an a priori hypothesis regarding the differential effectiveness of VR to develop basic, moderate, or complex social skills, but we test the relative effectiveness of VR.

Research Question 2: Are VR training programs more effective at developing basic, moderate, or complex social skills?

The measurement of training outcomes can influence observed results, which is perhaps best differentiated by Kirkpatrick's training evaluation model (Kirkpatrick & Kirkpatrick, 2015, 2016). Kirkpatrick proposed that training outcomes can be measured on four levels, and training effects are stronger on the lower levels because they are more immediate than higher levels.

First, reactions are self-report indicators of training success. Trainee reactions include direct opinions about the training program, but they also include self-reports of abilities after the training. Second, learning is the demonstration of training knowledge, skills, or abilities immediately after the training program, often demonstrated through post-test evaluations. Third, behaviors are observable demonstrations of learned knowledge, skills, or abilities after the training program. These behaviors are often recorded by on-the-job observations of trainees, but they may also be recorded by observations in other environments (e.g. home, in public). Fourth, results are the effects of the training on the intended ultimate outcomes. This may be employee performance or organizational revenue in organizational training programs, whereas it may be life or relationship satisfaction for social skills training programs. We test whether the proposed magnitude of effects on these four tiers of outcomes holds when studying VR training programs.

Hypothesis 6. The effectiveness of VR training programs for social skills development varies based on the measurement method in the following order: reactions, learning, behaviors, and results (strongest to weakest).

VR training programs for social skills have been analyzed using participants from the general population, but they have also been used to analyze the development of social skills in specialized populations, such as those with autism and Down syndrome (Ip et al., 2018; Park et al., 2011; Tsang & Man, 2013). In these latter instances, the training programs are often catered to the population of interest, as it is more homogeneous than the general population. For instance, individuals with autism typically struggle with understanding theory of mind – the understanding that other people's perspectives are different from their own. Relevant VR training programs can teach this concept to improve the social skills of many people with autism; however, this social skills training would help only a small portion of the general population. Because training programs for specialized populations are typically catered to these populations, we suggest that VR training programs are more effective when applied to these populations.

Hypothesis 7. VR trainings with specialized samples are more effective at developing social skills than those using general samples.

As mentioned above, the current article analyzes the effectiveness of VR training programs for social skill development as compared to relevant control groups. These control groups may differ from study to study, which may influence the magnitude of observed effects. In the current article, we differentiate control groups using three categories: no treatment, non-equivalent treatment, and equivalent treatment. No treatment control groups do not receive any intervention whatsoever. Non-equivalent treatment groups receive an intervention that is not matched for intensity and/or time compared to the VR intervention group, and these groups typically receive an intervention that is lesser than the VR intervention. Lastly, equivalent treatment groups receive an intervention that is matched for intensity and time compared to the VR intervention group. We expect the observed effects of VR training groups to appear larger when the control group receives smaller interventions.

Hypothesis 8. The demonstrated effectiveness of VR training programs for social skills development varies based on the control group in the following order: no treatment, non-equivalent treatment, and equivalent treatment (strongest to weakest).

The effectiveness of all computer-based training (CBT) programs is expected to improve alongside technological improvements (Klimova, 2016; Rauscher et al., 2016; Spadaro et al., 2017), and the same is expected for VR training programs (Howard, 2018; Howard and Rose, 2019). We include hypotheses about objective indicators of technological sophistication, such as the applied hardware, but the publication year of the source can be considered a proxy for the technological sophistication of the intervention, as technologies become more sophisticated as time progresses (Howard, 2018; Howard and Rose, 2019; Klimova, 2016; Rauscher et al., 2016; Spadaro et al., 2017). By understanding whether the effectiveness of VR training programs is changing over time, the current article may be able to resolve conflicting empirical findings. For example, if virtual reality training programs were found to be ineffective in the 1990s and similar programs were found to be effective in the 2010s with no other changes to experimental designs, the change in outcomes may due to the technological advancements made over the past decade. By positing this hypothesis, we serve to address possible inconsistencies in the literature.

Hypothesis 9. The effectiveness of VR training programs for social skills development has a positive correlation with publication year.

Lastly, the file-drawer problem is a widely-discussed phenomenon in meta-analyses, which proposes that published results (e.g. journal articles) are typically larger than unpublished results, including dissertations and conference presentations (Pashler & Wagenmakers, 2012; Simonsohn, Nelson, & Simmons, 2014). While the magnitude of the file-drawer problem has been questioned in recent years (Dalton, Aguinis, Dalton, Bosco, & Pierce, 2012), supporting its existence for a particular topic can caution readers about research on that topic. Specifically, if a file-drawer problem is found, readers should be aware that the published results may not be a comprehensive picture of a specific effect; studies finding null effects have been conducted but not published, and the file-drawer problem can prevent researchers from gaining access to the full picture surrounding certain phenomenon. Therefore, we hypothesize the following:

Hypothesis 10. The reported effects of VR training programs for social skills development are stronger in published journal articles compared to other outlets (e.g. dissertations and conference presentations).

2. Methods

We conducted a meta-analysis to test our hypotheses presented above. Meta-analyses statistically aggregate prior results to obtain summary statistics that describe a literature, and subgroup comparisons can be made by coding relevant studies. For example, researchers could code whether studies used immersive or non-immersive displays, and analyses could be conducted to compare the two groups. Because our hypotheses we directed towards an entire domain of research, VR training programs for social skill development, we believed that a meta-analysis was most appropriate to test our hypotheses. Likewise, coding relevant studies could address our research questions about program comparisons (e.g. immersive vs. non-immersive displays). To conduct our meta-analysis, we followed the Preferred Reporting Items for Systematic Review and Meta-Analysis Protocols (PRISMA-P; Liberati et al., 2009; Moher, Liberati, Tetzlaff, & Altman, 2009; Moher et al., 2015) as well as the Meta-Analytic Reporting Standards (MARS; Kepes, McDaniel, Brannick, & Banks, 2013), placing a particular focus on the latter. We also followed the recommendation of other prominent guides and prior meta-analyses of CBT programs (Howard, 2017a, 2018; Lahti, Hätönen, & Välimäki, 2014; Lou, Abrami, & d'Apollonia, 2001; Van der Kleij, Feskens, & Eggen, 2015).

2.1. *Identifying sources*

Typical approaches were used to identify all relevant sources, as recommended by prior guides and authors (Kepes et al., 2013; Lahti et al., 2014; Liberati et al., 2009; Lou, Abrami, & d'Apollonia, 2001; Moher et al., 2009). All searches, however, were restricted to the year 2010 to the present. As mentioned above, researchers have made leaps and bounds regarding the sophistication of VR technology within the past decade. VR technology of today is drastically different than the technology of 10 or 20 years ago. Perhaps the largest growth in VR technology occurred in the year 2012, when Oculus began gaining public attention, which lead to several other companies funding their own VR R&D efforts. We believed that only including studies since 2010 would provide an accurate understanding of the effects of current VR technology, as this timespan would include research efforts both slightly before and after the widespread popularity of VR.

Searches were conducted using EBSCO and Google Scholar in October and November of 2018. EBSCO searches return results for several other databases including, Academic Source Complete, ERIC, PsycINFO and many others. This allowed the current meta-analysis to discover relevant journal articles and dissertations, and all EBSCO results were included in our initial coding database. Alternatively, Google Scholar is much more comprehensive than most other academic databases, as it catalogues more journals and conference materials. At the same time, Google Scholar is more likely to return only tangentially related search results. For instance, the search "'Virtual Reality' 'Public Speaking'" returned nine results in EBSCO but 2,620 results in Google Scholar. We only included the first 1,000 results of our Google Scholar searches in our initial coding database. Results after the first several hundred were deemed tangentially related (e.g. terms not appeared in primary text but cited single source with search terms in title), and thereby the decision to only record the first 1,000 results was supported.

Using these two databases, each of our searches included two parts. The first part included the phrase "Virtual Reality", "Computer Simulation", or "Digital Simulation" and the second part included the word or phrase "Training", "Intervention", "Therapy", "Enhance", "Promote", "Support", "Social Skills", "Communication Skills" "Social Abilities", "Communication Abilities", "Verbal Skills", "Verbal Abilities", "Public Speaking", or "Social Anxiety" (quotation marks included). Using each possible combination of these two parts in both databases (3x14x2), these 84 searches returned 13,082 sources to be initially coded.[2]

2.2. *Inclusion criteria*

Two trained coders coded each phase of the current meta-analysis to reduce concerns with researcher subjectivity (Kepes et al., 2013; Liberati et al., 2009; Moher et al., 2009, 2015). Before each phase, the two coders jointly identified coding criteria, and they developed a coding book along with coding guidelines (detailed below). The coders trained each other on the criteria, identifying relevant sources to use as poignant examples. Once the coding began, these coders initially coded the same sources until a sufficient level of inter-rater agreement was achieved (Cohen $\kappa > 0.80$) for a set of 50 articles. Then, the two coders coded the sources separately, but they conferred on any source that was unclear or confusing. Each of these training and coding decisions were based on prior guides, and they are typical in meta-analyses (Kepes et al., 2013; Liberati et al., 2009; Moher et al., 2009, 2015).

All inclusion criteria and coding results are illustrated in Fig. 1. The sources were first coded for the following criteria: (a) written in the English language, (b) studied human participants, (c) included more than nine participants,[3] (d) provided quantitative results, (e) used VR,[4] (f) tested the performance of a training program with post-training measures, and (g) included a control/comparison group. Each of these criteria were chosen due to prior precedence (Howard, 2017a, 2018; Lahti et al., 2014; Van der Kleij et al., 2015). This reduced the list of 13,082 sources to 504. Most excluded sources called for research on VR but did not study it.

Next, the two coders coded whether the source included an indicator of social skills as an outcome, again coding the same sources until a sufficient level of inter-rater agreement was met (Cohen $\kappa > 0.80$). Most often, these indicators were self-assessments of social skills (e.g. social self-efficacy), tests of social skills (e.g. judged interviews), or other-source ratings of social skills (e.g. peer assessments). We did not consider indicators of social or public speaking anxiety to be representative of social skills, but rather emotional

[2] This extremely large number of articles may seem excessive for most meta-analyses, but these searches were intended to produce a database for multiple meta-analyses on the efficacy of VR for skill development. Also, many authors did not use the term VR to describe their training programs, even if they used the technology, and many authors called their training programs VR, even if they did not use the technology. Thus, search criteria needed to be comprehensive to discover a sufficient and encompassing number of studies.

[3] Prior studies have considered this cutoff to be a proxy for pilot tests of VR training programs. Studies using only nine participants are typically initial tests of early-stage VR training programs to identify their weaknesses, and these studies do not reflect the effectiveness of completed VR training programs. Excluding these studies can provide a more accurate understanding of VR training program effectiveness.

[4] We defined VR using the definitions provided above, and the VR program must have included an interactive three-dimensional environment. Given this criterion, we did not include many "virtual patient" studies in our analyses. These excluded studies presented a non-immersive animation of a person, and users could only interact with the program by selecting preselected sentences. This does not provide any further interaction with a virtual environment than a program that plays prerecorded videos. Thus, we did not consider this to be VR.

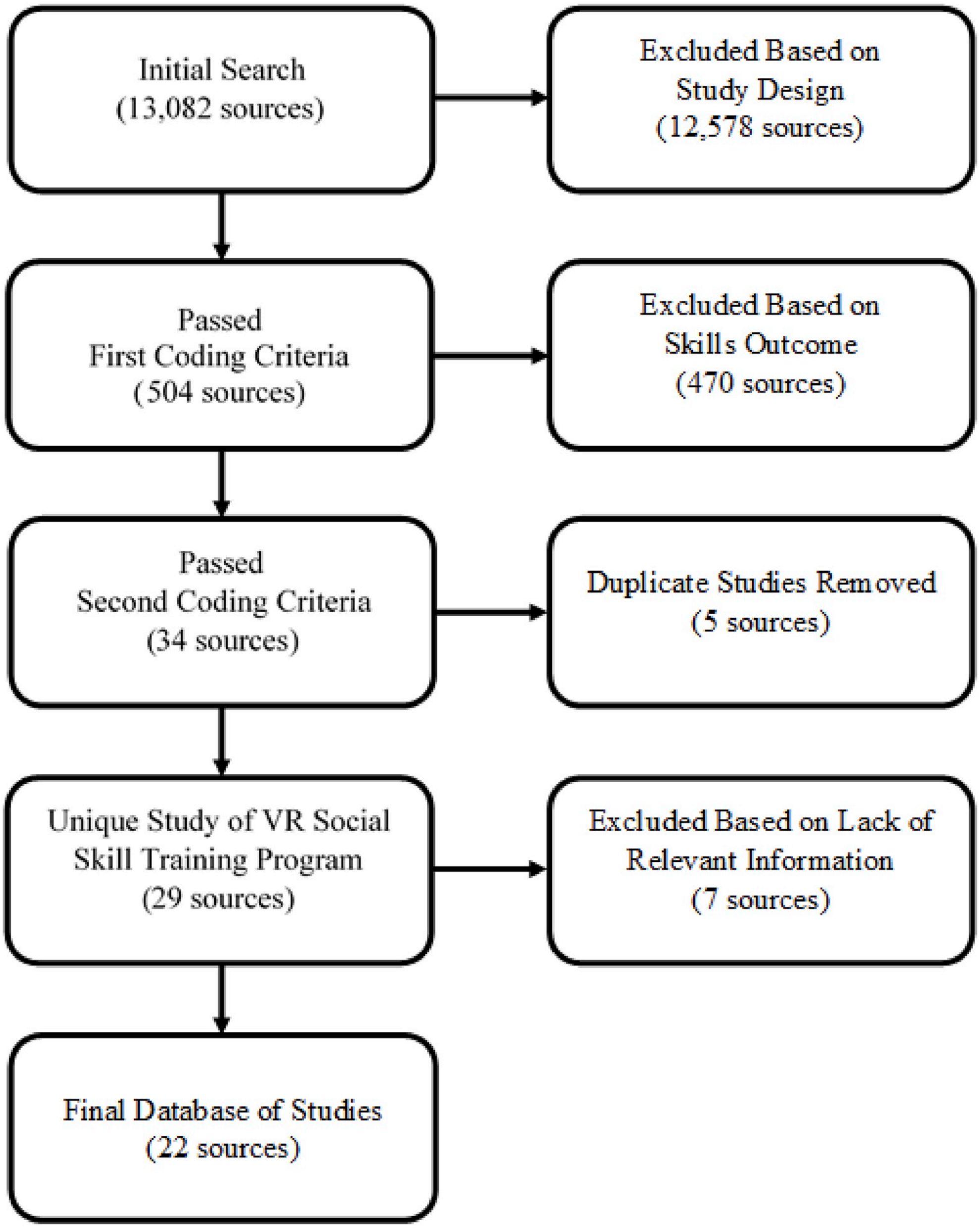

Fig. 1. Inclusion criteria Figure.

self-regulation. Although social or public speaking anxiety influences social skills, these phenomena are qualitatively different from social skills. Improvements to social or public speaking anxiety would be a beneficial outcome of a VR training program, but it does not guarantee that the program improved social skills. Also, we did not include sources that tested social outcome improvement from physical outcome training programs (e.g. increases in friendships from a treadmill training program). The dynamics of these programs are qualitatively different than social skill training programs, and thereby they were not included. This coding phase reduced the list of 504 sources to 34. Most excluded sources studied a training program for cognitive or physical abilities but not social skills.

Unlike the prior two phases, both coders coded each of the final 34 sources for the following attributes. First, effect sizes were recorded. Second, selected characteristics of the source, measures, research design, and applied training program were recorded. Each of these attributes served as possible moderating variables, and further details regarding each of these possible moderators are detailed below. At this point, five sources were removed because they were duplicates of other sources. This narrowed the number of sources to 29.

Lastly, after coding all sources, we contacted the authors of the sources for any relevant information not provided in the original source, including effect sizes and moderator attributes. At this time, we also asked authors for additional unreported data and/or results regarding VR training programs for social skills. Two follow-up emails were sent to each author. Not every author replied to our emails, resulting in six sources being excluded from our analyses. No authors provided unpublished material. Thus, the final analyses included 22 total sources, one of which provided two studies for a total of 23 total samples included in the current meta-analysis.

2.3. Moderator coding

Coding results for each source and effect are provided in Supplemental Material A.

2.3.1. Type of VR intervention

We recorded whether the VR intervention attempted to improve social skills by (a) only providing a digital space to practice social interactions, (b) providing instruction regarding social interactions, (c) providing instruction regarding methods to regulate emotions and social anxiety, or (d) providing instruction regarding both social interactions and regulating emotions. It is assumed that most sources categorized as (b) and (c) would also provide a digital space to practice social interactions. Therefore, articles were only coded as (a) if they did not also provide instruction regarding social interactions or regulation.

2.3.2. Game elements

Because most authors did not provide sufficient details to determine whether specific game elements were present, we instead coded whether the VR intervention was "gamified or "not gamified". We coded the VR intervention as "gamified" if (a) the authors directly stated it was a game or game-like training or (b) it appeared to include multiple game elements (gamified) as detailed by the author(s). We coded the VR intervention as "not gamified" if (a) the authors did not label it as a game or game-like training or (b) it included none or very few game elements beyond those required for VR programs (e.g. environment).

2.3.3. Type of display hardware

We recorded whether the VR intervention group used a computer monitor or an immersive display technology (e.g. surround-screen display, HMD). If the authors did not directly state the display hardware, it was assumed that a monitor was used.

2.3.4. Type of input hardware

We recorded whether the VR intervention group used a keyboard, mouse, and/or joystick or a specialized input technology (e.g. motion capture). If the authors did not state the input hardware, it was assumed that a keyboard, mouse, and/or joystick was used.

2.3.5. Complexity of outcome

We recorded whether the developed social skill was basic, moderate complexity, or complex. We defined basic social skills as those requiring a single social competency to perform (e.g. recognizing emotions), whereas a complex social skill requires most types of social competencies (e.g. holding an entire conversation). Moderate complexity social skills require a defined subset of competencies (e.g. giving a monologue).

2.3.6. Type of outcome (Kirkpatrick's hierarchy)

We recorded the type of outcome using Kirkpatrick's training evaluation model (Kirkpatrick & Kirkpatrick, 2015, 2016), which includes four categories: reactions, learning, behavior, and results.

2.3.7. Type of participants

We recorded whether the participants were from general or specialized populations. General populations included typical adults and children. All other populations were specialized populations, which included those with autism, aphasia, anxiety, and others.

2.3.8. Type of control group

We recorded whether the included control group was a "no-treatment", "non-equivalent treatment", or "equivalent treatment" group. A no-treatment control group did not receive any intervention. A non-equivalent treatment control group received an intervention, but it was not matched for the duration or intensity of the intervention received by the VR intervention group. An equivalent treatment group received an intervention, and it was matched for the duration and intensity of the intervention received by the VR intervention group.

2.3.9. Type of source

We recorded whether the source was a journal article or an unpublished source (conference presentation, thesis, or dissertation).

2.3.10. Publication year

We recorded the year the article was published, the presentation was presented, or the thesis/dissertation was completed.

2.4. Analyses

We began our analyses by calculating indices of publication bias. These include fail-safe k, Egger's test, the trim-and-fill methods, and weight-function model analyses. The fail-safe k provides an estimate of the number of undiscovered null studies that would need to be discovered before the observed effect would no longer be statistically significant (Van Wijk, Jansen, & Lyles, 2008; Zhao, Wayne, Glibkowski, & Bravo, 2007). If the fail-safe k is sufficiently large, it can be assumed that the results would not be greatly swayed by the discovery of more studies. Egger's test provides an estimate of the relationship between the observed effect sizes and the standard errors (Debray, Moons, & Riley, 2018; Kepes, Banks, McDaniel, & Whetzel, 2012). If a significant relationship exists, then it can be assumed that the included small-sample studies have larger effect sizes than the large-sample studies, and notable publication biases may exist in the dataset. The trim-and-fill method provides an estimate of how many studies are missing from the left and right of the meta-analytic effect, based on effect sizes and standard errors (Carpenter, 2012; Duval & Tweedie, 2000). This test can estimate the extent that the observed effects are over- or under-estimated. Lastly, the weight-function model analysis determines whether certain p-value intervals were significantly more or less likely to be observed in the dataset (Dear & Begg, 1992; Vevea & Woods, 2005). Certain intervals are typically associated with publication biases, which are tested in the current article.

We then tested for outliers and influential cases. We primarily considered three indicators in identifying these cases: studentized deleted residuals, Cook's distance, and covariance ratios (Cousineau & Chartier, 2010; Viechtbauer & Cheung, 2010). These statistics provide a numerical indicator for each study, which is then plotted on a line chart. Values that are notably higher than the others are considered outliers and/or influential cases. These charts are provided in Supplemental Material B, but they are also summarized below.

To calculate the primary analyses, we used a random-effects approach in Comprehensive Meta-Analysis V3. Random-effects meta-analyses are more resilient to outlier studies than fixed-effects approaches, and random-effects approaches provide more accurate results overall (Hedges & Vevea, 1998; Hunter & Schmidt, 2000). We reported all primary results as Cohen's *d*, which is an estimate of the mean difference between two groups (e.g. VR treatment vs. control) relative to their standard deviations. For example, a Cohen's *d* of one indicates that the means of the two groups were one standard deviation apart. We did not include any effect sizes that represent the relationship of more than two variables (e.g. ANCOVA) due to their statistical concerns (Anderson et al., 2010; Boxer,

Table 1
Publication bias results.

	I^2	k	Fail Safe k	Egger's test β_0	Egger's Test t	Implied Missing	
						Left of Mean(μ)	Right of Mean(μ)
1.) Overall Effectives of VR vs. Control	93.158	23	719	2.696	1.501	0	6
Type of Intervention							
2a.) Basic	95.529	4	34	-.715	.068	0	1
2b.) Knowledge-Based	93.917	15	403	3.252	1.463	0	4
2c.) Emotion-Based	0	3	0	-.636	.445	0	0
2d.) Both	–	1	–	–	–	–	–
Game Elements							
3a.) Game Elements Excluded	94.551	17	496	3.782	1.539	0	5
3b.) Game Elements Included	80.328	6	18	.933	.339	0	2
Second Life							
4a.) Did not Use Second Life	90.788	16	122	3.005	1.829	0	4
4b.) Used Second Life	95.301	7	233	5.960	1.007	0	2
Type of Display Hardware							
5a.) Computer Monitor	91.075	15	500	.895	.431	0	4
5b.) Immersive Display Hardware	93.800	8	13	4.528	1.412	0	3
Type of Input Hardware							
6a.) Basic Input	93.922	20	679	2.853	1.473	0	5
6b.) Specialized Input	0	3	0	.203	.107	0	0
Complexity of Outcome							
7a.) Basic Outcome	79.975	9	30	.407	.235	0	3
7b.) Moderate Outcome	92.826	7	43	.876	.131	0	2
7c.) Complex Outcome	94.073	13	266	4.420	1.560	0	4
Type of Outcome							
8a.) Reactions	92.586	8	73	3.797	.841	0	3
8b.) Learning	90.982	15	421	-.053	.025	0	3
8c.) Behavior	–	2	–	–	–	–	–
8d.) Results	31.847	3	0	1.670	1.148	2	0
Type of Participants							
9a.) General	96.339	12	446	5.864	1.814	0	1
9b.) Specialized	38.411	11	24	.728	.619	0	2
Type of Control Group							
10a.) No Treatment	91.499	9	165	2.330	.837	0	2
10b.) Non-Equivalent Treatment	60.545	5	0	.599	.363	0	0
10c.) Equivalent Treatment	93.723	11	224	.291	.073	0	2
Type of Source							
11a.) Journal Article	93.796	20	773	3.788	1.937	0	4
11b.) Thesis, Dissertation, or Conference Presentation	39.308	3	0	.047	.009	0	0

Groves, & Docherty, 2015; Rothstein & Bushman, 2015). We also did *not* correct for unreliability, due to precedence in prior meta-analyses of training and education programs (Griffeth, Horn, & Gaertner, 2000; Harrison, Newman, & Roth, 2006; Jones, Woods, & Guillaume, 2016; Zimmerman, 2008). Authors of these studies note that meta-analytic results should not be corrected for unreliability when the results serve a practical purpose, such as identifying the most effective training or educational program. Because all the studies included in the current meta-analysis compared a VR training program against an alternative training program, and thereby sought to discover the most effective training or educational program, it would be inappropriate to correct for unreliability in the current meta-analysis.

To test moderating effects, we calculated meta-analytic estimates separated by each level of the moderators and compared confidence intervals. We also calculated meta-regressions for each moderator, using the moderator as the sole predictor of VR training performance in each meta-regression. If the associated coefficient is statistically significant, then the moderator can be considered to significantly effect VR training program efficacy (Thompson & Higgins, 2002; Van Houwelingen, Arends, & Stijnen, 2002). We ran separate analyses because our number of sources was relatively small for meta-regressions, preventing the simultaneous analysis of multiple predictors. We also interpret marginally significant results ($0.10 > p > .05$) as supportive of hypotheses, again due to the relatively small number of sources for meta-regression.

Lastly, many authors recommend sensitivity analyses, which is the replication of meta-analytic results using varied approaches. Supplemental Material B includes a reanalysis of all primary analyses with all possible outliers and influential cases removed, whereas Supplemental Material C includes a reanalysis using a three-level meta-analytic approach. This approach identifies sources of dependency within and across studies, allowing it to address multiple effect sizes from a single study without needing to average them together (Cheung, 2015; Jak, 2015). None of the inferences below were altered when performing either of these analyses, and they are therefore provided in the supplementary material but not discussed in the primary text.

3. Results

3.1. Publication bias and outlier analyses

We began with investigating indices of publication bias, which are presented in Table 1. The overall fail-safe k was 719, indicating that 719 unreported null effects would need to be discovered for the overall result to become non-significant. When analyzing the fail-safe k values for the multiple levels of each moderator, no fail-safe k was smaller than 120 for a significant effect other than the effect of VR for specialized participants. Because these values are sufficiently large based on prior guidelines (Hunter & Schmidt, 2000; Van Wijk et al., 2008; Zhao et al., 2007), the observed significant effects are considered robust.

Next, the results of Egger's test did not suggest the presence of publication bias in the overall analyses or any subgroup analysis, but the trim-and-fill methods did indicate the presence of publication bias in the overall analyses as well as several subgroup analyses. Fig. 2 presents a depiction of the effect sizes. As apparent from the figure, multiple studies with positive and large effects also had small sample sizes, indicating that some reported results may have been "cherry picked". That is, the authors may have collected multiple small samples but only chose to report their significant findings. The weight function analysis was statistically significant for the overall effect ($\chi^2 = 6.864$, df = 1, p = .009), further supporting this finding.

To probe this possible publication bias, we tested for outliers and influential cases. Studentized deleted residuals, Cook's distance

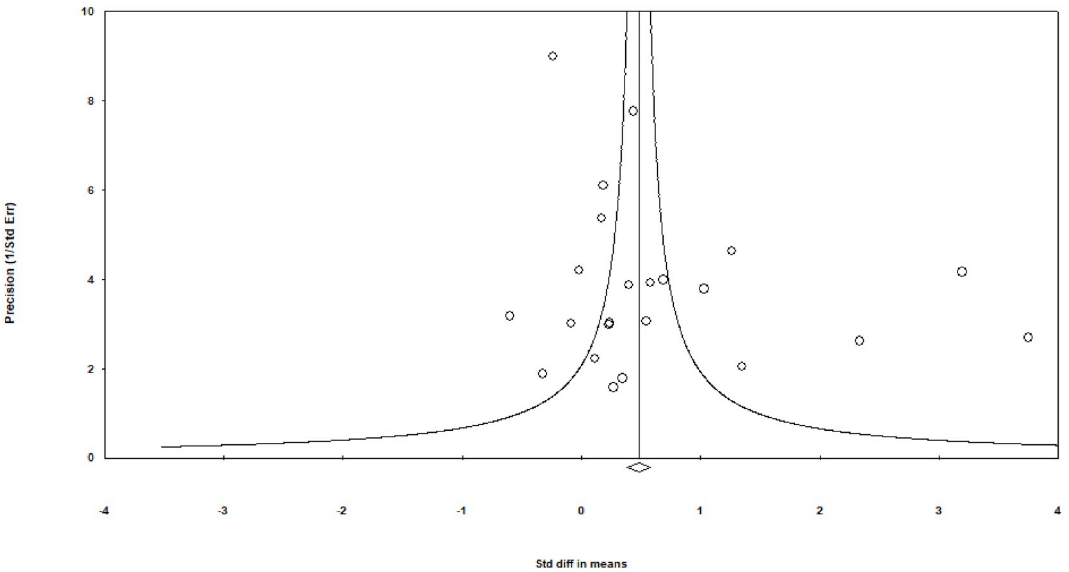

Fig. 2. Funnel plot of precision by standard difference in means.

covariance ratio values each indicated that two outliers may be present (Coffey et al., 2017; Heuett & Heuett, 2011). We chose to include these two sources in our primary reporting for three reasons: (a) neither of these studies were extreme outliers or influential cases, (b) random effects approaches in meta-analyses are resilient to outliers and influential cases, and (c) the number of included studies was relatively modest and these sources may only appear to be outliers due to a small number of comparison sources. Nevertheless, Supplemental Material B includes the current results recalculated without these two studies. When interpreting these recalculated results, none of the statistical significances or inferences changed, indicating that the current results are not driven by one or two sources.

Lastly, the I^2 values were large for the overall analysis and most subgroup analyses (>75). This indicates additional moderator variables are yet to be discovered and additional future directions can be pursued beyond the current meta-analysis.

3.2. Primary analyses

The results of our primary analyses are provided in Table 2. Overall, VR training programs were more effective than comparison groups for developing social skills, with a mean that was three-fourths of a standard deviation higher than the comparison group on average ($d = 0.691$, 95% CI [0.297, 1.085], $z = 3.435$, $p = .001$). Hypothesis 1 was supported.

Research Question 1 asked which approach to VR training programs was most effective: no instruction, knowledge-based instruction, emotion-based instruction, or both. Of the four categories, only one reached statistical significance: knowledge-based instruction ($d = 0.762$, 95% CI [0.275, 1.249], $z = 3.068$, $p = .002$). No instruction ($d = 1.046$, 95% CI [-0.727, 2.819], $z = 1.156$, $p = .247$), emotion-based instruction ($d = 0.133$, 95% CI [-0.153, 0.419], $z = 0.910$, $p = .363$), and both ($d = -0.019$, 95% CI [-0.485, 0.448], $z = -0.078$, $p = .938$) failed to reach significance. Four dummy-coded meta-regressions were tested, each using a different category as the reference group. No comparison was statistically significant, and the dummy codes failed to produce a significant overall model effect. Nevertheless, the effect sizes for no instruction and knowledge-based instruction were notably stronger than emotion-based instruction and both types of instruction together, providing implications for Research Question 1.

Hypothesis 2 proposed that VR training programs with game elements are more effective than those without game elements. Programs with game elements ($d = 0.531$, 95% CI [-0.096, 1.159], $z = 1.660$, $p = .097$) were marginally significant, whereas programs

Table 2
Primary meta-analytic results.

	# of Sources	k	N	d	95% C.I.	z-value	Sig
1.) Overall Effectives of VR vs. Control	22	23	1,530	.691	.297, 1.085	3.435	**.001**
Type of Intervention							
2a.) Basic	4	4	179	1.046	-.727, 2.819	1.156	.247
2b.) Knowledge-Based	14	15	1,088	.762	.275, 1.249	3.068	**.002**
2c.) Emotion-Based	3	3	191	.133	-.153, .419	.910	.363
2d.) Both	1	1	72	-.019	-.485, .448	-.078	.938
Game Elements							
3a.) Game Elements Excluded	17	17	1,311	.726	.258, 1.193	3.041	**.002**
3b.) Game Elements Included	5	6	256	.531	-.096, 1.159	1.660	.097
Second Life							
4a.) Did not Use Second Life	15	16	678	.551	.080, 1.022	2.292	**.022**
4b.) Used Second Life	7	7	853	.980	.271, 1.689	2.710	**.007**
Type of Display Hardware							
5a.) Computer Monitor	14	15	1,079	.845	.375, 1.314	3.526	**<.001**
5b.) Immersive Display Hardware	8	8	451	.422	-.229, 1.072	1.271	.204
Type of Input Hardware							
6a.) Basic Input	19	20	1,304	.780	.326, 1.234	3.366	**.001**
6b.) Specialized Input	3	3	226	.121	-.142, .383	.899	.369
Complexity of Outcome							
7a.) Basic Outcome	8	9	632	.395	-.015, .806	1.888	.059
7b.) Moderate Outcome	7	7	319	.667	-.269, 1.603	1.397	.163
7c.) Complex Outcome	13	13	935	.678	.167, 1.189	2.601	**.009**
Type of Outcome							
8a.) Reactions	8	8	549	.680	-.019, 1.380	1.906	.057
8b.) Learning	14	15	949	.787	.290, 1.284	3.105	**.002**
8c.) Behavior	2	2	188	-.028	-.571, .515	-.101	.920
8d.) Results	3	3	136	-.090	-.352, .173	-.670	.503
Type of Participants							
9a.) General	12	12	1038	.979	.335, 1.622	2.982	**.003**
9b.) Specialized	10	11	493	.333	.087, .578	2.655	**.008**
Type of Control Group							
10a.) No Treatment	9	9	654	.882	.263, 1.501	2.793	**.005**
10b.) Non-Equivalent Treatment	5	5	214	-.072	-.414, .269	-.416	.678
10c.) Equivalent Treatment	10	11	738	.773	.118, 1.447	2.250	**.024**
Type of Source							
11a.) Journal Article	19	20	1,435	.821	.395, 1.247	3.777	**<.001**
11b.) Thesis, Dissertation, or Conference Presentation	3	3	96	-.224	-.772, .325	-.799	.424

without game elements ($d = 0.726$, 95% CI [0.258, 1.193], $z = 3.041$, $p = .002$) reached statistical significance. A meta-regression did not find a significant difference between the two groups ($B_1 = -0.257$, $S.E. = 0.490$, 95% CI [-1.218, 0.705], $z = -0.523$, $p = .601$), and the confidence intervals entirely overlapped. Hypothesis 2 was not supported.

Hypothesis 3 proposed that VR training programs developed in Second Life are more effective than those developed via other programs. Programs developed in Second Life reached statistical significance ($d = 0.980$, 95% CI [0.271, 1.689], $z = 2.710$, $p = .007$) as did those developed using other programs ($d = 0.551$, 95% CI [0.080, 1.022], $z = 2.292$, $p = .022$). A meta-regression did not find a significant difference between the two groups ($B_1 = 0.429$, $S.E. = 0.430$, 95% CI [-0.413, 1.271], $z = 0.999$, $p = .318$), and their confidence intervals almost entirely overlapped. Hypothesis 3 was not supported.

Hypothesis 4 predicted that VR programs using immersive displays are more effective than those using monitors. Programs using immersive displays failed to reach statistical significance ($d = 0.422$, 95% CI [-0.229, 1.072], $z = 1.271$, $p = .204$), whereas those using computer monitors were statistically significant ($d = 0.845$, 95% CI [0.375, 1.314], $z = 3.526$, $p < .001$). A meta-regression did not find a significant difference between the groups ($B_1 = -0.424$, $S.E. = 0.405$, 95% CI [-1.218, 0.370], $z = -1.046$, $p = .296$). The confidence intervals of the two groups were notably distinct, but in an opposite direction than predicted. Hypothesis 4 was not supported.

Hypothesis 5 predicted that VR programs using specialized input devices are more effective than those using basic input devices (keyboard, mouse, joystick). Programs using specialized input devices failed to reach statistical significance ($d = 0.121$, 95% CI [-0.142, 0.383], $z = 0.899$, $p = .369$), whereas those using basic input devices were statistically significant ($d = 0.780$, 95% CI [0.326, 1.234], $z = 3.366$, $p = .001$). A meta-regression failed to find a significant difference between the two groups ($B_1 = -0.652$, $S.E. = 0.596$, 95% CI [-1.821, 0.517], $z = -1.093$, $p = .274$). The confidence intervals of the two groups were almost entirely distinct, but again in an opposite direction of the predicted effect. Hypothesis 5 was not supported.

Research Question 2 questioned whether VR training programs were more effective at developing basic or complex social skills. Programs applied to develop basic skills had the smallest observed effect and was marginally significant ($d = .395$, 95% CI [-0.015, 0.806], $z = 1.888$, $p = .059$). Programs applied to develop moderate complexity outcomes had a larger effect but was not statistically significant ($d = .667$, 95% CI [-0.269, 1.603], $z = 1.397$, $p = .163$). Programs applied to develop complex outcomes also had a larger effect and was statistically significant ($d = .678$, 95% CI [0.167, 1.189], $z = 2.601$, $p = .009$). Three dummy-coded meta-regressions did not find any statistically significant comparisons among these groups.

Hypothesis 6 predicted that the observed effect of VR training programs depends on the type of measured outcome, and the strength of the observed relationships would range from strongest to weakest in the following order: reactions, learning, behavior, and results. The effect for reactions approached statistical significance ($d = 0.680$, 95% CI [-0.019, 1.380], $z = 1.906$, $p = .057$); the effect for learning was statistically significant ($d = 0.787$, 95% CI [0.290, 1.284], $z = 3.105$, $p = .002$), the effect for behavior was not statistically significant ($d = -0.028$, 95% CI [-0.571, 0.515], $z = -0.101$, $p = .920$), and the effect for results was not statistically significant ($d = -0.090$, 95% CI [-0.352, 0.173], $z = -0.670$, $p = .503$). Three dummy-coded meta-regressions were tested. No comparison was statistically significant, and the dummy codes failed to produce a significant overall model effect. Nevertheless, the effect sizes for reactions and learning were notably larger than behavior and results, providing partial support Hypothesis 6.

Hypothesis 7 predicted that VR programs for specialized populations are more effective than those for general populations. VR programs for specialized populations reached statistical significance ($d = 0.333$, 95% CI [0.087, 0.578], $z = 2.655$, $p = .008$) as did VR programs for general populations ($d = 0.979$, 95% CI [0.335, 1.622], $z = 2.982$, $p = .003$). A meta-regression did not find a significant difference between the two groups ($B_1 = -0.616$, $S.E. = 0.413$, 95% CI [-1.426, 0.195], $z = -1.489$, $p = .136$). Although the confidence intervals were notably distinct, they were in the opposite direction than predicted. Hypothesis 7 was not supported.

Hypothesis 8 predicted that the observed effect of VR training programs would depend on the comparison group, ranging from strongest to weakest in the following order: no treatment, non-equivalent treatment, and equivalent treatment. VR programs compared to no treatment ($d = .882$, 95% CI [0.263, 1.501], $z = 2.793$, $p = .005$) and equivalent treatment ($d = 0.773$, 95% CI [0.118, 1.447], $z = 2.250$, $p = .024$) comparisons reached statistical significance, whereas VR programs compared to non-equivalent treatments did not reach statistical significance ($d = -0.072$, 95% CI [-0.414, 0.269], $z = -0.416$, $p = .678$). A series of three meta-regression produced a significant coefficient for the comparison of equivalent and non-equivalent treatment groups ($B_1 = 1.109$, $S.E. = 0.532$, 95% CI [0.066, 2.153], $z = -2.084$, $p = .037$) as well as a marginally significant coefficient for the comparison of no treatment and non-equivalent treatment groups ($B_1 = 1.077$, $S.E. = 0.558$, 95% CI [-0.016, 2.170], $z = -1.930$, $p = .054$). The overall model effect was marginally significant ($Q = 5.042$, $df = 2$, $p = .080$). Although these analyses produced many significant effects, they were not in the predicted order. Hypothesis 8 was not supported.

Hypothesis 9 predicted that the effects for VR training programs would have a positive relationship with year. The coefficient in a meta-regression very closely approached statistical significance ($B_1 = -0.154$, $S.E. = 0.079$, 95% CI [-0.308, 0.001], $z = -1.952$, $p = .051$), but it was yet again in the opposite direction than predicted. To make a dichotomous comparison, we separated sources by whether they were published before ($n = 12$) or after ($n = 11$) 2015. No sources were published in 2015. Sources published before 2015 had a statistically significant effect ($d = 1.022$, 95% CI [0.569, 1.474], $z = 4.424$, $p < .001$), whereas those published after 2015 did not have a significant effect ($d = 0.290$, 95% CI [-0.392, 0.973], $z = 0.834$, $p = .405$). Hypothesis 9 was not supported, due to the direction of the marginally significant results.

Hypothesis 10 predicted that reported results of VR training programs would be larger in published journal articles compared to theses, dissertations, and conference presentations. VR programs reported in journal articles were statistically significant ($d = 0.821$, 95% CI [0.395, 1.247], $z = 3.777$, $p < .001$), whereas VR programs reported in theses, dissertations, and conference presentations were not statistically significant ($d = -0.224$, 95% CI [-0.772, 0.325], $z = -0.799$, $p = .424$). A meta-regression found a marginally significant effect ($B_1 = -1.048$, $S.E. = 0.607$, 95% CI [-2.238, 0.142], $z = -1.726$, $p = .084$), and the confidence intervals did not overlap whatsoever. Hypothesis 10 was supported.

4. Discussion

A goal of the current article was to test the efficacy of VR training programs for social skill development. Our meta-analytic results showed that VR training programs were more effective overall than comparisons by almost three-fourths of a standard deviation. These results provide large support for the continued used of VR programs to develop social skills; however, the following results draw into question almost all current knowledge on this application of VR.

Of the four types of VR training programs, only those providing discrete information to improve social knowledge reached statistical significance. Programs that provided a practice space alone had a large observed effect, but they also had much variation in their effects that caused confidence intervals to include zero. Programs that provided instruction on emotional regulation as well as those that provided information to improve both knowledge and emotional regulation produced an extremely small and non-significant effect. Based on these results, only knowledge-based programs can be recognized to be effective. This finding has widespread implications for the continued creation of VR training programs for social skill development.

Practically, it appears appropriate to continue administering knowledge-based VR programs; programs that simply provide a practice space may be effective with further research and development; but much doubt is now cast on the ability of emotion-based programs to improve social skills. Theoretically, perspectives associated with the incorporation of new knowledge may be most relevant for the continued study and development of VR programs for social skills developments, which includes most extant perspectives on education (Harasim, 2017; Illeris, 2018). Research on coping may too prove to be useful, as this research has dichotomized coping strategies as problem-focused coping and emotion-focused coping (Baker & Berenbaum, 2007; Carver, Scheier, & Weintraub, 1989). Problem-focused coping refers to strategies aimed at discovering solutions to address a stressor, whereas emotion-focused coping refers to strategies aimed at discovering approaches to address negative emotions associated with a stressor. While each are more effective in certain situations, it is possible that problem-focused coping may be the best approach to improving social skills when undergoing a VR training. Learning social skills can be considered a stressor, and knowledge-based VR programs may be most effective because they help users utilize problem-focused coping strategies. Thus, prior research on problem-focused coping may be ideal to integrate with our understanding of VR.

The current article tested two aspects of the software. Gamified VR training programs were just as effective as VR training programs with few or no game elements, casting doubt on the effectiveness of gamification for the development of social skills. This finding calls into question much prior research on the benefits of gamification, and it also supports recent research suggesting that the benefits of game elements are overstated (Ferrara, 2013; Hamari, Koivisto, & Sarsa, 2014; Lister, West, Cannon, Sax, & Brodegard, 2014; Sardi, Idri, & Fernández-Alemán, 2017). We agree with this research and the other recent suggestions that (a) game elements should be cautiously incorporated into programs and (b) authors should test the effects of individual game elements rather than omnibus program comparisons. Also, trainings created in Second Life produced larger effects than those created in other programs, but this difference was not statistically significant. It is easier to program a training in Second Life than most other software, however, and thereby the non-significant difference between Second Life and other programs supports the continued use of Second Life. While other programs should also be utilized when needed, we see no reason to discontinue the use of Second Life.

The current article also tested two aspects of the hardware. Training programs using immersive displays were less effective than programs using monitors, although this difference was not statistically significant. Training programs using specialized input devices produced much smaller effects than those using a keyboard, mouse, and/or joystick, but this effect was again not statistically significant. These results suggest that advanced hardware devices do not improve the efficacy of VR training programs for social skill development, whether they are output or input devices. This finding runs contrary to one of the most widespread notions in human-computer interaction and human factors research: improvements to technologies lead to improved outcomes (Card, 2017; Preece, Rogers, & Sharp, 2015). We provide three possible justifications for this finding. First, advanced hardware could indeed be worse than traditional hardware for VR training programs to develop social skills, and the applications of these technologies for this purpose may result in task-technology misfit (Howard & Rose, 2019). Second, these programs may be developed with traditional hardware in mind, and researchers may be applying the inappropriate hardware for the chosen software. New programs may need to be developed specifically for the advanced hardware to improve observed effects. Third, users may not be acclimated to the advanced hardware, and there may be seductive details that distract trainees from the purposes of the program (Garner, Gillingham, & White, 1989; Howard, 2017b). If true, then these effects should fade over time as these technologies become more popular. Research is needed on these technologies, however, for these justifications to be supported.

The current article also supported that the study design has just as large – if not larger – of an effect on observed results than the training program itself. VR training programs with the intention to develop basic social skills were marginally significant; those to develop moderate-complexity social skills were not significant; and those to develop complex social skills were statistically significant. Researchers should not expect to obtain similar magnitudes of results across all applications of VR training programs to develop social skills, and they should interpret their results with these outcomes differences in mind. Regarding the measurement of outcomes, studies measuring reactions and learning produced much greater effects than those measuring behaviors and results, although comparisons of these measurement approaches were not significant. Researchers should nevertheless recognize differences in the magnitude of results based on the type of measure being applied and conceptually account for these differences.

Contrary to expectations, studies utilizing participants from specialized populations produce smaller effects than those from general populations, likely because these participants have additional learning difficulties. Researchers should therefore be cognizant that the chosen sample also influences results. Because these effects are smaller, larger sample sizes should be collected with specialized populations to ensure sufficient statistical power. Many studies on VR include sample sizes below 64 total participants (Bouchard et al., 2017; Marshall et al., 2016; Strickland et al., 2013), which is the necessary sample size to obtain power of .80

assuming moderate effects ($d = 0.50$; t-test). Future research should strive to collect sample sizes beyond this 64-participant cutoff, especially when using populations known to produce small effects.

Further, the use of non-equivalent control groups produced smaller effects than no treatment or equivalent control groups. While also contrary to expectations, these results can be justified. Typically, studies that utilized non-equivalent control groups administered a standard training program to both groups, and then an additional VR training program was administered to one of the groups. Because the VR group was, on average, no more effective than the non-equivalent control group, these results suggest that VR training programs may not provide benefits when applied in conjunction with other training programs. Alternatively, VR training programs were more effective than equivalent control groups, and, in these studies, the authors compared a VR training against a comparable non-VR training. This result suggests that VR training programs are more effective than other types of training programs for developing social skills. Thus, VR may be a potential replacement for alternative training types, but researchers and practitioners should be hesitant to use VR in conjunction with alternative training types.

The efficacy of VR training programs for social skills development had a relationship with the publication year; however, this relationship was negative, and the efficacy of these training programs is getting smaller over time. This finding suggests that modern researchers are pushing the possibilities of VR and applying the technology to develop a wider variety of social skills than earlier research. Unfortunately, the social skills targeted in these more recent articles are not as effectively developed via VR than those targeted in earlier articles. It should be recognized that, when testing these boundaries of VR, researchers should not expect large effect sizes. Instead, VR may only produce small to moderate effects in improving social skills that have been previously untested. For this reason, researchers should ensure that their sample sizes are large enough to detect significant effects, much like in the study of specialized populations. Researchers should also not ignore improvements that VR can provide towards previously untested social skills – even if these improvements are incremental.

Lastly, the results of VR training programs for social skills development were larger in journal articles compared to those reported in conference publications, theses, and dissertations, thereby indicating that unsuccessful attempts to develop these training programs are less likely to be published. This indicates that a file-drawer problem may exist in the study of VR training programs for social skills. Readers should inspect the current literature with this limitation in mind, and they should be aware that VR training programs, overall, may produce smaller effects for developing social skills than commonly assumed. Likewise, future researchers should not be surprised by small or even null effects when studying VR training programs.

4.1. Future research directions

Given these results and implications, many directions for future research should be considered. First and foremost, researchers need to develop and test the effects of additional VR training programs for social skills development. While the number of studies included in the current meta-analysis is adequate, there are much fewer tests of VR training programs to develop social skills compared to those to develop cognitive, emotional, or physical skills (Howard, 2018; Irish, 2013; Jensen & Konradsen, 2018). Due to this dearth of research, great uncertainty still exists regarding the ideal properties of VR training programs to develop social skills, which are discussed further below. Likewise, few results have been sufficiently replicated. While we showed that VR training programs are effective for social skill development, overall, few authors have repeatedly supported that a specific program is effective. Thus, few VR training programs can be reliably administered to any population to improve social skills, but researchers can cite the current results to justify new applications of the technology for this purpose.

VR training programs that provided instruction on emotional regulation abilities showed a very small effect on improving social skills. Prior meta-analyses suggested that these programs may be efficient in reducing participants' social anxiety (Anderson et al., 2013; Parsons & Rizzo, 2008; Powers & Emmelkamp, 2008), but the current meta-analysis is the first to show that these programs may not actually improve social skills. Authors should further probe this discovery and question the common assumption that reductions to social anxiety necessarily relate to social skill improvement. Currently, very few authors measure social skills when testing VR training programs that provide instruction on emotional regulation while interacting with others, and instead these authors solely measure indicators of social anxiety. We call on future research to change this trend, and researchers should measure social outcomes when testing these programs.

As mentioned above, theory associated with problem-focused coping may be particularly valuable to integrate with future research on knowledge-based VR training programs. We specifically urge future researchers to integrate recent work investigating populations most likely to employ certain coping strategies (Afshar et al., 2015; Dardas & Ahmad, 2015). These populations may particularly benefit from a VR training program that induced coping strategies that they do not commonly apply, and therefore those who often engage in emotion-focused coping may benefit the most from knowledge-based VR training programs.

VR also appears to be most ideal for developing more complex social skills, and future research should apply theories associated with complex outcomes to understand the relative effectiveness of VR training programs. For instance, experiential learning theories (Kolb, 2014; Kolb & Kolb, 2012; Moon, 2013) suggest that the most effective approach to learning is to perform the skills and abilities of interest, and it is often applied to understand the development of complex skills. These theories propose specific features that produce greater learning outcomes when performing skills (Kolb, 2014; Kolb & Kolb, 2012), which should be adapted in future research. Recent authors have expanded the complexity of these theories (Huang, Chen, & Chou, 2016; McCarthy, 2016; Seaman, Brown, & Quay, 2017), and these modern additions may be valuable in better understanding VR training programs. Seaman et al. (2017) identify unique and prominent future research directions regarding experiential learning theories, and each of these directions can be integrated into current research on VR training programs. Notably, Seaman et al. (2017) highlight the importance of integrating neuropsychological as well as sociocultural research with experiential learning theory – both of which have been integrated into research on VR (Hasler & Friedman, 2012; Parsons, 2015).

Likewise, the magnitude of effects depended on the type of outcome in the order specified by Kirkpatrick's hierarchy (Kirkpatrick & Kirkpatrick, 2015, 2016), but few authors measured the distal outcomes: behaviors and results. As these two categories are the ultimate indicators of training success, researchers must move beyond self-reports and tests administered immediately after training programs to ensure the effectiveness of their VR training programs. We suggest that researchers studying students should measure career outcomes, as the ultimate goal for these populations is often employment; researchers studying specialized populations should measure well-being outcomes, as the ultimate goal for these populations is often better mental health; and researchers studying employees should measure organizational performance, as the ultimate goal for these populations is often higher revenue. Researchers should also ensure that their measures are appropriate when performing these studies. Many studies administered self-report measures with little validity or reliability information. Because research in cyberpsychology and human-computer interaction is known to apply measures with questionable properties (Ellis, Kaye, Wilcockson, & Ryding, 2018; Ramirez & Choucri, 2016), it is paramount that researchers provide evidence that their measures are appropriate.

Current research on VR programs has many notable strengths, but many studies take an atheoretical approach. That is, many authors assume that VR will produce beneficial outcomes without applying a specific theory to justify its application. We call for future research to more closely integrate relevant theory into the application and study of VR for the training of social skills. In addition to experiential learning theories, motivational theories could be applied to explain increases to motivation and subsequently learning, such as self-determination theory (Ryan & Deci, 2000) or the theory of planned behavior (Ajzen, 1991). These theories may be able to explain why VR is also effective at developing basic social skills. VR may heighten trainee motivation, as users appear to naturally enjoy the technology (Mestre, Maïano, Dagonneau, & Mercier, 2011), and therefore VR may provide excitement to social skills that may otherwise be boring to train. Applying theories such as these could provide a better understanding of the link between VR, excitement, and improved outcomes. The integration of theory – beyond the examples provided above – appears pivotal for all future research on VR.

Lastly, future researchers should perform further meta-analyses on VR training programs for social skill developments. With future technological developments, the VR training programs may become more effective at developing social skills, and the current meta-analysis will need to be updated with new results – as is the case with all meta-analyses on emergent technologies. In performing these future meta-analyses, future authors should consider replicating the current result using different databases and search parameters. While the search parameters of the current meta-analysis included more sources than typical meta-analyses (13,082), utilizing different approaches to identify sources can ensure that the current results are robust and VR training programs are indeed effective at developing social skills. Also, social skills targeted in more recent articles are not as effectively developed via VR than those targeted in earlier articles, given the negative relationship between effect size and year. Once a larger base of research has been conducted in this domain, a future meta-analysis should compare the effectiveness of VR across multiple types of social skills. Instead of making broad groups based on complexity, as done in the current article, a future meta-analysis should group sources together based on the specific social skill being trained (e.g. eye contact, public speaking, etc.). In doing so, the most effective applications for VR can be clearly identified. While too few articles currently exist to perform such an analysis, it could nevertheless provide widespread implications and open possible avenues for future research.

To conclude, the goal of the current article was to test the efficacy of VR training programs for social skill development as well as identify aspects of the training program, participant population, and study design that influence program success. The results showed that VR training programs are effective, overall, for social skill development, but most of the factors that influence program success contradicted current assumptions. Future research should further probe these factors, and the current meta-analysis is only the beginning of this field of study.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.compedu.2019.103707.

References

Afshar, H., Roohafza, H. R., Keshteli, A. H., Mazaheri, M., Feizi, A., & Adibi, P. (2015). The association of personality traits and coping styles according to stress level. *Journal of Research in Medical Sciences: The Official Journal of Isfahan University of Medical Sciences, 20*(4), 353.

Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes, 50*(2), 179–211.

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

Anderson, C., Shibuya, A., Ihori, N., Swing, E., Bushman, B., Sakamoto, A., & Saleem, M. (2010). Violent video game effects on aggression, empathy, and prosocial behavior in Eastern and Western countries: A meta-analytic review. *Psychological Bulletin, 136*(2).

Armstrong, M. B., Landers, R. N., & Collmus, A. B. (2015). Gamifying recruitment, selection, training, and performance management: Game-thinking in human resource management. In D. Davis, & H. Gangadharbatla (Eds.), *Emerging research and trends in gamification* (pp. 140–165). Hershey, PA: IGI Global.

Baker, J. P., & Berenbaum, H. (2007). Emotional approach and problem-focused coping: A comparison of potentially adaptive strategies. *Cognition & Emotion, 21*(1), 95–118.

von Barnekow, A., Bonet-Codina, N., & Tost, D. (2017). Can 3D gamified simulations be valid vocational training tools for persons with intellectual disability? A pilot based on a real-life situation. *Methods of Information in Medicine, 56*(2), 162–170.

Baumeister, R. F. (2014). Self-regulation, ego depletion, and inhibition. *Neuropsychologia, 65*, 313–319.

Bedwell, W., Pavlas, D., Heyne, K., Lazzara, E., & Salas, E. (2012). Toward a taxonomy linking game attributes to learning: An empirical study. *Simulation & Gaming, 43*(6), 729–760.
Bouchard, S., Dumoulin, S., Robillard, G., Guitard, T., Klinger, E., Forget, H., & Roucaut, F. X. (2017). Virtual reality compared with in vivo exposure in the treatment of social anxiety disorder: A three-arm randomised controlled trial. *The British Journal of Psychiatry, 210*(4), 276–283.
Boxer, P., Groves, C. L., & Docherty, M. (2015). Video games do indeed influence children and adolescents' aggression, prosocial behavior, and academic performance: A clearer reading of Ferguson (2015). *Perspectives on Psychological Science, 10*(5), 671–673.
Buckley, P., & Doyle, E. (2016). Gamification and student motivation. *Interactive Learning Environments, 24*(6), 1162–1175.
Burke, M. J., Sarpy, S. A., Smith-Crowe, K., Chan-Serafin, S., Salvador, R. O., & Islam, G. (2006). Relative effectiveness of worker safety and health training methods. *American Journal of Public Health, 96*(2), 315–324.
Card, S. K. (2017). *The psychology of human-computer interaction*. CRC Press.
Carpenter, C. J. (2012). A trim and fill examination of the extent of publication bias in communication research. *Communication Methods and Measures, 6*(1), 41–55.
Carver, C. S., Scheier, M. F., & Weintraub, J. K. (1989). Assessing coping strategies: A theoretically based approach. *Journal of Personality and Social Psychology, 56*(2), 267.
Chen, H., Dey, A., Billinghurst, M., & Lindeman, R. (2017). November). Exploring the design space for multi-sensory heart rate feedback in immersive virtual reality. In *Proceedings of the 29th Australian conference on computer-human interaction* (pp. 108–116). ACM.
Cheung, M. W. L. (2015). *Meta-analysis: A structural equation modeling approach*. John Wiley & Sons.
Coffey, A. J., Kamhawi, R., Fishwick, P., & Henderson, J. (2017). The efficacy of an immersive 3D virtual versus 2D web environment in intercultural sensitivity acquisition. *Educational Technology Research & Development, 65*(2), 455–479.
Cordeil, M., Dwyer, T., Klein, K., Laha, B., Marriott, K., & Thomas, B. H. (2017). Immersive collaborative analysis of network connectivity: Cave-style or head-mounted display? *IEEE Transactions on Visualization and Computer Graphics, 23*(1), 441–450.
Cousineau, D., & Chartier, S. (2010). Outliers detection and treatment: A review. *International Journal of Psychological Research, 3*(1), 58–67.
Cutolo, F., Meola, A., Carbone, M., Sinceri, S., Cagnazzo, F., Denaro, E., & Ferrari, V. (2017). A new head-mounted display-based augmented reality system in neurosurgical oncology: A study on phantom. *Computer Assisted Surgery, 22*(1), 39–53.
Dalton, D. R., Aguinis, H., Dalton, C. M., Bosco, F. A., & Pierce, C. A. (2012). Revisiting the file drawer problem in meta-analysis: An assessment of published and nonpublished correlation matrices. *Personnel Psychology, 65*(2), 221–249.
Dardas, L. A., & Ahmad, M. M. (2015). Coping strategies as mediators and moderators between stress and quality of life among parents of children with autistic disorder. *Stress and Health, 31*(1), 5–12.
Davis-Kendrick, M. (2015). *Effects of simulation on critical thinking, communication, and decision-making abilities in third semester associate degree nursing students*. Doctoral dissertation, Jones International University.
Dear, K. B., & Begg, C. B. (1992). An approach for assessing publication bias prior to performing a meta-analysis. *Statistical Science*, 237–245.
Debray, T. P., Moons, K. G., & Riley, R. D. (2018). Detecting small-study effects and funnel plot asymmetry in meta-analysis of survival data: A comparison of new and existing tests. *Research Synthesis Methods, 9*(1), 41–50.
Detrick, W. S. (2012). *Effective use of a computer planetarium to improve students' understanding of moon phases and eclipses* (Doctoral dissertation).
Dicheva, D., Dichev, C., Agre, G., & Angelova, G. (2015). Gamification in education: A systematic mapping study. *Journal of Educational Technology & Society, 18*(3), 75–88.
Druckman, D., & Robinson, V. (1998). From research to application: Utilizing research findings in negotiation training programs. *International Negotiation, 3*(1), 7–38.
Duval, S., & Tweedie, R. (2000). Trim and fill: A simple funnel-plot–based method of testing and adjusting for publication bias in meta-analysis. *Biometrics, 56*(2), 455–463.
Ellis, D. A., Kaye, L. K., Wilcockson, T. D., & Ryding, F. C. (2018). Digital traces of behaviour within addiction: Response to Griffiths (2017). *International Journal of Mental Health and Addiction, 16*(1), 240–245.
ElShenawy, E. (2010). Does negotiation training improve negotiators' performance? *Journal of European Industrial Training, 34*(3), 192–210.
Felnhofer, A., Kothgassner, O. D., Hetterle, T., Beutl, L., Hlavacs, H., & Kryspin-Exner, I. (2014). Afraid to be there? Evaluating the relation between presence, self-reported anxiety, and heart rate in a virtual public speaking task. *Cyberpsychology, Behavior, and Social Networking, 17*(5), 310–316.
Ferrara, J. (2013). Games for persuasion: Argumentation, procedurality, and the lie of gamification. *Games and Culture, 8*(4), 289–304.
Fowler, C. (2015). Virtual reality and learning: Where is the pedagogy? *British Journal of Educational Technology, 46*(2), 412–422.
Freina, L., & Ott, M. (2015). A literature review on immersive virtual reality in education: State of the art and perspectives. *eLearning & Software for Education, 1*.
Garner, R., Gillingham, M. G., & White, C. S. (1989). Effects of 'seductive details' on macroprocessing and microprocessing in adults and children. *Cognition and Instruction, 6*(1), 41–57.
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*(3), 463–488.
Hagger, M. S., Wood, C., Stiff, C., & Chatzisarantis, N. L. (2010). Ego depletion and the strength model of self-control: A meta-analysis. *Psychological Bulletin, 136*(4), 495.
Hamari, J., Koivisto, J., & Sarsa, H. (2014, January). Does gamification work? A literature review of empirical studies on gamification. In *2014 47th Hawaii international conference on system sciences (HICSS)* (pp. 3025–3034). IEEE.
Harasim, L. (2017). *Learning theory and online technologies*. Routledge.
Harrison, D. A., Newman, D. A., & Roth, P. L. (2006). How important are job attitudes? Meta-analytic comparisons of integrative behavioral outcomes and time sequences. *Academy of Management Journal, 49*(2), 305–325.
Hasler, B. S., & Friedman, D. A. (2012). Sociocultural conventions in avatar-mediated nonverbal communication: A cross-cultural analysis of virtual proxemics. *Journal of Intercultural Communication Research, 41*(3), 238–259.
Hedges, L. V., & Vevea, J. L. (1998). Fixed-and random-effects models in meta-analysis. *Psychological Methods, 3*(4), 486.
Heuett, B. L., & Heuett, K. B. (2011). Virtual reality therapy: A means of reducing public speaking anxiety. *International Journal of Humanities and Social Science, 1*(16), 1–6.
Howard, M. C. (2017). A meta-analysis and systematic literature review of virtual reality rehabilitation programs. *Computers in Human Behavior, 70*, 317–327.
Howard, M. C. (2017). Investigating the simulation elements of environment and control: Extending the Uncanny Valley Theory to simulations. *Computers & Education, 109*, 216–232.
Howard, M. C. (2018). Virtual reality interventions for personal development: A meta-analysis of hardware and software. *Human-Computer Interaction*, 1–35.
Howard, M. C., & Rose, J. C. (2019). *Refining and extending task–technology fit theory: Creation of two task–technology fit scales and empirical clarification of the construct*. Information & Management.
Huang, T. C., Chen, C. C., & Chou, Y. W. (2016). Animating eco-education: To see, feel, and discover in an augmented reality-based experiential learning environment. *Computers & Education, 96*, 72–82.
Hughes, A. M., Gregory, M. E., Joseph, D. L., Sonesh, S. C., Marlow, S. L., Lacerenza, C. N., & Salas, E. (2016). Saving lives: A meta-analysis of team training in healthcare. *Journal of Applied Psychology, 101*(9), 1266.
Hunter, J. E., & Schmidt, F. L. (2000). Fixed effects vs. random effects meta-analysis models: Implications for cumulative research knowledge. *International Journal of Selection and Assessment, 8*(4), 275–292.
Hu, R., Wu, Y. Y., & Shieh, C. J. (2016). Effects of virtual reality integrated creative thinking instruction on students' creative thinking abilities. *Eurasia Journal of Mathematics, Science and Technology Education, 12*(3).
Contemporary theories of learning: Learning theorists. In Illeris, K. (Ed.), *Their own words*, (2018). Routledge.
Ip, H. H., Wong, S. W., Chan, D. F., Byrne, J., Li, C., Yuan, V. S., & Wong, J. Y. (2018). Enhance emotional and social adaptation skills for children with autism spectrum disorder: A virtual reality enabled approach. *Computers & Education, 117*, 1–15.

Irish, J. E. (2013). Can I sit here? A review of the literature supporting the use of single-user virtual environments to help adolescents with autism learn appropriate social communication skills. *Computers in Human Behavior, 29*(5), A17–A24.
Jak, S. (2015). *Meta-analytic structural equation modelling*. Dordrecht, Neth: Springer.
Jensen, L., & Konradsen, F. (2018). A review of the use of virtual reality head-mounted displays in education and training. *Education and Information Technologies, 23*(4), 1515–1529.
Jones, R. J., Woods, S. A., & Guillaume, Y. R. (2016). The effectiveness of workplace coaching: A meta-analysis of learning and performance outcomes from coaching. *Journal of Occupational and Organizational Psychology, 89*(2), 249–277.
Kampmann, I. L., Emmelkamp, P. M., Hartanto, D., Brinkman, W. P., Zijlstra, B. J., & Morina, N. (2016). Exposure to virtual social interactions in the treatment of social anxiety disorder: A randomized controlled trial. *Behaviour Research and Therapy, 77*, 147–156.
Kaplan, A. M., & Haenlein, M. (2009). The fairyland of Second Life: Virtual social worlds and how to use them. *Business Horizons, 52*(6), 563–572.
Kass, A., Burke, R., Blevis, E., & Williamson, M. (1994). Constructing learning environments for complex social skills. *The Journal of the Learning Sciences, 3*(4), 387–427.
Kepes, S., Banks, G. C., McDaniel, M., & Whetzel, D. L. (2012). Publication bias in the organizational sciences. *Organizational Research Methods, 15*(4), 624–662.
Kepes, S., McDaniel, M. A., Brannick, M. T., & Banks, G. C. (2013). Meta-analytic reviews in the organizational sciences: Two meta-analytic schools on the way to MARS (the Meta-Analytic Reporting Standards). *Journal of Business and Psychology, 28*(2), 123–143.
Kingston, D. C., Riddell, M. F., McKinnon, C. D., Gallagher, K. M., & Callaghan, J. P. (2016). Influence of input hardware and work surface angle on upper limb posture in a hybrid computer workstation. *Human Factors, 58*(1), 107–119.
Kirkpatrick, J. D., & Kirkpatrick, W. K. (2015). Creating a post-training evaluation plan. *Training & Development, 42*(1), 12.
Kirkpatrick, J. D., & Kirkpatrick, W. K. (2016). *Kirkpatrick's four levels of training evaluation*. Association for Talent Development.
Klaassen, R. G., de Vries, P., & Ceulemans, D. S. (2018). *The applicability of feedback of virtual speech app metrics in a presentation technique course SEFI 2018 conference*.
Klimova, B. (2016). Computer-based cognitive training in aging. *Frontiers in Aging Neuroscience, 8*, 313.
Kolb, D. A. (2014). *Experiential learning: Experience as the source of learning and development*. FT press.
Kolb, A. Y., & Kolb, D. A. (2012). Experiential learning theory. In *Encyclopedia of the sciences of learning* (pp. 1215–1219). Boston, MA: Springer.
Krasny, L., Williams, B. J., Provencal, S., & Ozonoff, S. (2003). Social skills interventions for the autism spectrum: Essential ingredients and a model curriculum. *Child and Adolescent Psychiatric Clinics of North America, 12*(1), 107–122.
Lahti, M., Hätönen, H., & Välimäki, M. (2014). Impact of e-learning on nurses' and student nurses knowledge, skills, and satisfaction: A systematic review and meta-analysis. *International Journal of Nursing Studies, 51*(1), 136–149.
Landers, R. N., & Armstrong, M. B. (2017). Enhancing instructional outcomes with gamification: An empirical test of the Technology-Enhanced Training Effectiveness Model. *Computers in Human Behavior, 71*, 499–507.
Landers, R. N., Armstrong, M. B., & Collmus, A. B. (2017). How to use game elements to enhance learning: Applications of the theory of gamified learning. In M. Ma, & A. Oiko-nomou (Eds.), *Serious games and edutainment applications* (Vol. 2, pp. 457–483). Cham, Switzerland: Springer.
Lane, H. C., Hays, M. J., Core, M. G., & Auerbach, D. (2013). Learning intercultural communication skills with virtual humans: Feedback and fidelity. *Journal of Educational Psychology, 105*(4), 1026.
Liberati, A., Altman, D. G., Tetzlaff, J., Mulrow, C., Gøtzsche, P. C., Ioannidis, J. P., & Moher, D. (2009). The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: Explanation and elaboration. *PLoS Medicine, 6*(7), e1000100.
Lister, C., West, J. H., Cannon, B., Sax, T., & Brodegard, D. (2014). Just a fad? Gamification in health and fitness apps. *JMIR Serious Games, 2*(2).
Lou, Y., Abrami, P. C., & d'Apollonia, S. (2001). Small group and individual learning with technology: A meta-analysis. *Review of Educational Research, 71*(3), 449–521.
Luu, T. P., He, Y., Brown, S., Nakagome, S., & Contreras-Vidal, J. L. (2016). Gait adaptation to visual kinematic perturbations using a real-time closed-loop brain–computer interface to a virtual reality avatar. *Journal of Neural Engineering, 13*(3), 036006.
Manca, D., Brambilla, S., & Colombo, S. (2013). Bridging between virtual reality and accident simulation for training of process-industry operators. *Advances in Engineering Software, 55*, 1–9.
Marshall, J., Booth, T., Devane, N., Galliers, J., Greenwood, H., Hilari, K., & Woolf, C. (2016). Evaluating the benefits of aphasia intervention delivered in virtual reality: Results of a quasi-randomised study. *PLoS One, 11*(8), e0160381.
Matson, J. L., Rotatori, A. F., & Helsel, W. J. (1983). Development of a rating scale to measure social skills in children: The Matson Evaluation of Social Skills with Youngsters (MESSY). *Behaviour Research and Therapy, 21*(4), 335–340.
McCarthy, M. (2016). Experiential learning theory: From theory to practice. *Journal of Business & Economics Research (JBER), 14*(3), 91–100.
McFall, R. M. (1982). *A review and reformulation of the concept of social skills*. Behavioral Assessment.
Merchant, Z., Goetz, E. T., Cifuentes, L., Keeney-Kennicutt, W., & Davis, T. J. (2014). Effectiveness of virtual reality-based instruction on students' learning outcomes in K-12 and higher education: A meta-analysis. *Computers & Education, 70*, 29–40.
Mestre, D. R., Maïano, C., Dagonneau, V., & Mercier, C. S. (2011). Does virtual reality enhance exercise performance, enjoyment, and dissociation? An exploratory study on a stationary bike apparatus. *Presence: Teleoperators and Virtual Environments, 20*(1), 1–14.
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.
Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *Annals of Internal Medicine, 151*(4), 264–269.
Moher, D., Shamseer, L., Clarke, M., Ghersi, D., Liberati, A., Petticrew, M., & Stewart, L. A. (2015). Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Systematic Reviews, 4*(1), 1.
Morgan, J. R., Price, M., Schmertz, S. K., Johnson, S. B., Masuda, A., Calamaras, M., et al. (2014). Cognitive processes as mediators of the relation between mindfulness and change in social anxiety symptoms following cognitive behavioral treatment. *Anxiety, Stress & Coping, 27*(3), 288–302.
Nacke, L. E., & Deterding, C. S. (2017). The maturing of gamification research. *Computers in Human Behavior*, 450–454.
North, M. M., North, S. M., & Coble, J. R. (2015). Virtual reality therapy: An effective treatment for the fear of public speaking. *International Journal of Virtual Reality (IJVR), 3*(3), 1–6.
Odom, S. L., Thompson, J. L., Hedges, S., Boyd, B. A., Dykstra, J. R., Duda, M. A., & Bord, A. (2015). Technology-aided interventions and instruction for adolescents with autism spectrum disorder. *Journal of Autism and Developmental Disorders, 45*(12), 3805–3819.
Oh, H., Fiore, A. M., & Jeoung, M. (2007). Measuring experience economy concepts: Tourism applications. *Journal of Travel Research, 46*(2), 119–132.
Ozonoff, S., & Miller, J. N. (1995). Teaching theory of mind: A new approach to social skills training for individuals with autism. *Journal of Autism and Developmental Disorders, 25*(4), 415–433.
Papadopoulos, L., Pentzou, A. E., Louloudiadis, K., & Tsiatsos, T. K. (2013). Design and evaluation of a simulation for pediatric dentistry in virtual worlds. *Journal of Medical Internet Research, 15*(10).
Park, K. M., Ku, J., Choi, S. H., Jang, H. J., Park, J. Y., Kim, S. I., et al. (2011). A virtual reality application in role-plays of social skills training for schizophrenia: A randomized, controlled trial. *Psychiatry Research, 189*(2), 166–172.
Parsons, T. (2015). Virtual reality for enhanced ecological validity and experimental control in the clinical, affective and social neurosciences. *Frontiers in Human Neuroscience, 9*, 660.
Parsons, T. D., & Rizzo, A. A. (2008). Affective outcomes of virtual reality exposure therapy for anxiety and specific phobias: A meta-analysis. *Journal of Behavior Therapy and Experimental Psychiatry, 39*(3), 250–261.
Pashler, H., & Wagenmakers, E. J. (2012). Editors' introduction to the special section on replicability in psychological science: A crisis of confidence? *Perspectives on Psychological Science, 7*(6), 528–530.
Pine, B. J., II, & Gilmore, J. H. (1999). *The experience economy: Work is theatre & every business a stage*. Boston: Harvard Business School Press.

Plunkett, L. (2014). *Facebook buys Oculus Rift for $2 billion*. Kotaku.com. Retrieved from: https://kotaku.com/facebook-buys-oculus-rift-for-2-billion-1551487939.
Pot-Kolder, R., Geraets, C., Veling, W., van Beilen, M., Staring, A., Gijsman, H., & van der Gaag, M. (2018). Virtual-reality-based cognitive behavioural therapy versus waiting list control for paranoid ideation and social avoidance in patients with psychotic disorders: A single-blind randomised controlled trial. *The Lancet Psychiatry, 5*(3), 217–226.
Powers, M. B., & Emmelkamp, P. M. (2008). Virtual reality exposure therapy for anxiety disorders: A meta-analysis. *Journal of Anxiety Disorders, 22*(3), 561–569.
Preece, J., Rogers, Y., & Sharp, H. (2015). *Interaction design: Beyond human-computer interaction*. John Wiley & Sons.
Qian, L., Barthel, A., Johnson, A., Osgood, G., Kazanzides, P., Navab, N., et al. (2017). Comparison of optical see-through head-mounted displays for surgical interventions with object-anchored 2D-display. *International Journal of Computer Assisted Radiology and Surgery, 12*(6), 901–910.
Ramirez, R., & Choucri, N. (2016). Improving interdisciplinary communication with standardized cyber security terminology: A literature review. *IEEE Access, 4*, 2216–2243.
Rauscher, L., Kohn, J., Käser, T., Mayer, V., Kucian, K., McCaskey, U., & von Aster, M. (2016). Evaluation of a computer-based training program for enhancing arithmetic skills and spatial number representation in primary school children. *Frontiers in Psychology, 7*, 913.
Real, F. J., DeBlasio, D., Beck, A. F., Ollberding, N. J., Davis, D., Cruse, B., & Klein, M. D. (2017). A virtual reality curriculum for pediatric residents decreases rates of influenza vaccine refusal. *Academic Pediatrics, 17*(4), 431–435.
Riggio, R. E. (1986). Assessment of basic social skills. *Journal of Personality and Social Psychology, 51*(3), 649.
Rogers, S. A. (2017). *A massively multiplayer online role-playing game with language learning strategic activities to improve English grammar, listening, reading, and vocabulary*. University of South Alabama.
Rothstein, H. R., & Bushman, B. J. (2015). Methodological and reporting errors in meta-analytic reviews make other meta-analysts angry: A commentary on Ferguson (2015). *Perspectives on Psychological Science, 10*(5), 677–679.
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.
Sardi, L., Idri, A., & Fernández-Alemán, J. L. (2017). A systematic review of gamification in e-Health. *Journal of Biomedical Informatics, 71*, 31–48.
Seaman, J., Brown, M., & Quay, J. (2017). The evolution of experiential learning theory: Tracing lines of research in the JEE. *Journal of Experiential Education, 40*(4), NP1–NP21.
Seth, A., Vance, J. M., & Oliver, J. H. (2011). Virtual reality for assembly methods prototyping: A review. *Virtual Reality, 15*(1), 5–20.
Simonsohn, U., Nelson, L. D., & Simmons, J. P. (2014). P-curve: A key to the file-drawer. *Journal of Experimental Psychology: General, 143*(2), 534.
Slaughter, V., & Gopnik, A. (1996). Conceptual coherence in the child's theory of mind: Training children to understand belief. *Child Development, 67*(6), 2967–2988.
Spadaro, S., Karbing, D., Fogagnolo, A., Ragazzi, R., Mojoli, F., Astolfi, L., & Volta, C. (2017). Simulation training for residents focused on mechanical ventilation: A randomized trial using mannequin-based versus computer-based simulation. *Simulation in Healthcare, 12*(6), 349.
Strickland, D. C., Coles, C. D., & Southern, L. B. (2013). JobTIPS: A transition to employment program for individuals with autism spectrum disorders. *Journal of Autism and Developmental Disorders, 43*(10), 2472–2483.
Sullivan, D. K., Goetz, J. R., Gibson, C. A., Mayo, M. S., Washburn, R. A., Lee, Y., & Donnelly, J. E. (2016). A virtual reality intervention (Second Life) to improve weight maintenance: Rationale and design for an 18-month randomized trial. *Contemporary Clinical Trials, 46*, 77–84.
Sweigart, L., & Hodson-Carlton, K. (2013). Improving student interview skills: The virtual avatar as client. *Nurse Educator, 38*(1), 11–15.
Thompson, S. G., & Higgins, J. P. (2002). How should meta-regression analyses be undertaken and interpreted? *Statistics in Medicine, 21*(11), 1559–1573.
Tsang, M. M., & Man, D. W. (2013). A virtual reality-based vocational training system (VRVTS) for people with schizophrenia in vocational rehabilitation. *Schizophrenia Research, 144*(1–3), 51–62.
Van Houwelingen, H. C., Arends, L., & Stijnen, T. (2002). Advanced methods in meta-analysis: Multivariate approach and meta-regression. *Statistics in Medicine, 21*(4), 589–624.
Van Wijk, R., Jansen, J. J., & Lyles, M. A. (2008). Inter-and intra-organizational knowledge transfer: A meta-analytic review and assessment of its antecedents and consequences. *Journal of Management Studies, 45*(4), 830–853.
Van der Kleij, F. M., Feskens, R. C., & Eggen, T. J. (2015). Effects of feedback in a computer-based learning environment on students' learning outcomes: A meta-analysis. *Review of Educational Research, 85*(4), 475–511.
Vevea, J. L., & Woods, C. M. (2005). Publication bias in research synthesis: Sensitivity analysis using a priori weight functions. *Psychological Methods, 10*(4), 428.
Viechtbauer, W., & Cheung, M. W. L. (2010). Outlier and influence diagnostics for meta-analysis. *Research Synthesis Methods, 1*(2), 112–125.
Vora, J., Nair, S., Gramopadhye, A. K., Duchowski, A. T., Melloy, B. J., & Kanki, B. (2002). Using virtual reality technology for aircraft visual inspection training: Presence and comparison studies. *Applied Ergonomics, 33*(6), 559–570.
Wang, C. X., Calandra, B., Hibbard, S. T., & Lefaiver, M. L. M. (2012). Learning effects of an experimental EFL program in Second Life. *Educational Technology Research & Development, 60*(5), 943–961.
Xu, Y., Park, H., & Baek, Y. (2011). A new approach toward digital storytelling: An activity focused on writing self-efficacy in a virtual learning environment. *Journal of Educational Technology & Society, 14*(4), 181.
Yuan, S. N. V., & Ip, H. H. S. (2018). Using virtual reality to train emotional and social skills in children with autism spectrum disorder. *London Journal of Primary Care*, 1–3.
Zhang, M., Zhang, Z., Chang, Y., Aziz, E. S., Esche, S., & Chassapis, C. (2018). Recent developments in game-based virtual reality educational laboratories using the Microsoft Kinect. *International Journal of Emerging Technologies in Learning (iJET), 13*(1), 138–159.
Zhao, H. A. O., Wayne, S. J., Glibkowski, B. C., & Bravo, J. (2007). The impact of psychological contract breach on work-related outcomes: A meta-analysis. *Personnel Psychology, 60*(3), 647–680.
Zhu, Q., Wang, T., & Jia, Y. (2007, November). Second life: A new platform for education. In *Information technologies and applications in education, 2007. ISITAE'07. First IEEE international symposium on* (pp. 201–204). IEEE.
Zimmerman, R. D. (2008). Understanding the impact of personality traits on individuals' turnover decisions: A meta-analytic path model. *Personnel Psychology, 61*(2), 309–348.