Computers in Human Behavior 130 (2022) 107197

Contents lists available at ScienceDirect

Computers in Human Behavior

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

A meta-analysis and systematic literature review of mixed reality rehabilitation programs: Investigating design characteristics of augmented reality and augmented virtuality

Matt C. Howard *, Maggie M. Davis

The University of South Alabama, Mitchell College of Business, USA

ARTICLE INFO

Keyword:
Mixed reality
Augmented reality
Augmented virtuality
Rehabilitation
Kinect
Hololens

ABSTRACT

Researchers have shown great interest in the creation of rehabilitation programs using mixed reality (MR), which includes both augmented reality (inclusion of virtual elements in a real environment) and augmented virtuality (inclusion of real elements in a virtual environment). Due to its recent development and systematic limitations of associated studies (e.g., small sample sizes), many pivotal research questions for the study of MR rehabilitation (MRR) programs remain unanswered, and the current article seeks to answer these questions via a meta-analysis and systematic literature review. Our random-effects meta-analysis, including 29 studies, showed that MRR programs produce significant improvements in users' outcomes, and they were significantly more effective than alternative rehabilitation programs when matched for duration. It also demonstrated that there was not a significant effect regarding all studied characteristics of MRR programs, including the incorporation of game elements, use of head-mounted displays, and task-technology fit. Our systematic literature review identified the most studied types of MRR programs, and it revealed that the most common research designs are the least methodologically robust (e.g., single group post-test only design). Our discussion suggests that MRR programs are still in the earliest stage of research, and the current results are encouraging for future development. Because none of our studied characteristics altered the effectiveness of MRR programs, we argue that characteristics that are generally consistent across all MRR programs are the drivers of their effectiveness, which can be categorized as fidelity and motivational characteristics.

Researchers have continuously applied cutting-edge technologies to improve physical rehabilitation program outcomes, and among the most lauded applications is mixed reality (MR) (Alamri et al., 2010; Mubin et al., 2019; Ying & Aimin, 2017). MR refers to the presentation of integrated real and virtual content, and it exists on a spectrum between entirely real environments and entirely virtual environments (Carmigniani & Furht, 2011; Milgram & Kishino, 1994). The two primary types of MR are augmented reality and augmented virtuality (Carmigniani et al., 2011; Van Krevelen & Poelman, 2010). Augmented reality is the inclusion of virtual elements in a real environment, such as the mobile game, Pokémon Go, that digitally imposes virtual Pokémon in the presentation of a real environment; whereas augmented virtuality is the inclusion of real elements in a virtual environment, such as Xbox Kinect games that digitally impose an image of the user in the presentation of a virtual environment. Both augmented reality and augmented virtuality have been repeatedly applied in physical rehabilitation programs to improve the physical capabilities of people with disabilities, and researchers often discuss them together within the scope of MR (Berton et al., 2020; Gorman & Gustafsson, 2020).

Despite the fanfare surrounding MR rehabilitation (MRR) programs, there remains significant uncertainty regarding their application. Most extant studies on MRR programs are largely intended to demonstrate that the applied approach is feasible and does not pose severe concerns for users (Dorado et al., 2019; Hoermann et al., 2017; LaPiana et al., 2020). Researchers have yet to provide, however, robust evidence that MRR programs are capable of either improving physical abilities or improving physical abilities more than comparable alternatives. Likewise, characteristics of MRR programs can greatly vary, and few insights have been provided regarding which characteristics produce the most beneficial outcomes. For instance, while authors have provided arguments regarding whether MR presented via head-mounted display or screen is more effective for physical rehabilitation purposes, it is still

* Corresponding author. Mitchell College of Business, University of South Alabama, 5811 USA Drive S, Rm. 337, Mobile, AL, 36688, USA.
E-mail address: MHoward@SouthAlabama.edu (M.C. Howard).

https://doi.org/10.1016/j.chb.2022.107197
Received 17 August 2021; Received in revised form 28 December 2021; Accepted 12 January 2022
Available online 15 January 2022

unknown whether significant outcome differences occur in MRR programs when applying these displays. These cumulative uncertainties pose significant barriers to the study of MRR programs, as researchers and practitioners may continue to be limited in the development of MRR programs that produce beneficial outcomes. Equally important, these uncertainties hamper theoretical development regarding MRR programs. As these questions represent essential introductions to theoretical development, researchers cannot confidently apply a specific theoretical lens until these basic insights are established. Thus, resolving these tensions in the current literature is an essential step for the refinement of MRR programs and associated theory.

To address present uncertainties and advance the study of MRR programs, the current article reports a meta-analysis and systematic literature review. Our meta-analysis tests whether MRR programs significantly improve participants' physical capabilities, whether MRR programs significantly improve physical capabilities more than alternative interventions, and which characteristics of these programs systematically produce more beneficial outcomes. Our studied characteristics include aspects of the software (e.g., augmented reality vs. augmented virtuality, purpose, game elements), hardware (e.g., output device), context (e.g., task-technology fit), and research design (e.g., one-group vs. two group). We guide researchers and practitioners towards effective types of MRR programs and the contexts for which these programs may be effective. Alternatively, our systematic literature review characterizes the current field of study on MRR programs. It provides frequencies of studied MRR program characteristics, showing which types of MRR programs and contexts are dominant in empirical research. We likewise assess the frequency of applications relative to the applied research design. This analysis indicates whether researchers are equally applying research designs that adequately assess efficacy across all types of MRR programs or whether certain types of programs are assessed more frequently with research designs that can skew our understanding of MRR programs. Lastly, our discussion integrates our observations to direct future researchers towards theories and explanatory mechanisms that may be essential in understanding MRR programs.

Given these objectives, the current article has four primary implications. First, studies on the efficacy of MR rehabilitation programs often suffer from small sample sizes (e.g., 5 or 10) (Almasi et al., 2020; Blomqvist et al., 2021; Hoda et al., 2015). While they can assess whether some type of change occurs to participants' physical abilities, they are often unable to adequately assess whether these changes are statistically significant. By performing a meta-analysis, we aggregate these samples to more robustly assess whether changes in physical abilities from MRR programs are statistically significant. Second, it is difficult for a single study to compare MRR program elements (e.g., hardware, software, participant population), as researchers would need separate samples for each condition. A meta-analysis, however, can test whether certain MRR program elements systematically produce larger benefits to physical abilities, addressing these research questions. Third, our systematic literature review determines which types of programs are most investigated, which outcomes are most targeted, and which research designs are most used. We identify which characteristics have been sufficiently studied as well as characteristics requiring further investigation. Fourth, our discussion highlights novel theories and mechanisms that may explain the effectiveness of MRR programs, integrating our observations with future research directions. Together, these implications enable future research on MRR programs to progress in a more unified and sophisticated manner.

1. Background

Physical rehabilitation refers to a medical specialty concerned with optimizing physical abilities and reducing disability in people with health conditions (e.g., stroke, chronic pain, and brain injury), often with the goal of restoring or enhancing physical abilities to perform daily activities (e.g., dressing and eating) (WHO, 2021). Prior research has categorized physical abilities as balance, gait, strength/flexibility, motor control, and general (Blankevoort et al., 2010; Howard, 2017; Wiśniowska-Szurlej et al., 2019). Balance refers to the control of postural stability, and it is often assessed by visual or digital pad ratings of sitting or standing sway (Lv et al., 2016). Gait refers to walking abilities, and it is often assessed by the 10-m walk test and 6-min walk test (Lord et al., 2013). Strength/flexibility refers to generated force and range of motion, and example measures are number of exercise reputations, grip strength, sit and reach, and capable degrees of movement in arms or legs (Roberts et al., 2011). Motor control refers to fine movements and speed of action, and it is often assessed by pointing, touching, and grasping objects (Woodbury et al., 2010). General refers to a combination of these abilities, and it is often assessed by participant-reported measures of capabilities for completing tasks requiring multiple physical abilities together (e.g., preparing for work and grocery shopping) (Wiener et al., 1990). Given the essential nature of physical abilities for everyday functioning and overall well-being, researchers have continuously applied cutting-edge technologies to improve physical rehabilitation efficacy with MRR programs being among the most recent innovations. Today, MRR programs have been used to improve physical abilities in all of these categories.

Great variation can be seen across the applications of MRR programs. One program may present a virtual gym environment on a television, and an image of the user may appear in this virtual gym via a camera (augmented virtuality). A virtual trainer could provide instruction, and the user could see themselves performing rehabilitation activities. Another program could be presented via head-mounted display, in which a transparent screen covers the users' eyes and a projector integrates virtual images with the real environment (augmented reality). The head-mounted display could present targets that must be reached by the user, and feedback could be provided based on the number of goals met. Despite the varied nature,[1] authors have repeatedly argued that these differing programs can improve rehabilitation outcomes, and they often discuss augmented reality and augmented virtuality in tandem within the scope of MR.

MRR programs do have certain consistent characteristics. Like all rehabilitation programs, MRR programs instruct users to perform certain activities to develop user physical capabilities (e.g., gait, balance, motor control, strength/flexibility, or general). These programs are typically used over the course of several weeks, and the amount of time per session can vary from minutes (e.g., 15–30) to hours (e.g., 1–2) depending on the users' needs (Assis et al., 2016; Koroleva et al., 2020; Vukićević et al., 2019). These programs may be utilized in conjunction with an instructor who provides direction and feedback; however, MRR programs can typically recognize users' movements because they digitally blend reality and virtuality, and they often include the capability to provide automated direction and feedback. There may be no in-person instructor involved when using an MRR program (Trojan et al., 2014; Viglialoro et al., 2019). This feature enables users to utilize MRR programs in their homes, which can enable more widespread access to rehabilitation opportunities than traditional interventions.

Presently, no theoretical lens is dominant to understand the improved outcomes of MRR programs. Instead, researchers often imply that advancements to technologies should result in advancements to rehabilitation outcomes, which is common in the broader study of rehabilitation (Cervera et al., 2018; Howard, 2017) as well as MR (Garzón et al., 2019; Tekedere & Göke, 2016). In specific cases, some

[1] Both described programs use cameras or motion sensors as input devices rather than controllers. We intended to compare programs that utilize cameras and motion sensors against those that utilize controllers; however, too few studies utilized controllers to be included as category in the meta-analysis or systematic literature review. Thus, we solely refer to MRR programs as utilizing cameras and motion sensors due to their widespread popularity.

researchers argue that MRR programs can provide rehabilitation instruction that may be difficult via other approaches, such as home coaching, but such proposals are likewise made without integrating theory. We believe, however, that MRR programs improve physical capabilities due to two primary theoretical mechanisms, which have been occasionally discussed in prior studies (e.g., de Souza et al., 2019; Li et al., 2021).

First, most MRR programs enable the user to perform naturalistic rehabilitation tasks, whether using augmented reality or augmented virtuality, which reduces the "translation" of behaviors from the training environment to the transfer environment. When users perform physical rehabilitation activities using an MRR program, they can reach for familiar objects, step over common obstacles, and perform an array of other daily activities. The majority of MRR programs also allow users to view their limbs to perform behaviors, further increasing the similarity of behaviors between training and transfer environments (Eichler et al., 2019; Jeon & Kim, 2020). The extent that rehabilitation activities replicate behaviors intended to be performed after the intervention is called fidelity,[2] which is known to produce a positive relation with rehabilitation program effectiveness (Hardeman et al., 2008; Hochmitz & Yuviler-Gavish, 2011; Howard & Jacobs, 2016). The heightened fidelity produced by typical MRR programs is believed to, in part, cause them to be effective rehabilitation interventions.

Second, users tend to naturally enjoy using MR devices, and they often express favorable reactions to programs utilizing the technological medium (Jang & Park, 2019; Rasimah et al., 2011). In modern applications outside of rehabilitation, researchers regularly propose that the novelty of these programs is a large cause of these positive reactions, as users enjoy exploring the capabilities of MR programs that they have otherwise never experienced (Feng & Xie, 2019; Yim et al., 2017). This enjoyment may cause users to remain motivated towards otherwise boring rehabilitation activities. As a baisc example, a person may become bored with reaching for balls to develop their motor abilities, but they may enjoy reaching for digital balls presented on a head-mounted display. In turn, this increased motivation may cause users to work harder at developing physical abilities, ultimately resulting in MRR programs producing improved physical outcomes. Therefore, the heightened motivation produced by modern MRR programs is likewise believed to, in part, cause them to be effective rehabilitation interventions.

Due to these mechanisms, we propose that MRR programs are beneficial for improving physical abilities, and we distinguish two important aspects of MRR program performance that have yet to be sufficiently tested in prior research. Researchers have yet to firmly establish that (1) MRR programs can improve physical capabilities or (2) MRR programs can improve physical capabilities more than alternative interventions (Almasi et al., 2020; Blomqvist et al., 2021; Dorado et al., 2019; Hoermann et al., 2017; LaPiana et al., 2020). Investigations into these research questions include small sample sizes, among other concerns, which precludes robust insights into the effectiveness of MRR programs. For this reason, the current article performs a meta-analysis of prior investigations into the two hypotheses below to aggregate prior findings, which can address concerns of extant studies and provide firmer insights into the effectiveness of MRR programs. It should be noted, however, that it may be sufficient for practice to show that MRR programs produce comparable effects. If the case, then MRR programs may be suitable alternatives to traditional interventions, which could be applied based on contextual needs.

[2] "Fidelity", can also refer to the extent that an intervention was implemented as it was conceptualized, but this type of fidelity is often labeled, "treatment fidelity", instead of simply, "fidelity" (Hildebrand et al., 2012). For this reason, we use the term, "fidelity", to refer to similarity of behaviors between rehabilitation and transfer activities.

Hypothesis 1. Mixed reality rehabilitation programs improve physical capabilities.

Hypothesis 2. Mixed reality rehabilitation programs are more effective than alternative interventions at improving physical capabilities.

As mentioned, researchers typically discuss augmented reality and augmented virtuality in tandem under the scope of MR (Carmigniani et al., 2011; Van Krevelen & Poelman, 2010). Researchers rarely discuss differential effects of augmented reality and augmented virtuality for rehabilitation, but it is important to ensure that the effects of these mediums are similar enough to discuss them together. One medium may produce significantly better rehabilitation outcomes than the other, and it may be inappropriate to discuss them collectively under the broader scope of MR. We therefore test whether augmented reality and augmented virtuality produce differing rehabilitation outcomes. Because we do not have an a priori hypothesis regarding their relative effectiveness, we test this via a research question rather than a directional hypothesis.

Research Question 1. Do augmented reality and augmented virtuality produce differing rehabilitation outcomes?

1.1. Influences on Mixed Reality Rehabilitation Program Effectiveness

Each of the hypotheses below assess whether certain characteristics alter MRR program efficacy. We begin by discussing software (purpose, game elements), followed by hardware (display device), and concluded by context (task-technology fit). These characteristics were chosen based on their popularity in prior research and practice. For instance, past studies have frequently discussed the inclusion of game elements in MRR programs (Goršič et al., 2017; Lange et al., 2012; Ma et al., 2018) but have less frequently discussed aspects such as the location of use or realism of virtual avatars. By identifying whether these commonly discussed characteristics significantly impact MRR program efficacy, we provide broader implications. Furthermore, design characteristics are often intended to enhance the beneficial properties of MR, notably fidelity and enjoyment. We therefore draw on the same theoretical mechanisms discussed above when detailing the proposed benefits of our studied design characteristics.

Researchers have put extensive efforts into creating MRR programs that are customized to target specific rehabilitation activities, as developing physical abilities is often an idiosyncratic process that requires personalized rehabilitation routines (Gonze et al., 2020; Meurer et al., 2017). For instance, a person developing gait abilities after experiencing a stroke can suffer severe injury if a rehabilitation program includes excessively difficult activities and causes the person to strain or fall, whereas the same person could suffer stunted physical rehabilitation if the program does not sufficiently exert their physical capabilities. On the other hand, researchers have observed that many MR programs developed for alternative purposes may be effective for rehabilitation, as they require many of the same movements as standard rehabilitation programs (Chang et al., 2011; Da Gama et al., 2015; Webster & Celik, 2014). For example, many popular Xbox Kinect games require users to lean, squat, reach, and jump – which are all activities that would be beneficial in rehabilitating gait, balance, motor control, and strength/flexibility. These researchers speculate that these games may provide effective MRR program outcomes.

While MR programs developed for general purposes may improve physical abilities, research has consistently shown that interventions are more effective if they are customized to the participant population (Collins et al., 2014; Fritz et al., 2020). Customized MRR programs are often designed to elicit behaviors that better imitate actions performed in the real-world, and they are likewise often designed to target specific behaviors that users have difficulties performing. This suggests that customized MRR programs have greater fidelity than MRR programs created for general purpose, and MRR programs developed for rehabilitation purposes are expected to be more effective than MRR programs developed for general purposes.

Hypothesis 3. Mixed reality programs developed specifically for rehabilitation purposes are more effective than mixed reality rehabilitation programs developed for general purposes (e.g., enjoyment).

The most discussed software feature of any MR application is likely the inclusion of game elements (Ioannides et al., 2017; Muñoz et al., 2016). Across most contexts, from education to rehabilitation, authors have argued that the inclusion of game elements can improve user outcomes because game elements enhance user enjoyment and motivation. A traditional rehabilitation program may cause the user to lose interest, thus limiting the development of physical abilities. Including game elements, however, can enhance excitement during the rehabilitation process, causing the user to remain engaged and persist in developing abilities.

Many taxonomies exist to identify and define different types of game elements, including assessment, conflict, and fantasy (Bedwell et al., 2012; Toda et al., 2019). While some authors have begun to investigate the impact of individual game elements (Huang & Hew, 2015; Landers & Marin, 2021), most authors utilize omnibus approaches wherein the influence of "gamifying" a program is assessed. Gamifying a program typically refers to adding multiple game elements to a program in an omnibus manner (Landers et al., 2015, 2017). For instance, a developer could make an MRR program more challenging, implement a method to record score, and present user scores together on a leaderboard. While each of these changes constitutes separate game elements, the collective addition would be considered gamifying a program. The omnibus investigation of gamified programs is especially evident in the study of rehabilitation, whether using MR or another medium (Goršič et al., 2017; Lange et al., 2012; Ma et al., 2018). Thus, we assess the overall influence of including game elements in MRR programs.

Hypothesis 4. Mixed reality rehabilitation programs with game elements are more effective than mixed reality rehabilitation programs without game elements.

Alternatively, the most discussed hardware feature is likely the use of head-mounted displays (Guinet et al., 2019, 2021; Xue et al., 2019). Head-mounted displays are devices worn on the head with a visor that covers the user's eyes. The visor could be a screen in the case of head-mounted displays for virtual reality, or it could be a translucent display with projected images in the case of head-mounted displays for MR. These devices are presently more associated with virtual reality, as several companies now mass-produce head-mounted displays for consumer use (e.g., Oculus and Valve); however, companies have developed head-mounted displays for MR applications, such as Microsoft's HoloLens. Authors have argued that users naturally enjoy head-mounted displays (Debarba et al., 2018; Held et al., 2020; Jiang et al., 2020; Xue et al., 2019), and the use of devices such as the HoloLens can improve MRR program outcomes due to increases in motivation. The increased mobility of head-mounted displays can also allow more effective rehabilitation exercises (Held et al., 2020). While screen-based MRR programs can only present information in one location, head-mounted displays move with the user and can present information anywhere. MRR programs using head-mounted displays can enable the user to perform more realistic practice behaviors for certain physical abilities, such as gait, resulting in greater physical fidelity. These two benefits of head-mounted displays are believed to improve user outcomes when developing physical abilities, especially those requiring significant movement, such as gait. Therefore, we test whether MRR programs using head-mounted displays are indeed more effective than alternative MRR programs.

Hypothesis 5. Mixed reality rehabilitation programs using head-mounted displays are more effective than mixed reality rehabilitation programs using other output devices.

While many authors have argued that MRR programs are beneficial for users across all applications, others have suggested that their outcomes are dependent upon the context (Assis et al., 2019; Held et al., 2020). In other words, MRR programs are effective in contexts that benefit from their implementation. This notion matches task-technology fit theory, which proposes that technologies positively impact user outcomes if the technology's capabilities align with the task's needs (Goodhue & Thompson, 1995; Howard & Rose, 2019). This theory further proposes that the construct of task-technology fit, the extent that the technology aligns with the task, is the primary predictor of user outcomes. When performing our meta-analysis, we code the task-technology fit of the MRR programs relative to each measured outcome, and we test whether pairings with high task-technology fit are more effective than pairings with low task-technology fit. For instance, authors often test the efficacy of an MRR program to develop specific abilities, such as balance, but they may measure multiple different outcomes, such as balance and gait, as indicators of program efficacy (Ku et al., 2019; Lozano-Quilis et al., 2014). In this case, the MRR program has high task-technology fit with balance outcomes, but it has low task-technology fit with gait outcomes. By coding these outcomes, we meta-analytically assess whether task-technology fit theory identifies effective applications of MRR programs.

Hypothesis 6. Mixed reality programs' task-technology fit has a positive relation with effectiveness.

1.2. *Characterizing the field of study*

In addition to determining the influence of the studied design characteristics via meta-analysis, we also assess their frequency of application in studies of MRR programs by conducting a systematic literature review. For instance, we identify whether head-mounted displays or screens are more common output devices in studies on MRR programs. We also identify the frequency of study for two elements not discussed above.

First, physical abilities developed in rehabilitation programs have been categorized as balance, gait, strength/flexibility, motor control, and general (Blankevoort et al., 2010; Howard, 2017; Wiśniowska-Szurlej et al., 2019). We do not have an a priori belief that MRR programs would have differential effectiveness based on the type of physical ability, but it is important to identify whether certain physical abilities are more commonly developed via MRR programs. This would indicate that either these abilities are more suitable for development via MRR programs and/or more research attention is needed for MRR programs to develop less-studied physical abilities. Second, while most studies of MRR programs utilize samples that benefit from their use (e.g., stroke survivors), some studies utilize able-bodied individuals to assess whether the program is feasible. A significant portion of extant knowledge regarding MRR programs may be based on studies using populations not intended for their use (e.g., able-bodied individuals), and it cannot be guaranteed that these results generalize to desired populations. If true, then prior findings would need to be replicated with participants from targeted populations.

Research Question 2. What are the common software, hardware, and context characteristics of mixed reality rehabilitation programs?

An array of research designs has been applied to assess the effectiveness of MRR programs, which have varying levels of methodological robustness. We focus on the three most common approaches for the study of MRR programs to characterize the methodological sophistication of the field and determine whether MRR programs are investigated via methods that can accurately detect their effectiveness. We also assess whether certain research designs are applied more commonly to investigate certain types of MRR programs, which may indicate whether certain program characteristics are habitually studied via less-than-ideal approaches. It should be noted that these research designs are often believed to represent a progression of research, whereas the more-robust designs are applied after a certain MRR program has been supported via the less-robust designs. Differences in the frequency of application for a MRR program characteristic may not indicate shortcomings of researchers, but it may instead suggest that a certain research stream is still being developed and necessitates further research.

The least-robust common research design is the “single-group post” design, wherein all participants undergo an MRR program and outcomes are solely measured afterwards. This research design is used to assess whether the MRR program is feasible and/or detrimental to participants, but it cannot robustly assess whether the MRR program improves physical abilities or improves physical abilities more than alternative interventions. The moderately-robust common research design is the “single-group pre-post” design, wherein all participants undergo an MRR program and outcomes are measured both before and after the intervention. This research design is used to assess whether the MRR program improves physical abilities, but it cannot robustly assess whether the MRR program improves physical abilities more than alternative programs. The most robust common research design is the “two-group design”, wherein one group undergoes an MRR program and another group undergoes an alternative intervention. This research design can assess both whether the MRR programs improves outcomes and whether the MRR program improves physical abilities more than alternative programs. Of these three designs, it is considered the optimal research design for assessing relative improvement, as it can control for preexisting differences between the treatment and control groups. We assess the frequency that these research designs are applied as well as the frequency that they are applied relative to each MRR program characteristic described above.

Research Question 3. What frequency are the common research designs of mixed reality rehabilitation program evaluations?

2. Method

To answer our hypotheses and research questions, we performed a meta-analysis and systematic literature view. Meta-analyses aggregate the results of prior studies to obtain overall estimates, but they can also determine whether certain characteristics systematically produce larger results. Systematic literature reviews provide quantitative insights into specified characteristics of prior studies, but they do not involve the aggregation of effect sizes. We followed guidelines of prior authors to conduct these analyses with a particular focus on the preferred reporting for systematic reviews and meta-analyses (PRISMA) (Borenstein et al., 2021; Hunter & Schmidt, 2004; Moher et al., 2015; Shamseer et al., 2015; Wiernik & Dahlke, 2020).

2.1. Identifying sources

We used standard practices to discover all relevant published and unpublished sources. Searches were conducted in May of 2021 in four databases: EBSCO, IEEE Xplore, Cochrane Database, and Google Scholar. EBSCO searches return results for several other databases, including Academic Source Complete, ERIC, and PsycINFO. All databases catalogue ample unpublished materials (e.g., conference presentations, theses, dissertations, and grant reports), enabling the identification of sufficient original studies. We recorded all results when searching EBSCO, IEEE Xplore, and Cochrane Database. Google Scholar, however, produced excessive results for some searches (e.g., >15,000 sources), and we only retained the first 1000 results of our Google Scholar searches. This decision was supported when coding our search results (detailed below), as Google Scholar results became almost entirely irrelevant after the first few hundred results. We are confident that this search strategy uncovered most relevant sources.

Each search within these databases included two parts. The first part included: “Mixed Reality”, “Augmented Reality”, “Augmented Virtuality”, “Extended Reality”, “HoloLens”, or “Kinect”. The second part always included the word: “Rehabilitation”. We also conducted forwards and backwards searches of prominent articles on MRR programs by identifying sources cited by the discovered sources as well as relevant sources that cited the discovered sources, and we contacted all authors included in our final dataset for unpublished studies on MRR programs. All searches were restricted to the year 2010 to the present. Dramatic advancements have been made to MR technologies within the past decade, such as the Kinect (2010), Kinect 2 (2014), HoloLens (2016), Azure Kinect (2019), and HoloLens 2 (2019). Including MRR studies prior to the development of these technologies may cause MRR programs to appear ineffective, although modern applications may pose great benefits. Including only sources from the year 2010 to the present can provide more accurate assessments. Our searches returned 4641 initial sources.

2.2. Inclusion criteria

Two coders jointly developed coding guidelines, developed a coding rulebook, and trained each other on coding results. The two trained coders then initially coded the same 20 sources independently until a sufficient level of inter-rater agreement was achieved (Cohen’s $\kappa >$ 0.80), and then continued to independently code each source while conferring on any disagreements. This process is typical in meta-analyses (Borenstein et al., 2021).

In the first coding phase, the sources were coded for the following criteria: (a) written in the English language, (b) purported to study MR, (c) studied physical rehabilitation, and (c) provided quantitative results. The first coding phase reduced the list from 4641 sources to 142 sources. For most of these 142 sources, the authors provided sufficient information to confirm that they studied a MR program (e.g., program names, pictures, descriptions), but some authors did not provide sufficient information to verify. These authors were contacted for information regarding their MR program, but sources were excluded if the authors did not respond or were unable to provide information. During this process, authors from all 142 sources were also contacted for any additional unpublished or published study results. We believed that directly contacting authors is more effective than posting to listservs or social media, as people are more likely to respond to direct inquiries rather than impersonalized and general requests (Muñoz-Leiva et al., 2010). This process reduced the list from 142 sources to 108.

In the second coding phase, the remaining sources were coded based on their research design. Two group and one group pre-post studies were included within meta-analyses,[3] whereas all studies with quantitative results were included within the systematic literature review. All sources were coded for the characteristics described below, and those included in the meta-analysis also had their effect sizes recorded. Some studies did not report effect sizes that could be meta-analyzed. In these cases, the authors were contacted and asked to provide relevant statistics. If they did not reply, the article was not included in the meta-analytic calculations, but it was included in the systematic literature review. This process resulted in 27 sources with 29 total studies included within our meta-analysis, whereas 108 sources with 114 total studies were included in the systematic literature review. Fig. 1 includes a depiction of our coding process.

2.3. Source and study characteristic coding

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

[3] One group post studies cannot be included in meta-analyses, as any reported effect size from these studies (e.g., one-sample *t*-test) are not comparable in meta-analytic designs (Borenstein et al., 2021; Hunter & Schmidt, 2004).

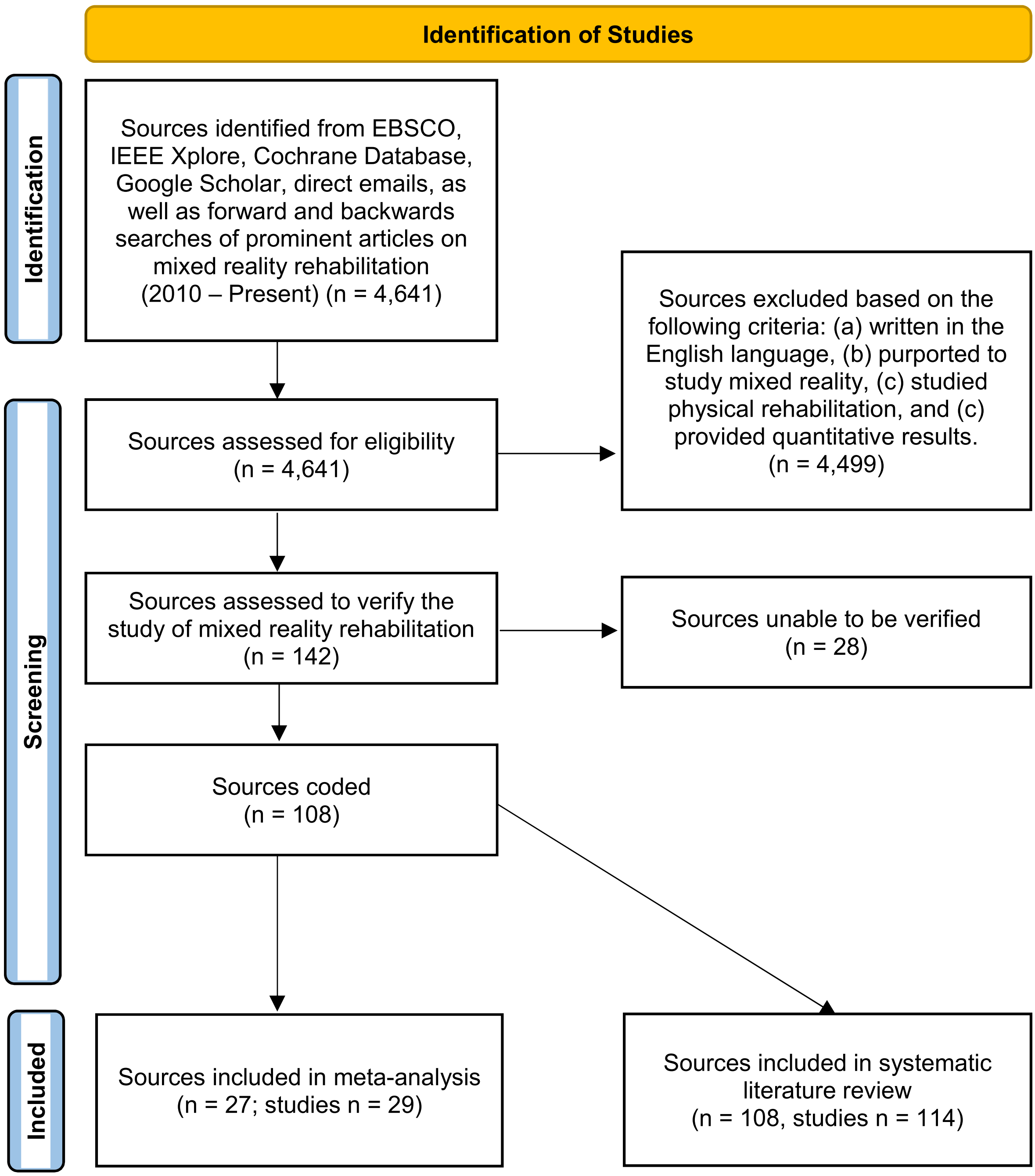

Fig. 1. Visual representation of coding process.

2.3.1. *Type of mixed reality*

Sources were coded on whether the MRR program utilized augmented reality or augmented virtuality.

2.3.2. *Purpose of program*

Sources were coded on whether the MRR program was specifically developed for rehabilitation purposes or general purposes (e.g., Kinect games). We also recorded the name of the applied MR software.

2.3.3. *Game elements*

Sources were coded on whether the MRR program included game elements. This was decided by both the presented images and/or video of the MRR program as well as the authors' description of the MRR program. If the program appeared or was described to contain game elements commonly included in various taxonomies (e.g., score) (Bedwell et al., 2012; Toda et al., 2019), then it was coded as including game elements.

2.3.4. *Output device*

Sources were coded on whether the MRR program was presented via a head-mounted display (e.g., HoloLens) or other output device (e.g., screen). We also recorded the name of the applied MR hardware.

2.3.5. *Task-technology fit*

Coders assessed the task-technology fit of the MRR program relative to each outcome, coding sources as either high or low task-technology fit. This decision was made based on the characteristics of the MRR program and the measured outcome. For instance, an MRR program that instructed users to stretch would have low task-technology fit with the outcome of gait. Alternatively, an MRR program that provided visual feedback to reaching behaviors would have high task-technology fit with the outcome of motor control.

2.3.6. Outcome

Coders recorded the outcome associated with each effect size as balance, gait, motor control, strength/flexibility, or general. Balance included outcomes such as sway while standing on a trackpad; gait included outcomes such as the 10-m walk test; motor control included outcomes such as the peg test; strength/flexibility included outcomes such as shoulder range of motion; and general included outcomes such as assessments of functioning in daily living.

2.3.7. Participant population

Sources were coded on whether they sampled from a population of individuals with physical disabilities or able-bodied individuals. For samples of individuals with physical disabilities, we also recorded the type of physical disability (e.g., stroke).

2.3.8. Type of research design

Sources were coded on the type of research design. One group post studies included participants that underwent an MRR program alone, and participants' outcomes were measured solely after the intervention. One group pre-post studies included participants that underwent an MRR program alone, and participants' outcomes were compared before and after the intervention. Two group studies included two groups in analyses (e.g., MRR program and alternative treatment) and analyses compared differences between the two groups.

2.3.9. Type of control group

Two group studies were coded on their control group. No treatment control studies utilized a control group that did not receive an intervention and a treatment group that received a MR intervention; unequal treatment control studies utilized a control group that received an intervention and a treatment group received the same intervention with an additional MR intervention; and equal treatment control studies utilized a control group that received an intervention that was equal in time to the MR intervention that the treatment group received.

2.4. Primary analyses

To begin, we calculated indices of publication biases, including fail-safe k, Egger's test, trim-and-fill method, and weight-function model analysis. These analyses were conducted using Comprehensive *Meta*-Analaysis V3 and the Metafor package in R 3.5.1. A summary of these results is provided below, and Supplemental Material B provides their full reporting. We also calculated indices of outliers and influential cases, including standardized residuals, Cook's distance, covariance rations, and others. We primarily used standardized residuals to identify outliers, which are summarized below and fully reported in Supplemental Material B.

We calculated our primary results using a random-effects model with effect sizes weighted by sample size in Comprehensive Meta-Analysis V3 adopting the Hedges and Olkin approach. Random-effects meta-analyses are more resilient to outlier studies than fixed-effects meta-analyses, and it provides more accurate estimates across most situations (Borenstein et al., 2021). Comprehensive Meta-Analysis V3 was used to convert effect sizes into a common metric, which enables the current meta-analysis to include an array of effect sizes reported in primary sources. We did not include effect sizes that represent the relation of more than two variables (e.g., ANCOVA) due to possible statistical biases (Boxer et al., 2015; Ferguson, 2015). Pre-post correlations are required to include one-group pre-post studies in meta-analyses as repeated measures designs, but few authors report these correlations. We treated these designs as independent groups in the primary analyses to avoid biasing effects of assuming the strength of pre-post correlations, and our sensitivity analyses includes results when assuming a pre-post correlation of .70 (described below). Our results are reported as Cohen's d, which represents the mean difference between the analyzed comparisons (e.g., treatment vs. control group). A Cohen's d of one indicates that the means of two comparisons were one standard deviation apart.

Multiple effect sizes reported in a single study were aggregated together via an approach recommended by Borenstein et al. (2021). Borenstein et al.'s (2021) approach incorporates the estimated average intercorrelations of outcomes in the aggregation formula, which can more appropriately address the effect sizes' covariance. Because many studies comparing group differences (e.g., treatment vs. control) do not report outcome correlations, Borenstein et al. (2021) suggest calculating multiple sets of meta-analytic estimates with different estimated average intercorrelations. As discussed in our sensitivity analysis section below, we calculated effects based on an estimated average intercorrelation of .00, .25, 0.50, 0.75, and 1.00. Because most studies reported multiple outcomes intended to represent the same or very similar physical abilities (e.g., gait), we believed that the most appropriate estimated average correlation would meet a strength expected in tests of convergent validity. Likewise, Borenstein et al. (2021) considers larger assumed correlations to produce more conservative estimates, which are typically preferred when assumptions must be made regarding the nature of effects. Therefore, the primary text reports all effects when estimating an average outcome intercorrelation of 0.75.

We did not perform any corrections for unreliability for two reasons. Most studies did not report internal consistency estimates, which would make it difficult to accurately calculate an estimate that would adequately represent all studies. Likewise, many authors have recommended against correcting for unreliability when effect sizes represent objective measurements due to practical implications of identifying the relative effectiveness of interventions in their original units of measurement (Harrison et al., 2006; Howard & Gutworth, 2020; Jones et al., 2016).

To compare source characteristics, we calculated meta-analytic estimates separated by each condition of the source characteristic. We calculated a series of meta-regressions in Comprehensive Meta-Analysis V3 to obtain statistical significance tests for these comparisons, wherein each source characteristic was the sole predictor in each. If the associated confidence interval excludes zero, then the characteristic is considered to significantly influence MRR program effectiveness. We assessed each predictor separately due to concerns about statistical power when including multiple predictors in meta-regressions (Schmidt, 2017; Viechtbauer et al., 2015), but our sensitivity analyses include a meta-regression with all predictors together.

2.5. Sensitivity analyses

Our primary meta-analytic results were calculated by aggregating multiple effect sizes from the same study together using the approach of Borenstein et al. (2021), and we report this analysis in the primary text due to precedence and familiarity. Two alternative approaches are to average effect sizes together as well as three-level meta-analysis, which identifies dependences within and across studies to aggregate effect sizes (Cheung, 2015; Jak, 2015). We conducted the former analysis using Comprehensive Meta-Analysis V3 and the latter analysis using the metaSEM package in R 3.5.1. Supplemental Material C provides the former results, whereas Supplemental Material D provides the latter results. All inferences were consistent between our primary results and alternative results. Further, pre-post correlations are required to include one-group pre-post studies in meta-analyses. Our primary analyses treat these effects as independent; however, we also performed sensitivity analyses using Comprehensive Meta-Analysis V3 and assuming a pre-post correlation of 0.70, which is a typical value reported in rehabilitation research (Myers et al., 2015; Paul et al., 2007; Storm et al., 2020). These analyses are also included in Supplemental Material D, and they again do not produce any differing inferences from our primary

Table 1
Primary meta-analytic findings.

	# of Sources	K	n	d	95%C.I.	z-value	Sig.	I^2	Meta-Reg. 95% C.I.
1.) Total	27	29	398	.60	.39, .81	5.56	<.01	28.48	–
2a.) One Group	17	19	149	.80	.42, 1.07	5.70	<.01	30.88	-.89, −.12
2b.) Two Group	10	10	249	.22	-.05, .49	1.59	.11	0.00	
2b.1.) No Treatment	2	2	45	.11	-.56, .77	.31	.76	0.00	All
2b.2.) Unequal Treat.	3	3	117	.06	-.32, .43	.29	.77	0.00	95% C.I. Include 0
2b.3.) Equal Treat.	5	5	87	.51	.05, .94	2.18	.03	0.00	
3a.) Aug. Virtuality	14	15	235	.64	.31, .98	3.74	<.01	47.91	-.43, .44
3b.) Aug. Reality	13	14	163	.57	.31, .83	4.24	<.01	0.00	
4a.) Rehab. Specific	20	21	309	.54	.33, .76	4.90	<.01	18.44	-.39, .63
4b.) Gen. Purpose	7	8	89	.74	.18, 1.31	2.58	.01	50.11	
5a.) W/Game Ele.	19	21	211	.68	.42, .94	5.14	<.01	25.77	-.20, .68
5b.) W/o Game Ele.	8	8	187	.42	.08, .77	2.41	.02	26.34	
6a.) HMD	4	4	41	.51	-.10, 1.12	1.65	.10	37.56	-.48, .74
6b.) Screen	23	25	357	.62	.39, .85	5.27	<.01	29.71	
7a.) Low TTF	12	13	278	.45	.11, .80	2.59	.01	42.79	-.52, .31
7b.) High TTF	26	27	440	.56	.32, .79	4.65	<.01	41.49	
8a.) Balance	10	10	114	.48	.21, .75	3.42	<.01	0.00	All
8b.) Gait	14	15	273	.62	.32, .92	4.03	<.01	45.17	95% C.I. Include 0
8c.) Motor Control	8	9	79	.61	.25, .97	3.30	<.01	13.65	
8d.) Strength/Flex	5	5	135	.56	-.43, 1.54	1.10	.27	78.69	
8e.) General	12	12	255	.42	.18, .67	3.37	<.01	16.71	

k = Number of Studies; n = Total Sample Size; d = Standardized Mean Difference; 95%C.I. = Confidence Interval of Standardized Mean Difference; z-value = Z-Value of Standardized Mean Difference; Sig. = *P*-Value of Standardized Mean Difference; Meta-Reg. 95%C.I. = Confidence Interval of Beta Coefficient Representing Difference of Effects Between Conditions.

analyses. Likewise, our primary results were calculated by aggregating studies together via Borenstein et al.'s (2021) approach with an assumed inter-outcome correlation of 0.75. As recommended by Borenstein et al. (2021), Supplemental Material C also includes a replication of our results with assuming other inter-outcome correlations (0.00, 0.25, 0.50, 0.75, and 1.00). Once again, these results do not produce differing inferences from our primary analyses.

We test most of our hypotheses by analyzing together studies that utilize one group pre-post designs and studies that utilize two group designs (Table 1), as such an approach can test whether certain MRR program attributes systematically produce larger effects. Because one group pre-post designs test the change produced by an intervention and two group designs test the relative change produced by an intervention, some researchers may be wary of analyzing these two designs together. For this reason, we also provide our results separated by studies that utilized one group pre-post designs (Table 2) and studies that utilized two group designs (Table 3). These analyses do not produce significantly different inferences from our primary analyses, but we do discuss some emergent trends from these analyses in our discussion.

We studied both augmented reality and augmented virtuality together under the scope of MR, but some authors may be only interested in a single medium. Supplemental Material C provides a replication of all results when only including studies that tested augmented reality and not augmented virtuality – and vice versa. Also, the meta-regression analyses provided in the primary text assessed the effect of each predictor individually. Supplemental Material C provides a meta-regression of all predictors together, which produced largely the same results with two additional statistically significant effects. Our discussion highlights that these additional findings may be promising directions for future research. Together, all sensitivity analyses largely

Table 2
Analysis of all results restricted to two group design studies alone.

	# of Sources	k	n	d	95%C.I.	z-value	Sig.	I^2	Meta-Reg. 95% C.I.
1.) Total	10	10	249	.22	-.05, .49	1.59	.11	0.00	–
2a.) One Group	–	–	–	–	–	–	–	–	–
2b.) Two Group	10	10	249	.22	-.05, .49	1.59	.11	0.00	
2b.1.) No Treatment	2	2	45	.11	-.56, .77	.31	.76	0.00	All
2b.2.) Unequal Treat.	3	3	117	.06	-.32, .43	.29	.77	0.00	95% C.I. Include 0
2b.3.) Equal Treat.	5	5	87	.51	.05, .94	2.18	.03	0.00	
3a.) Aug. Virtuality	5	5	159	.11	-.22, .44	.64	.52	0.00	-.24, .89
3b.) Aug. Reality	5	5	90	.44	-.03, .90	1.85	.07	0.00	
4a.) Rehab. Specific	8	8	211	.26	-.02, .55	1.80	.07	0.00	−1.14, .44
4b.) Gen. Purpose	2	2	38	-.08	-.82, .65	-.22	.83	0.00	
5a.) W/Game Ele.	5	5	90	.37	-.11, .84	1.52	.13	0.00	-.35, .80
5b.) W/o Game Ele.	5	5	159	.15	-.18, .47	.89	.37	0.00	
6a.) HMD	–	–	–	–	–	–	–	–	–
6b.) Screen	10	10	249	.22	-.05, .49	1.59	.11	0.00	
7a.) Low TTF	6	6	197	.21	-.12, .53	1.26	.21	0.00	−2.03, .45
7b.) High TTF	9	9	231	.18	-.18, .54	.97	.33	33.92	
8a.) Balance	2	2	45	.49	-.18, 1.16	1.44	.15	0.00	Strength vs. Balance, Gait, and General Exclude 0
8b.) Gait	6	6	197	.19	-.08, .47	1.39	.17	0.00	
8c.) Motor Control	3	3	44	.29	-.29, .87	.98	.33	0.00	
8d.) Strength/Flex	3	3	122	-.18	-.84, .48	-.54	.59	25.68	
8e.) General	6	6	202	.26	-.04, .57	1.73	.08	10.32	

k = Number of Studies; n = Total Sample Size; d = Standardized Mean Difference; 95%C.I. = Confidence Interval of Standardized Mean Difference; z-value = Z-Value of Standardized Mean Difference; Sig. = *P*-Value of Standardized Mean Difference; Meta-Reg. 95%C.I. = Confidence Interval of Beta Coefficient Representing Difference of Effects Between Conditions.

Table 3
Analysis of all results restricted to one group design studies alone.

	# of Sources	k	n	d	95%C.I.	z-value	Sig.	I^2	Meta-Reg. 95% C.I.
1.) Total	17	19	149	.80	.42, 1.07	5.70	<.01	30.88	–
2a.) One Group	17	19	149	.80	.42, 1.07	5.70	<.01	30.88	–
2b.) Two Group	–	–	–	–	–	–	–	–	
3a.) Aug. Virtuality	9	10	76	.90	.50, 1.30	4.42	<.01	36.90	-.75, .37
3b.) Aug. Reality	8	9	73	.70	.30, 1.09	3.48	<.01	28.66	
4a.) Rehab. Specific	12	13	98	.71	.41, 1.01	4.61	<.01	27.08	-.35, .91
4b.) Gen. Purpose	5	6	51	1.04	.43, 1.66	3.31	<.01	40.97	
5a.) W/Game Ele.	14	16	121	.78	.47, 1.09	4.95	<.01	33.04	-.84, .63
5b.) W/o Game Ele.	3	3	28	.90	.20, 1.61	2.52	.01	42.34	
6a.) HMD	4	4	41	.51	-.10, 1.12	1.65	.10	37.56	-.26, 1.03
6b.) Screen	13	15	108	.88	.57, 1.18	5.67	<.01	26.60	
7a.) Low TTF	6	7	81	.81	.16, 1.46	2.43	.02	62.11	-.57, .52
7b.) High TTF	17	18	209	.70	.47, .93	5.91	<.01	10.71	
8a.) Balance	8	8	107	.48	.18, .78	3.11	<.01	0.00	Balance vs. Strength Exclude 0
8b.) Gait	8	9	117	.98	.59, 1.37	4.89	<.01	31.03	
8c.) Motor Control	5	6	42	.80	.28, 1.32	3.00	<.01	33.14	
8d.) Strength/Flex	2	2	13	1.37	.63, 2.12	3.61	<.01	0.00	
8e.) General	6	6	83	.65	.29, 1.00	3.57	<.01	0.00	

k = Number of Studies; n = Total Sample Size; d = Standardized Mean Difference; 95%C.I. = Confidence Interval of Standardized Mean Difference; z-value = Z-Value of Standardized Mean Difference; Sig. = *P*-Value of Standardized Mean Difference; Meta-Reg. 95%C.I. = Confidence Interval of Beta Coefficient Representing Difference of Effects Between Conditions.

replicated the primary results, providing support that our observations were due to the substantive nature of MRR programs rather than our analytic decisions.

3. Results

3.1. Meta-analytic results

We first assessed indices of outliers, influential cases, and publication bias. Our results did not indicate the presence of any outliers or influential cases, as no standardized residual was above three – among other supporting indicators. Our Egger's tests and trim-and-fill analyses indicated the presence of publication bias for many analyses. Supplemental Material B includes a funnel plot to visualize the potential biases associated with our overall analyses. While we underwent extensive efforts to obtain unpublished results on MRR programs, such as obtaining conference presentations and contacting authors, statistically significant results were nevertheless more likely to be discovered. The current results should be interpreted with this bias in mind.

Hypothesis 1 proposed that MRR programs can improve users' physical capabilities. We meta-analyzed all one group pre-post studies to assess the change in users' outcomes from undergoing an MRR program, and we found that MRR programs statistically significantly change users' outcomes by .80 of a standard deviation (k = 19, n = 149, d = 0.80, *95%C.I.* [0.42, 1.07]). Hypothesis 1 was supported. Alternatively, Hypothesis 2 proposed that MRR programs improve users' physical capabilities more than alternative interventions. We meta-analyzed all two group studies to assess differences in the interventions' changes in outcomes, and we did not find a statistically significant differences in outcomes between MRR programs and alternative interventions (k = 10, n = 249, d = 0.22, *95%C.I.* [-0.05, 0.49]). To probe this effect further, we differentiated the types of two group studies based on their comparison group. MRR programs did not produce differing outcomes to control groups in no treatment control studies (k = 2, n = 45, d = 0.11, *95%C.I.* [-0.56, 0.77]) or unequal treatment control studies (k = 3, n = 117, d = 0.06, *95%C.I.* [-0.32, 0.43]), but they did produce significantly larger effects in equal treatment control studies (k = 5, n = 87, d = 0.51, *95%C.I.* [0.05, 0.94]). Based on these findings, we consider Hypothesis 2 to be partially supported.

We also assessed whether there was a difference in effect sizes between studies utilizing augmented virtuality or augmented reality to assess Research Question 1. Effects for augmented virtuality (k = 15, n = 235, d = 0.64, *95%C.I.* [0.31, 0.98]) and augmented reality (k = 14, n = 163, d = 0.57, *95%C.I.* [0.31, 0.83]) were statistically significant, and there was no statistically significant difference in the effect sizes (B = 0.00, *S.E.* = 0.22, *95%C.I.* [-0.43, 0.44]). This finding justifies the decision to study both augmented virtuality and augmented reality under the scope of MR.

Next, we proposed several features that may influence MRR program effectiveness. Hypothesis 3 predicted that MRR programs specifically made for rehabilitation purposes would produce larger effects than those made for other purposes. While both the former (k = 21, n = 309, d = 0.54, *95%C.I.* [0.33, 0.76]) and latter (k = 8, n = 89, d = 0.74, *95%C.I.* [0.18, 1.31]) produced statistically significant effects, there was no difference in the magnitude of their effect sizes (B = 0.12, *S.E.* = 0.26, *95%C.I.* [-0.39, 0.62]). Hypothesis 3 was not supported. Hypothesis 4 predicted that MRR programs with game elements produce larger effects than those without game elements. Programs with game elements (k = 21, n = 211, d = 0.68, *95%C.I.* [0.42, 0.94]) as well as those without (k = 8, n = 187, d = 0.42, *95%C.I.* [0.08, 0.77]) both produced statistically significant effects. While the difference in their effects trended in the hypothesized direction, their difference was not statistically significant (B = 0.24, *S.E.* = 0.22, *95%C.I.* [-0.20, 0.68]). Hypothesis 4 was not supported. Hypothesis 5 proposed that MRR programs using head-mounted displays produced larger effects than those utilizing screens. MRR programs using head-mounted displays (k = 4, n = 41, d = 0.51, *95%C.I.* [-0.10, 1.12]) and those using screens (k = 25, n = 357, d = 0.62, *95%C.I.* [0.39, 0.85]) both produced comparable effects, and the difference was again not statistically significant (B = 0.13, *S.E.* = 0.31, *95% C.I.* [-0.48, 0.74]). Hypothesis 5 was not supported. Lastly, Hypothesis 6 proposed that MRR programs with high task-technology fit would produce larger effects than those with low task-technology fit. Both MRR programs with high (k = 27, n = 440, d = 0.56, *95%C.I.* [0.32, 0.79]) and low (k = 13, n = 278, d = 0.45, *95%C.I.* [0.11, 0.80]) task-technology fit produced statistically significant results, but their difference was not statistically significant (B = −0.10, *S.E.* = 0.21, *95%C.I.* [-0.52, 0.31]). Hypothesis 6 was not supported.

3.2. Systematic literature review results

Table 4 provides the results of our systematic literature review. The first goal was to determine which MRR program characteristics are studied more often (Research Question 2). More studies were conducted on augmented reality (k = 84, 72%) than augmented virtuality (k = 31,

Table 4
Primary systematic literature review findings.

	Research Design			
	Two Group	One Group, Pre-Post	One Group, Post	Total
1.) Total	18 (16%)	25 (22%)	71 (62%)	114
2a.) Aug. Virtuality	5 (16%)	13 (42%)	13 (42%)	31 (27%)
2b.) Aug. Reality	13 (15%)	12 (14%)	59 (70%)	84 (73%)
3a.) Rehab. Specific	16 (15%)	19 (18%)	70 (67%)	105 (92%)
3b.) Gen. Purpose	2 (22%)	6 (66%)	1 (11%)	9 (8%)
4a.) W/Game Ele.	9 (11%)	20 (24%)	56 (66%)	85 (75%)
4b.) W/o Game Ele.	9 (31%)	5 (17%)	15 (52%)	29 (25%)
5a.) HMD	3 (12%)	5 (19%)	18 (69%)	26 (23%)
5b.) Screens	15 (17%)	20 (23%)	53 (60%)	88 (77%)
6a.) Balance	2 (13%)	9 (56%)	5 (31%)	16 (14%)
6b.) Gait	1 (13%)	3 (38%)	4 (50%)	8 (7%)
6c.) Motor Control	12 (15%)	10 (13%)	56 (72%)	78 (68%)
6d.) Strength/Flex	3 (60%)	0 (0%)	2 (40%)	5 (4%)
6e.) General	0 (0%)	3 (43%)	4 (57%)	7 (6%)
7a.) Phys. Impaired	15 (18%)	24 (29%)	43 (52%)	82 (64%)
7b.) Able Bodied	3 (6%)	2 (4%)	42 (89%)	47 (36%)

Note: Percentages listed in the first three columns represent the percent that the cell accounts for in the row total. Percentages listed in the fourth column (total) represents the percent that the cell accounts for in the column total. Number of studies in the augmented reality and augmented virtuality rows sum to greater than the total number of studies because one source studied both augmented reality and augmented virtuality in the same study. Likewise, number of studies when summing all outcome frequencies together is greater than the total number of studies because many articles investigated more than one type of outcome.

27%), rehabilitation-specific programs (k = 105, 92%) than programs for other purposes (k = 9, 8%), programs with game elements (k = 85, 75%) than without game elements (k = 29, 25%), programs using screens (k = 88, 77%) than head-mounted displays (k = 26, 23%), and studies using individuals with a physical impairment (k = 82, 64%) than able-bodied individuals (k = 47, 36%). More studies were conducted on motor control (k = 78, 68%) than all other outcomes combined (k = 36, 32%). The one group post research design (k = 71, 62%) was much more common than the one group pre-post (k = 25, 22%) or two group (k = 18, 16%) research designs.

We tested whether certain research designs were more commonly applied with specific types of programs (Research Question 3). Chi-square tests found significant relations of research design with the use of augmented reality or augmented virtuality ($\chi^2 = 10.87$, $df = 4$, $p < .01$), the use of rehabilitation-specific or general programs ($\chi^2 = 13.33$, $df = 4$, $p < .01$), the use of game elements ($\chi^2 = 6.81$, $df = 4$, $p = .03$), the use of individuals with a physical disability or able-bodied individuals ($\chi^2 = 18.49$, $df = 4$, $p < .01$), and the type of outcome ($\chi^2 = 26.62$, $df = 13$, $p < .01$). There was no significant relation of research design with or the use of head-mounted displays or screens ($\chi^2 = 0.76$, $df = 4$, $p = .68$). When further inspecting the significant results, studies on augmented reality were relatively more likely to have utilized a one group post design, whereas studies on augmented virtuality were relatively more likely to have utilized a one group pre-post design. Similarly, studies on rehabilitation-specific programs were more likely to have utilized a one group post design, whereas studies using general purpose programs were relatively more likely to have used one group pre-post designs. Studies without game elements were more likely to have used the two group research design. Studies using individuals with a physical impairment were more likely to have applied the one group pre-post and two group designs, whereas studies using able-bodied individuals were more likely to have applied a one group post design. The outcomes were relatively consistent regarding research design, but researchers studying motor control were relatively more likely to have utilized a one group post design.

In considering these results, we reassessed our hypotheses via meta-regression while controlling for research design for the MRR program characteristics that had a significant association with research design. Research design was still statistically significant in each of these analyses, but no other characteristics were statistically significant, providing further support for the consistency of the findings. Hypotheses 3, 4, 5, and 6 were still not supported.

Lastly, we recorded the type of physical impairment for samples of individuals of physical disabilities, and we also recorded the name of the applied hardware and software. The most studied samples were able-bodied individuals (k = 47, 37%), stroke survivors (k = 23, 20%), and those with motor control disabilities (k = 10, 9%). The most used hardware were custom hardware (k = 44, 38%), Microsoft Kinect (k = 43, 37%), and Microsoft HoloLens (k = 10, 9%). The most used software was custom software (k = 99, 86%), Microsoft Kinect games (k = 9, 8%), TheraMem (k = 3, 3%) and GenVirtual (k = 3, 3%). While the elements of these samples and MRR programs could be coded, there was significant variation across the studies.

4. Discussion

The current article had four primary goals: determine whether (1) MRR programs improve physical abilities, (2) MRR programs improve physical abilities more than alternative interventions, (3) certain characteristics of MRR programs influence effectiveness, and (4) certain characteristics of MRR programs and study designs are utilized more frequently than others. Via a meta-analysis and systematic literature review, we achieved these four goals.

Our meta-analysis showed that MRR programs can improve physical capabilities. On average, users' physical abilities improve by about three-fourths of a standard deviation from undergoing a MRR program. While our results did not support that MRR programs produce significantly larger effects compared to no treatment and unequal treatment control groups, they did produce significantly larger effects than equal treatment control groups. The nuance of this finding is discussed below in our methodological insights section, and it should be recognized that these findings provide initial support for the efficacy of MRR programs when compared to alternative interventions. Further, no characteristic of MRR programs systematically produced significantly larger results. Augmented reality programs produced comparable results to augmented virtuality programs; rehabilitation-specific programs produced comparable results to general-purpose programs; programs with game elements produced comparable results to programs without game elements; and programs with high task-technology fit produced comparable results to those with low task-technology fit. The similarity of outcomes was also seen when restricting analyses to one group or two group studies alone, further supporting the consistency of observed effects. These results indicate that many of the dominant research questions in the study of MRR programs may not influence effectiveness.

Several characteristics differed in their frequency of study. In rehabilitation, augmented reality is studied more frequently than augmented virtuality; rehabilitation-specific programs are studied more frequently than programs used for general purposes; programs with game elements are studied more often than programs without game elements; and programs using screens are studied more frequently than programs using HMDs. The most popular outcome was motor control followed by balance. Lastly, studies comparing the effectiveness of MRR programs relative to alternatives largely included individuals with physical disabilities, but many studies assessing the feasibility of MRR programs utilized able-bodied individuals. Each of these findings pose many implications for research and practice discussed below.

4.1. Research implications and future directions

4.1.1. Theoretical insights

We mentioned that authors rarely apply theoretical frameworks to guide the development and application of their MRR programs. While our results did not identify any characteristics of MRR programs that systematically produce improved results, they can nevertheless guide future researchers' attention towards theories that may be suitable for MRR program development.

We proposed that MRR programs produce higher fidelity and user enjoyment compared to alternative interventions. We also proposed that the studied characteristics of MRR programs are intended to enhance these natural attributes of MRR programs, increasing fidelity and user enjoyment. Our results showed that MRR program characteristics associated with entertainment (e.g., game elements, head-mounted displays) were no more effective than MRR programs with alternative characteristics (e.g., no game elements, screens), and MRR program characteristics associated with fidelity (e.g., rehabilitation-specific programs, head-mounted displays) were no more effective than MRR programs without these elements (e.g., general programs, screens). We also found that task-technology fit did not significantly influence our observed results.

It should be questioned, therefore, why MRR programs show such a robust consistency across a myriad of applications, rather than demonstrating differential effectiveness based on program characteristics. We believe that the natural fidelity and increased enjoyment produced by MRR programs are primary drivers of their effectiveness. Characteristics intended to further increase fidelity or enjoyment may struggle to do so, and/or increasing fidelity or enjoyment further than the typical MRR program may produce diminishing returns. If the case, then the inherent fidelity and enjoyment of MRR programs causes them to demonstrate consistent efficacy. We believe that this sentiment is supported by current research; many authors stress the realism and entertainment produced by MR regardless of the application, suggesting that they are relatively universal aspects of MR programs (de Souza et al., 2019; Hidayat et al., 2019; Savela et al., 2020).[4] For this reason, we discuss the theoretical associations of fidelity and enjoyment to prompt directions for future research – directions that can identify avenues to increase MRR program efficacy. Fig. 2 provides a visual representation of our theoretical proposals. We then discuss methodological and practice considerations produced by the current findings.

4.1.2. Fidelity

Two types of fidelity have been identified in modern research: physical fidelity and cognitive fidelity (Hardeman et al., 2008; Hochmitz & Yuviler-Gavish, 2011; Howard, 2017). Physical fidelity refers to the extent that the training environment appears similarly to the transfer environment, whereas cognitive fidelity refers to the extent that the training environment activates the same psychological mechanisms as the transfer environment. Both have been repeatedly linked to improved intervention outcomes (Mancuso et al., 2020; Riva et al., 2020).

MRR programs are believed to have strong physical and cognitive fidelity (Eichler et al., 2019; Gavish et al., 2011; Jeon & Kim, 2020), and identical elements theory is among the most popular theories to explain the beneficial effects of physical and cognitive fidelity (Benzing & Schmidt, 2018; Bonney et al., 2017; Nguyen et al., 2019). Identical elements theory proposes that fidelity has a positive relation with transfer effectiveness, as users identify how to perform real-world behaviors by identifying the common elements in training behaviors. Programs with strong physical and cognitive fidelity do not require as much "translation" between the training environment and the transfer environment, and users can more accurately develop correct approaches for performing behaviors. Other rehabilitation programs may instruct users to perform behaviors under circumstances rarely seen in naturalistic settings, such as reaching behaviors that are unrealistic in daily life (e.g., free of distractions). MRR programs, on the other hand, can instruct users to perform familiar behaviors in realistic environments, such as reaching for familiar objects in virtual environments that replicate common distractions, and practitioners can be more assured that the relevant physical capabilities are being developed. Moving forward, researchers may benefit from applying theories associated with fidelity, such as identical elements theory, and they should assess whether the realism of mixed-reality environments is a cause of MRR program effectiveness.

This similarity between training and transfer environments may also cause programs with strong fidelity to activate the same neurological pathways as performing the behaviors in naturalistic environments (Díez-Cirarda et al., 2017; Howard, 2017; Maggio et al., 2019). Physical rehabilitation not only involves developing the physical capabilities to perform tasks, but it also requires the reactivation of cognitive structures that may have laid dormant or been damaged since the user's physical impairment (Dong et al., 2007; Luukkainen-Markkula et al., 2009). For some users, such as stroke survivors, this reactivation may be the most important benefit of MRR programs. Current research has provided initial evidence for the benefits of neurological activation in MRR programs. Notably, Koroleva et al. (2020) demonstrated that brain-derived neurotrophic factor, a potential marker of neurorehabilitation, had a clear association with active treatment in MRR programs, and Assis et al. (2019) supported, via resting-state fMRI, that brain regions associated with motor activity had higher connectivity values after participants experienced an MRR program. Future researchers should explore this link between the mind and body in MRR programs, especially those with high fidelity.

Further, cognitive fidelity may provide an additional benefit not shared with physical fidelity. When performing activities of daily living, such as navigating around a grocery store, people must attend to multiple stimuli in the environment and cognitively regulate their behaviors in response to stimuli. Strategies to control these psychological processes (e.g., attention and thinking), known as metacognitive abilities, are often pivotal to successfully performing behaviors in realistic environments (Bivona et al., 2008; Kizony et al., 2006; Skidmore et al., 2011). MRR programs with strong cognitive fidelity can force users to attend to multiple stimuli, which enables them to simultaneously practice metacognitive abilities. Future researchers should apply theories associated with metacognition to determine ideal avenues to improve

[4] These discussed characteristics could not be studied in our meta-analysis. These characteristics appear in almost all MRR programs, preventing the comparison of studies with and without these characteristics. Likewise, these characteristics are difficult to quantify in MRR programs, as almost all MRR programs contain them to a significant extent. Thus, we discuss them as possible causes of MRR program effectiveness to investigate in future research.

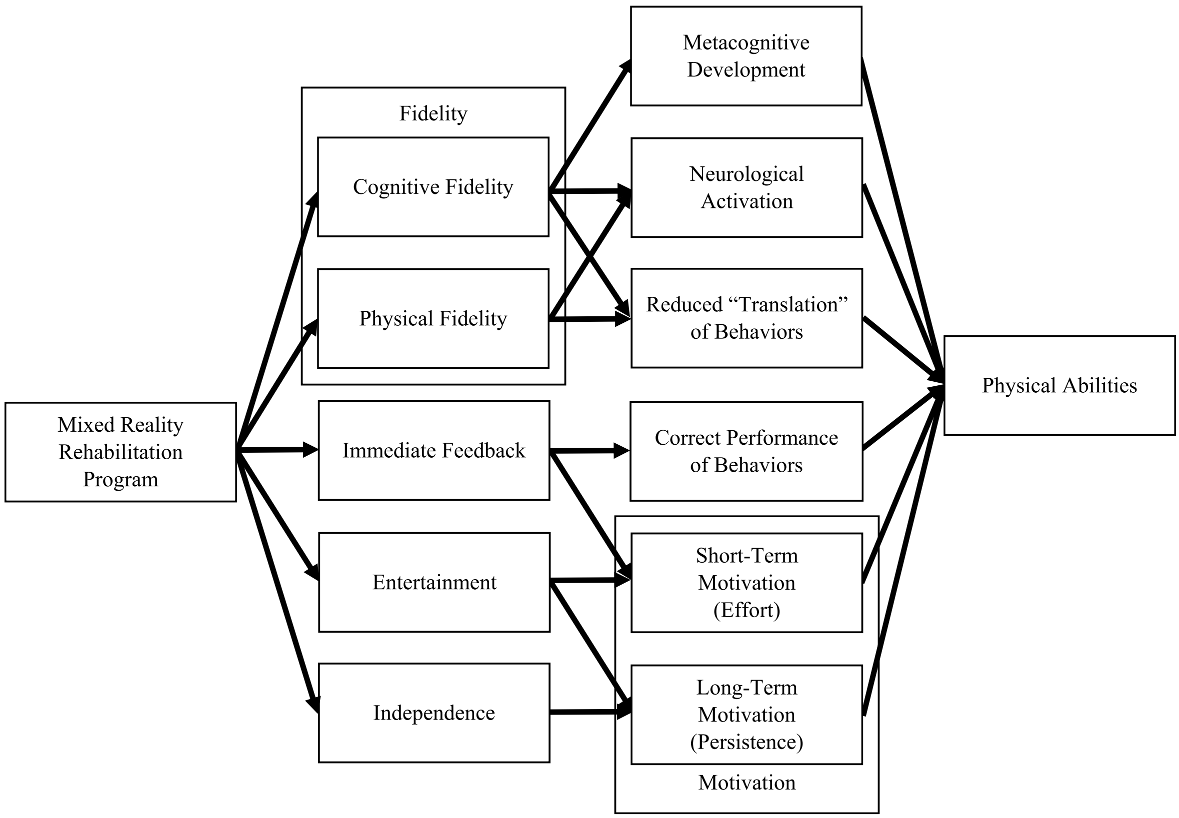

Fig. 2. Visual representation of mixed reality rehabilitation program mediating mechanisms and outcomes.

MRR program effectiveness. As discovered by Al Banna et al. (2016), the most popular theoretical perspective to study metacognition in post-stroke individuals is Crosson et al.'s (1989) three-level model of self-awareness. This model proposes that individuals' self-awareness can be hierarchically organized as: intellectual awareness, the ability to recognize that a physical ability is impaired and the implications of this impairment; emergent awareness, the ability to recognize when a problem occurs with a physical ability; and anticipatory awareness, the ability to anticipate a problem with a physical ability. Most users have already achieved the initial level of self-awareness before enrolling, but MRR programs are apt for enabling the achievement of levels two and three. Users have ample opportunities to practice behaviors in naturalistic settings, and they can begin to identify and anticipate when problems occur, such as experiencing sensory overload from surrounding stimuli. Research is needed, therefore, to support whether MRR programs can develop these important metacognitive abilities.

Together, both physical and cognitive fidelity may improve outcomes due to identical elements and neurorehabilitation, whereas cognitive fidelity may also improve outcomes due to the additional mechanism of metacognitive development. The consistency of fidelity in MRR programs may explain why differential effectiveness was not observed, but this cannot be verified until future research uses them as theoretical lenses. Therefore, future research on these mechanisms is pivotal for enhancing knowledge and development of MRR programs.

4.1.3. Motivational characteristics

Several consistent characteristics of MRR programs may increase user motivation. Among the most important is the provision of immediate feedback. Because MRR programs must include some type of image recognition to mix reality and virtuality, associated software is typically able to identify the movement of users (Cipresso et al., 2018; Yip et al., 2019). By utilizing this capability, program developers typically incorporate feedback within MRR programs, such that users' movements are detected, the movements are compared to proper rehabilitation actions, and feedback for improvement is provided (An et al., 2019; Assis et al., 2016; Dahl-Popolizio et al., 2014). This capability ensures that users are performing movements correctly, which can both improve the development of physical capabilities and reduce injury.

Some authors have also noted that immediate and personalized feedback can adhere to the recommendations of goal-setting theories, which is known to enhance user motivation (García-Bravo et al., 2020; Jeon & Kim, 2020; Lee et al., 2016). For instance, a significant amount of research has supported the usefulness of specific, measurable, achievable, realistic, and timely (SMART) goals for rehabilitation purposes (Bovend'Eerdt et al., 2009; Bowman et al., 2015). The ability to provide instantaneous feedback satisfies the SMART goal requirements of measurable and timely, which may be difficult to satisfy in alternative rehabilitation programs. Instant feedback is also an important aspect of many important elements of gamified programs, such as score and leaderboards (Jahn et al., 2021; Landers et al., 2017; Moro et al., 2019). The common inclusion of instant feedback may also explain why game elements did not have a significant impact in our meta-analysis; the ability to provide instant feedback may already satisfy a primary benefit of gamified programs, causing the addition of game elements to have a smaller impact than usual. If this is the case, then instant feedback may cause users to undergo similar processes as the inclusion of game elements, such as increased motivation. Future researchers should assess, therefore, whether the game elements of score and leaderboards have significant influence on user motivation in MRR programs. If not, then it can be supported that the instantaneous feedback provided by MRR programs effectively serve these functions.

Further, the increased excitement produced by MRR programs may cause users to remain motivated towards their rehabilitation goals when compared to alternative programs, and this excitement may even cause users to persist longer with treatment. Future researchers should assess excitement and motivation as mediators between MRR program application and user outcomes. If supported, developers may not need to add extraneous exciting elements to MRR programs, and they can instead focus on other elements that may be lacking in MRR programs. Future researchers should also assess the efficacy of MRR programs with and without specific game elements, as guided by associated taxonomies (Bedwell et al., 2012; Toda et al., 2019), to determine whether any produce additional motivation beyond the natural novelty of these programs. If none significantly contribute to excitement, then it can be strongly supported that the inherent excitement in these programs is sufficient.

Lastly, MRR programs often provide automated instruction to users based on the recognition of their movements, which allows users to perform rehabilitation activities in the home environment (Trojan et al., 2014; Viglialoro et al., 2019). This capability allows individuals to have easier access to rehabilitation, as users do not need to commute to perform rehabilitation activities. It also gives users greater autonomy and flexibility over rehabilitation activities because they can choose when and where to use the MRR programs. Many motivational theories, such as self-determination theory (Deci & Ryan, 2008), argue that autonomy is a primary predictor of motivation. The independence facilitated by MRR programs along with associated accessibility and autonomy may increase user motivation in the short-term (i.e., individual sessions) and long-term (i.e., course of entire rehabilitation). Future researchers should assess whether feelings of accessibility and autonomy are explanatory mechanisms of MRR program effectiveness. Researchers should likewise apply modern motivational theory, including self-determination theory, to test whether the motivating aspects of these programs (e.g., autonomy) can be leveraged to produce more beneficial outcomes.

4.1.4. Methodological insights

Our results show that most research on MRR programs utilizes the one group post design, which can only test whether the MRR program is feasible or poses severe concerns. It cannot determine whether the MRR program improves user outcomes or user outcomes beyond other interventions (Goodwin & Goodwin, 2016; Leong & Austin, 2006). Likewise, we also show that most studies utilizing two group or one group pre-post research designs obtain samples of individuals with physical impairments, whereas approximately half of studies utilizing one group post designs obtain samples of able-bodied individuals. Such trends indicate that the study of MRR programs currently exist in early stages, as researchers are pilot testing their MRR programs rather than robustly assessing program effectiveness (Galna et al., 2014). It is encouraging, therefore, to observe that MRR programs are indeed successful at improving user physical capabilities despite being within early stages of investigation.

Moving forward, however, researchers should avoid studies with able-bodied individuals alone. MRR programs deemed to be safe for able-bodied individuals cannot be assured to be safe for those with physical impairments, and results found with able-bodied individuals cannot be assumed to generalize to those with physical impairments. If such samples are utilized, researchers should replicate their results using a sample of individuals with physical disabilities.

Similarly, researchers should adopt more robust research designs, such as two-group and one group pre-post studies. For instance, one group pre-post studies can still be utilized to assess whether MRR programs are feasible or pose severe concerns, while simultaneously assessing whether the program is able to improve physical capabilities (Goodwin & Goodwin, 2016; Leong & Austin, 2006). By moving towards these more robust designs, researchers can provide stronger evidence for the effectiveness of MRR programs. Perhaps more importantly, researchers can ensure that certain threats to internal validity are addressed in their studies. For instance, one-group designs – whether post or pre-post – cannot account for history of maturation effects (Goodwin & Goodwin, 2016; Leong & Austin, 2006). Researchers should utilize two-group pre-post designs whenever possible to address concerns with one group designs in addition to determining whether MRR programs are more effective than alternatives.

We observed that MRR programs do not produce significantly larger effects than no treatment and unequal treatment intervention, but these programs do produce significantly larger effects than equal treatment interventions. When conducting physical rehabilitation research including participants with a physical impairment,[5] researchers are expected to provide a standard of care. In no treatment control studies, participants may not require an intervention to reach adequate physical abilities, and MR interventions may not benefit these participants because they do not need their physical abilities improved. In unequal treatment control studies, participants in the control group receive an intervention, whereas participants in the treatment group receive the same intervention as well as the MRR program. The participants may be already receiving an intervention to reach adequate physical abilities, and MR interventions may not benefit these participants because their physical abilities may have already reached an adequate level. In equal treatment control studies, however, participants need an intervention to achieve adequate physical abilities, and our effects show that MR interventions are more effective at enabling participants to reach this level. Based on these findings, currently tested MRR programs may not significantly benefit participants who have already reached an adequate level of physical capabilities, but they may benefit participants who do not have adequate physical capabilities more so than traditional interventions. Thus, future research is needed to differentiate and test these two different research questions – whether MRR programs are indeed better at enabling participants to reach a sufficient level of physical abilities and whether these programs can improve physical abilities after participants have reached adequate levels.

Researchers have primarily focused on motor control and, to a lesser extent, balance in MRR programs. This is likely because the most common approach for MRR programs is to present the MR environment on a screen (e.g., television), which lends itself to rehabilitation activities performed in a stationary location, such as reaching or squatting. Future research should broaden studied outcomes in two manners. First, with the increasing availability of MR head-mounted displays, researchers should develop programs to enhance physical abilities that benefit from the mobility of head-mounted displays, such as gait. Some authors have begun to utilize these devices for these purposes, showing great potential in their application. Second, because we showed that task-technology fit did not significantly impact the observed outcomes, it is possible that MRR programs are able to develop more physical abilities than traditionally assumed, but many authors do not measure a broad range of abilities in assessing the efficacy of programs. An MRR program to develop balance abilities, for example, may likewise be effective in developing gait abilities, but such possibilities are presently unknown given current research designs. Therefore, we urge future researchers to assess a broader range of outcomes.

4.2. Practice implications

Continued development of MRR programs greatly enhances access to rehabilitation for populations that may otherwise be unable to seek long-term treatment for physical injury or disability. For example, access to rehabilitation services is more complex in rural or remote areas (Dew et al., 2013; Wilson et al., 2009), which is an issue that is present on a global scale (Landry et al., 2009; Roots et al., 2013; Winn et al., 2014). Specifically, The World Health Organization (WHO, 2011) estimates that over 15% of the global population lives with a disability with over 75% of those individuals residing in a low to middle income countries. As research continues to identify characteristics and techniques to improve MRR program outcomes, this can improve health outcomes for

[5] No study included in the meta-analysis included able-bodied individuals as participants.

individuals in more remote areas of the world.

MRR programs may also become a lower cost alternative to traditional treatment programs as many of these programs do not require an in-person therapist to conduct treatment, which vastly enhances accessibility for individuals across all levels of socioeconomic status. Poverty is also associated as a barrier to physical rehabilitation for individuals due to associations with limited social support and lack of transportation (Bright et al., 2018). Physical rehabilitation clinics and home health agencies should specifically focus on integrating MMR programs into the treatment components of post-surgical care and physical rehabilitation to enhance health outcomes for individuals within this specific population.

Regarding the motivational benefits provided by the gaming elements of the MRR programs, there is a chance that the increased motivation towards meeting treatment goals may extend into additional aspects of individuals' lives. The impact of physical disability often extends beyond physical limitations by impacting relationships, mental health, and well-being (Tough et al., 2017). When individuals are exposed to a gaming environment provided by the MRR programs that enhances motivation towards treatment, they may begin to experience motivation towards achieving health and healing in other aspects of life. As individuals begin to experience the benefits of improved physical health outcomes, it is expected that mental and social health outcomes will also indirectly improve as research shows that these areas of a person's life are highly interconnected (Ohrnberger et al., 2017; Repetti et al., 2002). Thus, as individuals begin to feel stronger both mentally and physically, it is expected that the motivation will be extended to enhanced social relationships and professional productivity, ultimately contributing greatly to regional, national, and even global economic health.

4.3. Limitations

The quality of any meta-analysis is dependent upon the included articles. Studies on MRR programs were shown to use differing research approaches, and we provided our results separated by research design in Tables 2 and 3 None of our separated results differed from the aggregated results presented in Table 1 regarding the program characteristics, supporting the consistency of our observed effects. Analytical decisions can also sway meta-analytic results. We provide ample supplemental materials wherein we recalculate our results using alternative analytic approaches (Supplemental Material C and D). As our results were consistent across these recalculations, we provide firm support that our results reflect substantive effects of MRR programs rather than analytical biases. We further implemented standard search procedures to discover published and unpublished works, but meta-analyses always benefit from the inclusion of more sources to address publication biases. Future researchers should replicate the current results once additional studies have been performed. This recommendation is consistent with prior work on virtual reality rehabilitation programs, wherein early meta-analyses with modest sample sizes (Saposnik et al., 2011) laid the foundation for subsequent works with much larger sample sizes (Howard, 2017; Karamians et al., 2020). The current meta-analysis and systematic literature review should be seen as the first of many such studies on MRR programs.

Meta-analyses obtain estimates of average effects by aggregating studies. Meta-analytic results are not estimates of ideal interventions or reflective of interventions' potential. Attempts to maximize the potential of MRR programs in individual studies still pose great merit, and the current meta-analytic results are not the ultimate indicator of MRR program effectiveness.

Our meta-analysis was specifically focused on overall effectiveness and design characteristics of MRR programs, and we included some contextual considerations to better understand these characteristics (e.g., outcomes). The effectiveness of a MRR program is influenced by more than design characteristics alone, however, as it is also influenced by contextual and user characteristics. Incorporating these characteristics is outside the scope of the current meta-analysis, but future researchers should assess the extent that these characteristics influence MRR program outcomes. Notably, we argued that current MRR programs may not be particularly effective for improving physical abilities of participants that already possess adequate physical abilities. Future research should more directly test this notion and assess whether participants' physical abilities prior to undergoing a MRR program predicts effectiveness. Likewise, we argued that MRR programs may produce improved outcomes due to user excitement, enjoyment, and motivation. A users' interest in MR may therefore predict MRR program outcomes, suggesting that this direction of future research may be quite fruitful.

Lastly, the outcomes studied in the current meta-analysis were able to be coded within the five categories derived from prior research, but there were some differences among the measures within each category. The physical capabilities to perform a sitting balance test is different than the physical capabilities to perform a standing balance test, and future researchers should investigate the effect of specific measures when assessing the effectiveness of physical rehabilitation interventions. Certain outcome measures may be apt for identifying changes in physical abilities, whereas others may be relatively ineffective. Similar sentiments could be expressed regarding the participant populations and applied MRR programs, and future researchers should build upon the present investigation by focusing on attributes of MRR programs and further differentiating their characteristics. Supplemental Material A could serve as a beneficial starting point to conduct these additional investigations.

5. Conclusion

MRR programs improve user outcomes, but many design characteristics of MRR programs do not influence effectiveness. Characteristics consistent across most MRR programs may be the cause of their consistent effectiveness, and we detailed theoretical considerations regarding fidelity and enjoyment that can guide researchers towards future directions. These directions can identify theories that enable researchers to understand physical and cognitive mechanisms involved in effective MRR programs resulting in frameworks akin to Fig. 2. These frameworks can serve as guides for future development and application of superior MRR programs. Thus, the current article should be viewed as the beginning of many studies on MRR.

Credit author statement

The primary author was primarily responsible for all aspects of the present submission. This included the search procedures, coding, analyses, and writing. The secondary author was also responsible for the coding and writing.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.chb.2022.107197.

References

Al Banna, M., Redha, N. A., Abdulla, F., Nair, B., & Donnellan, C. (2016). Metacognitive function poststroke: A review of definition and assessment. *Journal of Neurology, Neurosurgery & Psychiatry, 87*(2), 161–166.

Alamri, A., Cha, J., & El Saddik, A. (2010). AR-REHAB: An augmented reality framework for poststroke-patient rehabilitation. *IEEE Transactions on Instrumentation and Measurement, 59*(10), 2554–2563.

Almasi, S., Shahmoradi, L., Ansari, N. N., & Honarpishe, R. (2020). Kinect-based virtual rehabilitation for upper extremity motor recovery in chronic stroke. In *2020 international serious games symposium (ISGS)* (pp. 51–60). IEEE.

An, C. M., & Kim, D. H. (2019). Clinical application of AR system in early rehabilitation program after stroke: 2 case study. *Journal of Kansai Physical Therapy, 31*(3), 141–146.

Assis, G., Brandao, A., Corrêa, A. G. D., & Castellano, G. (2019). Evaluation of a protocol for fMRI assessment associated with augmented reality rehabilitation of stroke subjects. *SBC Journal on Interactive Systems, 10*(1), 35–42.

Assis, G. A. D., Corrêa, A. G. D., Martins, M. B. R., Pedrozo, W. G., & Lopes, R. D. D. (2016). An augmented reality system for upper-limb post-stroke motor rehabilitation: A feasibility study. *Disability and Rehabilitation: Assistive Technology, 11*(6), 521–528.

Bedwell, W. L., Pavlas, D., Heyne, K., Lazzara, E. H., & Salas, E. (2012). Toward a taxonomy linking game attributes to learning: An empirical study. *Simulation & Gaming, 43*(6), 729–760.

Benzing, V., & Schmidt, M. (2018). Exergaming for children and adolescents: Strengths, weaknesses, opportunities and threats. *Journal of Clinical Medicine, 7*(11), 422.

Berton, A., Longo, U. G., Candela, V., Fioravanti, S., Giannone, L., Arcangeli, V., & Denaro, V. (2020). Virtual reality, augmented reality, gamification, and telerehabilitation: Psychological impact on orthopedic patients' rehabilitation. *Journal of Clinical Medicine, 9*(8), 2567.

Bivona, U., Ciurli, P., Barba, C., Onder, G., Azicnuda, E. V. A., Silvestro, D., & Formisano, R. (2008). Executive function and metacognitive self-awareness after severe traumatic brain injury. *Journal of the International Neuropsychological Society, 14*(5), 862–868.

Blankevoort, C. G., Van Heuvelen, M. J., Boersma, F., Luning, H., De Jong, J., & Scherder, E. J. (2010). Review of effects of physical activity on strength, balance, mobility and ADL performance in elderly subjects with dementia. *Dementia and Geriatric Cognitive Disorders, 30*(5), 392–402.

Blomqvist, S., Seipel, S., & Engström, M. (2021). Using augmented reality technology for balance training in the older adults: A feasibility pilot study. *BMC Geriatrics, 21*(1), 1–13.

Bonney, E., Jelsma, L. D., Ferguson, G. D., & Smits-Engelsman, B. C. (2017). Learning better by repetition or variation? Is transfer at odds with task specific training? *PLoS One, 12*(3), Article e0174214.

Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (2021). *Introduction to meta-analysis*. John Wiley & Sons.

Bovend'Eerdt, T. J., Botell, R. E., & Wade, D. T. (2009). Writing SMART rehabilitation goals and achieving goal attainment scaling: A practical guide. *Clinical Rehabilitation, 23*(4), 352–361.

Bowman, J., Mogensen, L., Marsland, E., & Lannin, N. (2015). The development, content validity and inter-rater reliability of the SMART-goal evaluation method: A standardised method for evaluating clinical goals. *Australian Occupational Therapy Journal, 62*(6), 420–427.

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.

Bright, T., Wallace, S., & Kuper, H. (2018). A systematic review of access to rehabilitation for people with disabilities in low-and middle-income countries. *International Journal of Environmental Research and Public Health, 15*(10), 2165.

Carmigniani, J., & Furht, B. (2011). Augmented reality: An overview. *Handbook of Augmented Reality*, 3–46.

Carmigniani, J., Furht, B., Anisetti, M., Ceravolo, P., Damiani, E., & Ivkovic, M. (2011). Augmented reality technologies, systems and applications. *Multimedia Tools and Applications, 51*(1), 341–377.

Cervera, M. A., Soekadar, S. R., Ushiba, J., Millán, J. D. R., Liu, M., Birbaumer, N., & Garipelli, G. (2018). Brain-computer interfaces for post-stroke motor rehabilitation: A meta-analysis. *Annals of Clinical and Translational Neurology, 5*(5), 651–663.

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

Cheung, M. W. L. (2015). *Meta-analysis: A structural equation modeling approach*. John Wiley & Sons.

Cipresso, P., Giglioli, I. A. C., Raya, M. A., & Riva, G. (2018). The past, present, and future of virtual and augmented reality research: A network and cluster analysis of the literature. *Frontiers in Psychology, 9*, 2086.

Collins, L. M., Nahum-Shani, I., & Almirall, D. (2014). Optimization of behavioral dynamic treatment regimens based on the sequential, multiple assignment, randomized trial (SMART). *Clinical Trials, 11*(4), 426–434.

Crosson, B., Barco, P. P., Velozo, C. A., Bolesta, M. M., Cooper, P. V., Werts, D., & Brobeck, T. C. (1989). Awareness and compensation in postacute head injury rehabilitation. *The Journal of Head Trauma Rehabilitation, 4*(3), 46–54.

Da Gama, A., Fallavollita, P., Teichrieb, V., & Navab, N. (2015). Motor rehabilitation using kinect: A systematic review. *Games for Health Journal, 4*(2), 123–135.

Dahl-Popolizio, S., Loman, J., & Cordes, C. C. (2014). Comparing outcomes of Kinect videogame-based occupational/physical therapy versus usual care. *Games for Health: Research, Development, and Clinical Applications, 3*(3), 157–161.

Debarba, H. G., de Oliveira, M. E., Lädermann, A., Chagué, S., & Charbonnier, C. (2018). Augmented reality visualization of joint movements for physical examination and rehabilitation. In *2018 IEEE conference on virtual reality and 3D user interfaces (VR)* (pp. 537–538). IEEE.

Deci, E. L., & Ryan, R. M. (2008). Self-determination theory: A macrotheory of human motivation, development, and health. *Canadian Psychology/Psychologie canadienne, 49*(3), 182.

Dew, A., Bulkeley, K., Veitch, C., Bundy, A., Gallego, G., Lincoln, M., & Griffiths, S. (2013). Addressing the barriers to accessing therapy services in rural and remote areas. *Disability & Rehabilitation, 35*(18), 1564–1570.

Díez-Cirarda, M., Ojeda, N., Peña, J., Cabrera-Zubizarreta, A., Lucas-Jiménez, O., Gómez-Esteban, J. C., & Ibarretxe-Bilbao, N. (2017). Increased brain connectivity and activation after cognitive rehabilitation in Parkinson's disease: A randomized controlled trial. *Brain Imaging and Behavior, 11*(6), 1640–1651.

Dong, Y., Winstein, C. J., Albistegui-DuBois, R., & Dobkin, B. H. (2007). Evolution of FMRI activation in the perilesional primary motor cortex and cerebellum with rehabilitation training-related motor gains after stroke: A pilot study. *Neurorehabilitation and Neural Repair, 21*(5), 412–428.

Dorado, J., del Toro, X., Santofimia, M. J., Parreño, A., Cantarero, R., Rubio, A., & Lopez, J. C. (2019). A computer-vision-based system for at-home rheumatoid arthritis rehabilitation. *International Journal of Distributed Sensor Networks, 15*(9), 1550147719875649.

Eichler, S., Salzwedel, A., Rabe, S., Mueller, S., Mayer, F., Wochatz, M., … Völler, H. (2019). The effectiveness of telerehabilitation as a supplement to rehabilitation in patients after total knee or hip replacement: Randomized controlled trial. *JMIR Rehabilitation and Assistive Technologies, 6*(2), Article e14236.

Feng, Y., & Xie, Q. (2019). Demystifying novelty effects: An analysis of consumer responses to YouTube videos featuring augmented reality out-of-home advertising campaigns. *Journal of Current Issues & Research in Advertising, 40*(1), 36–53.

Ferguson, C. J. (2015). Pay no attention to that data behind the curtain: On angry birds, happy children, scholarly squabbles, publication bias, and why betas rule metas. *Perspectives on Psychological Science, 10*(5), 683–691.

Fritz, J. M., Rhon, D. I., Teyhen, D. S., Kean, J., Vanneman, M. E., Garland, E. L., & Greene, T. H. (2020). A sequential multiple-assignment randomized trial (SMART) for stepped care management of low back pain in the military health system: A trial protocol. *Pain Medicine, 21*(Supplement_2), S73–S82.

Galna, B., Jackson, D., Schofield, G., McNaney, R., Webster, M., Barry, G., … Rochester, L. (2014). Retraining function in people with Parkinson's disease using the microsoft kinect: Game design and pilot testing. *Journal of NeuroEngineering and Rehabilitation, 11*(1), 1–12.

García-Bravo, S., Cano-de-la-Cuerda, R., Domínguez-Paniagua, J., Campuzano-Ruiz, R., Barreñada-Copete, E., López-Navas, M. J., & Cuesta-Gómez, A. (2020). Effects of virtual reality on cardiac rehabilitation programs for ischemic heart disease: A randomized pilot clinical trial. *International Journal of Environmental Research and Public Health, 17*(22), 8472.

Garzón, J., Pavón, J., & Baldiris, S. (2019). Systematic review and meta-analysis of augmented reality in educational settings. *Virtual Reality, 23*(4), 447–459.

Gavish, N., Gutierrez, T., Webel, S., Rodriguez, J., & Tecchia, F. (2011). Design guidelines for the development of virtual reality and augmented reality training systems for maintenance and assembly tasks. *BIO web of conferences, 1*, 29 (EDP Sciences).

Gonze, B. D. B., Padovani, R. D. C., Simoes, M. D. S., Lauria, V., Proença, N. L., Sperandio, E. F., & Dourado, V. Z. (2020). Use of a smartphone app to increase physical activity levels in insufficiently active adults: Feasibility sequential multiple assignment randomized trial (SMART). *JMIR Research Protocols, 9*(10), Article e14322.

Goodhue, D. L., & Thompson, R. L. (1995). Task-technology fit and individual performance. *MIS Quarterly*, 213–236.

Goodwin, C. J., & Goodwin, K. A. (2016). *Research in psychology methods and design*. John Wiley & Sons.

Gorman, C., & Gustafsson, L. (2020). The use of augmented reality for rehabilitation after stroke: A narrative review. *Disability and Rehabilitation: Assistive Technology*, 1–9.

Goršič, M., Cikajlo, I., Goljar, N., & Novak, D. (2017). A multisession evaluation of an adaptive competitive arm rehabilitation game. *Journal of NeuroEngineering and Rehabilitation, 14*(1), 1–15.

Guinet, A. L., Bouyer, G., Otmane, S., & Desailly, E. (2019). Reliability of the head tracking measured by Microsoft HoloLens during different walking conditions. *Computer Methods in Biomechanics and Biomedical Engineering, 22*(sup1), S169–S171.

Guinet, A. L., Bouyer, G., Otmane, S., & Desailly, E. (2021). Validity of HoloLens augmented reality head mounted display for measuring gait parameters in healthy adults and children with cerebral palsy. *Sensors, 21*(8), 2697.

Hardeman, W., Michie, S., Fanshawe, T., Prevost, A. T., Mcloughlin, K., & Kinmonth, A. L. (2008). Fidelity of delivery of a physical activity intervention: Predictors and consequences. *Psychology and Health, 23*(1), 11–24.

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.

Held, J. P. O., Yu, K., Pyles, C., Veerbeek, J. M., Bork, F., Heining, S. M., & Luft, A. R. (2020). Augmented reality–based rehabilitation of gait impairments: Case report. *JMIR mHealth and uHealth, 8*(5), Article e17804.

Hidayat, I. K., Rini, D. R., Novica, D. R., & Daud, K. A. M. (2019). Implementing augmented reality and gamification in tourism. *International Journal of Technology Management and Information System, 1*(1), 39–45.

Hildebrand, M. W., Host, H. H., Binder, E. F., Carpenter, B., Freedland, K. E., Morrow-Howell, N., & Lenze, E. J. (2012). Measuring treatment fidelity in a rehabilitation intervention study. *American Journal of Physical Medicine & Rehabilitation/Association of Academic Physiatrists, 91*(8), 715.

Hochmitz, I., & Yuviler-Gavish, N. (2011). Physical fidelity versus cognitive fidelity training in procedural skills acquisition. *Human Factors, 53*(5), 489–501.

Hoda, M., Hoda, Y., Alamri, A., Hafidh, B., & Saddik, A. E. (2015). A novel study on natural robotic rehabilitation exergames using the unaffected arm of stroke patients. *International Journal of Distributed Sensor Networks, 11*(6), 590584.

Hoermann, S., Ferreira dos Santos, L., Morkisch, N., Jettkowski, K., Sillis, M., Devan, H., & Cutfield, N. J. (2017). Computerised mirror therapy with augmented reflection technology for early stroke rehabilitation: Clinical feasibility and integration as an adjunct therapy. *Disability & Rehabilitation, 39*(15), 1503–1514.

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., & Gutworth, M. B. (2020). A meta-analysis of virtual reality training programs for social skill development. *Computers & Education, 144*, 103707.
Howard, M. C., & Jacobs, R. R. (2016). The multiphase optimization strategy (MOST) and the sequential multiple assignment randomized trial (SMART): Two novel evaluation methods for developing optimal training programs. *Journal of Organizational Behavior, 37*(8), 1246–1270.
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, 56*(6), 103134.
Huang, B., & Hew, K. F. (2015). Do points, badges and leaderboard increase learning and activity: A quasi-experiment on the effects of gamification. In *Proceedings of the 23rd international conference on computers in education* (pp. 275–280).
Hunter, J. E., & Schmidt, F. L. (2004). *Methods of meta-analysis: Correcting error and bias in research findings*. Sage.
Ioannides, M., Magnenat-Thalmann, N., & Papagiannakis, G. (Eds.). (2017). *Mixed reality and gamification for cultural heritage* (Vol. 2). Cham, Switzerland: Springer.
Jahn, K., Kordyaka, B., Machulska, A., Eiler, T. J., Gruenewald, A., Klucken, T., & Niehaves, B. (2021). Individualized gamification elements: The impact of avatar and feedback design on reuse intention. *Computers in Human Behavior, 119*, 106702.
Jak, S. (2015). *Meta-analytic structural equation modelling*. New York, NY: Springer.
Jang, Y., & Park, E. (2019). An adoption model for virtual reality games: The roles of presence and enjoyment. *Telematics and Informatics, 42*, 101239.
Jeon, S., & Kim, J. (2020). Effects of augmented-reality-based exercise on muscle parameters, physical performance, and exercise self-efficacy for older adults. *International Journal of Environmental Research and Public Health, 17*(9), 3260.
Jiang, T., Yu, D., Wang, Y., Zan, T., Wang, S., & Li, Q. (2020). HoloLens-based vascular localization system: Precision evaluation study with a three-dimensional printed model. *Journal of Medical Internet Research, 22*(4), Article e16852.
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.
Karamians, R., Proffitt, R., Kline, D., & Gauthier, L. V. (2020). Effectiveness of virtual reality-and gaming-based interventions for upper extremity rehabilitation poststroke: A meta-analysis. *Archives of Physical Medicine and Rehabilitation, 101*(5), 885–896.
Kizony, R., Weiss, P. L. T., Shahar, M., & Rand, D. (2006). TheraGame: A home based virtual reality rehabilitation system. *International Journal on Disability and Human Development, 5*(3), 265–270.
Koroleva, E. S., Tolmachev, I. V., Alifirova, V. M., Boiko, A. S., Levchuk, L. A., Loonen, A. J., & Ivanova, S. A. (2020). Serum BDNF's role as a biomarker for motor training in the context of AR-based rehabilitation after ischemic stroke. *Brain Sciences, 10*(9), 623.
Ku, J., Kim, Y. J., Cho, S., Lim, T., Lee, H. S., & Kang, Y. J. (2019). Three-dimensional augmented reality system for balance and mobility rehabilitation in the elderly: A randomized controlled trial. *Cyberpsychology, Behavior, and Social Networking, 22*(2), 132–141.
Landers, R. N., Bauer, K. N., & Callan, R. C. (2017). Gamification of task performance with leaderboards: A goal setting experiment. *Computers in Human Behavior, 71*, 508–515.
Landers, R. N., Bauer, K. N., Callan, R. C., & Armstrong, M. B. (2015). Psychological theory and the gamification of learning. In *Gamification in education and business* (pp. 165–186). Cham: Springer.
Landers, R. N., & Marin, S. (2021). Redesigning job tasks and work itself through workplace gamification: A review, research agenda, and recommendations for practice. *Organizational Gamification*, 63–89.
Landry, M., Ricketts, T., Fraher, E., & Verrier, M. (2009). Physical therapy health human resource ratios: a comparative analysis of the United States and Canada. *Physical Therapy, 89*(2), 149–161.
Lange, B., Koenig, S., McConnell, E., Chang, C. Y., Juang, R., Suma, E., & Rizzo, A. (2012). Interactive game-based rehabilitation using the Microsoft Kinect. In *Virtual reality conference, IEEE* (pp. 171–172). IEEE Computer Society.
LaPiana, N., Duong, A., Lee, A., Alschitz, L., Silva, R. M., Early, J., & Mourad, P. (2020). Acceptability of a mobile phone–based augmented reality game for rehabilitation of patients with upper limb deficits from stroke: Case study. *JMIR Rehabilitation and Assistive Technologies, 7*(2), Article e17822.
Lee, M., Son, J., Kim, J., Pyun, S. B., Eun, S. D., & Yoon, B. (2016). Comparison of individualized virtual reality-and group-based rehabilitation in older adults with chronic stroke in community settings: A pilot randomized controlled trial. *European Journal of Integrative Medicine, 8*(5), 738–746.
Leong, F. T., & Austin, J. T. (2006). *The psychology research handbook: A guide for graduate students and research assistants*. Sage.
Li, C., Song, X., Chen, S., Wang, C., He, J., Zhang, Y., … Shull, P. (2021). Long-term effectiveness and adoption of a cellphone augmented reality system on patients with stroke: Randomized controlled trial. *JMIR Serious Games, 9*(4), Article e30184.
Lord, S., Galna, B., & Rochester, L. (2013). Moving forward on gait measurement: Toward a more refined approach. *Movement Disorders, 28*(11), 1534–1543.
Lozano-Quilis, J. A., Gil-Gómez, H., Gil-Gómez, J. A., Albiol-Pérez, S., Palacios-Navarro, G., Fardoun, H. M., & Mashat, A. S. (2014). Virtual rehabilitation for multiple sclerosis using a kinect-based system: Randomized controlled trial. *JMIR serious games, 2*(2), e12.
Luukkainen-Markkula, R., Tarkka, I. M., Pitkänen, K., Sivenius, J., & Hämäläinen, H. (2009). Rehabilitation of hemispatial neglect: A randomized study using either arm activation or visual scanning training. *Restorative Neurology and Neuroscience, 27*(6), 665–674.
Lv, Z., Penades, V., Blasco, S., Chirivella, J., & Gagliardo, P. (2016). Evaluation of Kinect2 based balance measurement. *Neurocomputing, 208*, 290–298.
Maggio, M. G., Maresca, G., De Luca, R., Stagnitti, M. C., Porcari, B., Ferrera, M. C., … Calabrò, R. S. (2019). The growing use of virtual reality in cognitive rehabilitation: Fact, fake or vision? A scoping review. *Journal of the National Medical Association, 111*(4), 457–463.
Mancuso, V., Stramba-Badiale, C., Cavedoni, S., Pedroli, E., Cipresso, P., & Riva, G. (2020). Virtual reality meets non-invasive brain stimulation: Integrating two methods for cognitive rehabilitation of mild cognitive impairment. *Frontiers in Neurology, 11*.
Ma, M., Proffitt, R., & Skubic, M. (2018). Validation of a Kinect V2 based rehabilitation game. *PLoS One, 13*(8), Article e0202338.
Meurer, W. J., Seewald, N. J., & Kidwell, K. (2017). Sequential multiple assignment randomized trials: An opportunity for improved design of stroke reperfusion trials. *Journal of Stroke and Cerebrovascular Diseases, 26*(4), 717–724.
Milgram, P., & Kishino, F. (1994). A taxonomy of mixed reality visual displays. *IEICE - Transactions on Info and Systems, 77*(12), 1321–1329.
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), 1–9.
Moro, S., Ramos, P., Esmerado, J., & Jalali, S. M. J. (2019). Can we trace back hotel online reviews' characteristics using gamification features? *International Journal of Information Management, 44*, 88–95.
Mubin, O., Alnajjar, F., Jishtu, N., Alsinglawi, B., & Al Mahmud, A. (2019). Exoskeletons with virtual reality, augmented reality, and gamification for stroke patients' rehabilitation: Systematic review. *JMIR Rehabilitation and Assistive Technologies, 6*(2), Article e12010.
Muñoz-Leiva, F., Sánchez-Fernández, J., Montoro-Ríos, F., & Ibáñez-Zapata, J.Á. (2010). Improving the response rate and quality in Web-based surveys through the personalization and frequency of reminder mailings. *Quality and Quantity, 44*(5), 1037–1052.
Muñoz, J. C., Cowling, M. A., & Birt, J. (2016). Using gamification and mixed reality visualization to improve conceptual understanding in ICT system analysis and design. In *33rd international conference on innovation, practice and research in the use of educational technologies in tertiary education* (pp. 455–460).
Myers, N. L., Sciascia, A. D., Westgate, P. M., Kibler, W. B., & Uhl, T. L. (2015). Increasing ball velocity in the overhead athlete: A meta-analysis of randomized controlled trials. The. *The Journal of Strength & Conditioning Research, 29*(10), 2964–2979.
Nguyen, L., Murphy, K., & Andrews, G. (2019). Immediate and long-term efficacy of executive functions cognitive training in older adults: A systematic review and meta-analysis. *Psychological Bulletin, 145*(7), 698–733.
Ohrnberger, J., Fichera, E., & Sutton, M. (2017). The relationship between physical and mental health: A mediation analysis. *Social Science & Medicine, 195*, 42–49.
Paul, S. M., Siegel, K. L., Malley, J., & Jaeger, R. J. (2007). Evaluating interventions to improve gait in cerebral palsy: A meta-analysis of spatiotemporal measures. *Developmental Medicine and Child Neurology, 49*(7), 542–549.
Rasimah, C. M. Y., Ahmad, A., & Zaman, H. B. (2011). Evaluation of user acceptance of mixed reality technology. *Australasian Journal of Educational Technology, 27*(8).
Repetti, R. L., Taylor, S. E., & Seeman, T. E. (2002). Risky families: Family social environments and the mental and physical health of offspring. *Psychological Bulletin, 128*(2), 330.
Riva, G., Mancuso, V., Cavedoni, S., & Stramba-Badiale, C. (2020). Virtual reality in neurorehabilitation: A review of its effects on multiple cognitive domains. *Expert Review of Medical Devices, 17*(10), 1035–1061.
Roberts, H. C., Denison, H. J., Martin, H. J., Patel, H. P., Syddall, H., Cooper, C., & Sayer, A. A. (2011). A review of the measurement of grip strength in clinical and epidemiological studies: Towards a standardised approach. *Age and Ageing, 40*(4), 423–429.
Roots, R. K., & Li, L. C. (2013). Recruitment and retention of occupational therapists and physiotherapists in rural regions: A meta-synthesis. *BMC Health Services Research, 13*(1), 1–13.
Saposnik, G., Levin, M., & & Stroke Outcome Research Canada (SORCan) Working Group. (2011). Virtual reality in stroke rehabilitation: A meta-analysis and implications for clinicians. *Stroke, 42*(5), 1380–1386.
Savela, N., Oksanen, A., Kaakinen, M., Noreikis, M., & Xiao, Y. (2020). Does augmented reality affect sociability, entertainment, and learning? A field experiment. *Applied Sciences, 10*(4), 1392.
Schmidt, F. L. (2017). *Statistical and measurement pitfalls in the use of meta-regression in meta-analysis*. Career Development International.
Shamseer, L., Moher, D., 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: Elaboration and explanation. *BMJ, 349*.
Skidmore, E. R., Holm, M. B., Whyte, E. M., Dew, M. A., Dawson, D., & Becker, J. T. (2011). The feasibility of meta-cognitive strategy training in acute inpatient stroke rehabilitation: Case report. *Neuropsychological Rehabilitation, 21*(2), 208–223.
de Souza, R. F., Farias, D. L., da Rosa, R. C. L. F., & Damasceno, E. F. (2019). Analysis of low-cost virtual and augmented reality technology in case of motor rehabilitation. In *2019 21st symposium on virtual and augmented reality (SVR)* (pp. 161–164). IEEE.
Storm, F. A., Petrarca, M., Beretta, E., Strazzer, S., Piccinini, L., Maghini, C., & Biffi, E. (2020). Minimum clinically important difference of gross motor function and gait endurance in children with motor impairment: A comparison of distribution-based approaches. *BioMed Research International, 2020*, 1–9.
Tekedere, H., & Göke, H. (2016). Examining the effectiveness of augmented reality applications in education: A meta-analysis. *International Journal of Environmental & Science Education, 11*(16), 9469–9481.
Toda, A. M., Klock, A. C., Oliveira, W., Palomino, P. T., Rodrigues, L., Shi, L., & Cristea, A. I. (2019). Analysing gamification elements in educational environments using an existing Gamification taxonomy. *Smart Learning Environments, 6*(1), 1–14.

Tough, H., Siegrist, J., & Fekete, C. (2017). Social relationships, mental health and wellbeing in physical disability: A systematic review. *BMC Public Health, 17*(1), 1–18.

Trojan, J., Diers, M., Fuchs, X., Bach, F., Bekrater-Bodmann, R., Foell, J., … Flor, H. (2014). An augmented reality home-training system based on the mirror training and imagery approach. *Behavior Research Methods, 46*(3), 634–640.

Van Krevelen, D. W. F., & Poelman, R. (2010). A survey of augmented reality technologies, applications and limitations. *International Journal of Virtual Reality, 9*(2), 1–20.

Viechtbauer, W., López-López, J. A., Sánchez-Meca, J., & Marín-Martínez, F. (2015). A comparison of procedures to test for moderators in mixed-effects meta-regression models. *Psychological Methods, 20*(3), 360.

Viglialoro, R. M., Condino, S., Turini, G., Carbone, M., Ferrari, V., & Gesi, M. (2019). Review of the augmented reality systems for shoulder rehabilitation. *Information, 10*(5), 154.

Vukićević, S., Đorđević, M., Glumbić, N., Bogdanović, Z., & Đurić Jovičić, M. (2019). A demonstration project for the utility of kinect-based educational games to benefit motor skills of children with ASD. *Perceptual & Motor Skills, 126*(6), 1117–1144.

Webster, D., & Celik, O. (2014). Systematic review of Kinect applications in elderly care and stroke rehabilitation. *Journal of NeuroEngineering and Rehabilitation, 11*(1), 1–24.

Wiener, J. M., Hanley, R. J., Clark, R., & Van Nostrand, J. F. (1990). Measuring the activities of daily living: Comparisons across national surveys. *Journal of Gerontology, 45*(6), S229–S237.

Wiernik, B. M., & Dahlke, J. A. (2020). Obtaining unbiased results in meta-analysis: The importance of correcting for statistical artifacts. *Advances in Methods and Practices in Psychological Science, 3*(1), 94–123.

Wilson, R. D., Lewis, S. A., & Murray, P. K. (2009). Trends in the rehabilitation therapist workforce in underserved areas: 1980-2000. *The Journal of Rural Health, 25*(1), 26–32.

Winn, C., Chisholm, B., & Hummelbrunner, J. (2014). Factors affecting recruitment and retention of rehabilitation professionals in northern ontario, Canada: A cross-sectional study. *Rural and Remote Health, 14*, 2619.

Wiśniowska-Szurlej, A., Ćwirlej-Sozańska, A., Wołoszyn, N., Sozański, B., & Wilmowska-Pietruszyńska, A. (2019). Association between handgrip strength, mobility, leg strength, flexibility, and postural balance in older adults under long-term care facilities. *BioMed Research International, 2019*, 1–9.

Woodbury, M., Velozo, C. A., Thompson, P. A., Light, K., Uswatte, G., Taub, E., & Wolf, S. L. (2010). Measurement structure of the wolf motor function test: Implications for motor control theory. *Neurorehabilitation and Neural Repair, 24*(9), 791–801.

World Health Organization. (2011). *The world report on disability*. Retrieved from http://www.who.int/disabilities/world_report/2011/en/.

World Health Organization. (2021). Rehabilitation. Retrieved from https://www.who.int/news-room/fact-sheets/detail/rehabilitation.

Xue, H., Sharma, P., & Wild, F. (2019). User satisfaction in augmented reality-based training using microsoft HoloLens. *Computers, 8*(1), 9.

Yim, M. Y. C., Chu, S. C., & Sauer, P. L. (2017). Is augmented reality technology an effective tool for e-commerce? An interactivity and vividness perspective. *Journal of Interactive Marketing, 39*, 89–103.

Ying, W., & Aimin, W. (2017). Augmented reality based upper limb rehabilitation system. In *2017 13th IEEE international Conference on electronic measurement & instruments (ICEMI)* (pp. 426–430). IEEE.

Yip, J., Wong, S. H., Yick, K. L., Chan, K., & Wong, K. H. (2019). Improving quality of teaching and learning in classes by using augmented reality video. *Computers & Education, 128*, 88–101.