



# Measuring the effectiveness of virtual training: A systematic review

Paweł Strojny<sup>a,b,\*</sup>, Natalia Dużmańska-Misiarczyk<sup>a,1</sup>

<sup>a</sup> Institute of Applied Psychology, Faculty of Management and Social Communication, Jagiellonian University, Kraków, Poland

<sup>b</sup> Nano Games Sp. z o.o., Kraków, Poland



## ARTICLE INFO

### Keywords:

Virtual reality  
Virtual environment  
Training  
Learning effectiveness

## ABSTRACT

The amount of research on virtual reality learning tools increases with time. Despite the diverse environments and theoretical foundations, enough data have been accumulated in recent years to provide a systematic review of the methods used. We pose ten questions concerning the methodological aspects of these studies. We performed a search in three databases according to the PRISMA guidelines and evaluated several characteristics, with particular emphasis on researchers' methodological decisions. We found an increase over time in the number of studies on the effectiveness of VR-based learning. We also identified shortcomings related to how the duration and number of training sessions are reported. We believe that these two factors could affect the effectiveness of VR-based training. Furthermore, when using the Kirkpatrick model, a significant imbalance can be observed in favor of outcomes from the 'Reaction' and 'Learning' levels compared to the 'Behavior' and 'Results' levels. The last of these was not used in any of the 330 reviewed studies. These results highlight the importance of research on the effectiveness of VR training. Taking into account the identified methodological shortcomings will allow for more significant research on this topic in the future.

## 1. Introduction

In some domains, virtual learning already plays an important role in education and training (e.g., in medicine; see Zhao, Jiang, & Ding, 2020). Much attention is now focused on developing the technical aspects of software and solving the problems that hinder its wider use. However, from the perspective of learning, another aspect cannot be neglected, namely the development of methods for testing the educational effectiveness of these tools. As we will show below, a rapid increase in the amount of research related to the use of virtual environments in education can be observed in recent years, even though interest in other research areas related to educational technologies has stabilized (Chen, Zou, et al., 2020; Chen, Xie, Zou, & Hwang, 2020). In light of these insights, it can be concluded that this field is in the process of establishing itself and now is the right time to consolidate the data obtained over the last two decades and analyze them in terms of the methods used to evaluate the effectiveness of virtual education. Systematization of the knowledge regarding the methods used so far will allow the strengths and weaknesses of each of them to be identified and will stimulate more thoughtful and theory-driven choices in future research. This, in turn, will contribute to the next step: redirecting the interest of researchers,

decision makers, and users away from the technical aspects and towards issues related to educational effectiveness, understood as causing change in end users. This is what this work is dedicated to. We review the research in this area and propose solutions that could contribute to the increased reliability of future studies and the effectiveness of educational tools and learning processes (see Figs. 8–16).

### 1.1. Virtual, augmented, and mixed reality in training and education

In modern times, life is almost impossible without any form of virtual environment (VE) or virtual reality (VR). These are defined as computer-generated displays that allow the user to perceive, feel, and interact with an environment that is similar to the physical one by using multiple sensory channels, input and output devices, and simulated scenarios (Jayaram et al., 1997; Parsons et al., 2017; Schroeder, 2008). These technologies may also be described in the reality-virtuality continuum (Milgram & Kishino, 1994). In this concept, there are two extremal points on a scale: pure reality and pure virtuality. Everything that falls between these categories is defined as mixed reality (MR). A distinction between augmented reality and augmented virtuality is also used by academics: augmented reality is the real world enhanced by virtual objects, while

\* Corresponding author. Institute of Applied Psychology UJ, ul. Stanisława Łojasiewicza 4, 30-348, Kraków.

E-mail address: [p.strojny@uj.edu.pl](mailto:p.strojny@uj.edu.pl) (P. Strojny).

<sup>1</sup> These authors contributed equally to this work, thus they should be treated as equal first authors.

augmented virtuality is a VE in which some elements of the real world are present. However, when placing environments on the reality-virtuality continuum, emphasis is most often placed on visual stimuli, but other modalities, such as sound, motion, haptics, flavor, and smell, can also be delivered through such environments (Skarbez et al., 2021).

Virtual stimuli can be delivered with the use of various technologies, one of which is head-mounted displays (HMDs), which deliver a stereoscopic image, thus immersing the user in a three-dimensional environment. By blocking peripheral vision, HMDs increase the realism of the environment (Shibata, 2002). Another example of a VE is the Cave Automatic Virtual Environment (CAVE), which consists of several projectors in a room that display various angles of the simulated environment (Cruz-Neira et al., 1992). VE can also be delivered with the use of desktop applications or fixed-base simulators, such as driving simulators.

VEs have many features that make them good tools for learning purposes. They have been widely used in vocational training, especially in professions where hands-on practice is costly or dangerous (e.g., medicine, Seymour et al., 2002; military, Smith, 2010; firefighting, Narciso et al., 2020). VEs allow for knowledge and skills acquisition in conditions similar to those in which this knowledge will be applied in the future, i.e., learning through experience (Psotka, 1995). Furthermore, realistic VEs allow for easier knowledge or skills transfer, which is extremely important from the perspective of effective learning (see, e.g., Carlson et al., 2015). VR technologies allow a sense of 'being there' to be created, namely presence, which is delivered through immersion, the attribute of technology which allows an illusion of reality to be experienced (Slater & Wilbur, 1997). This VE characteristic may evoke engagement and thus be beneficial for learning (Psotka, 1995). In addition, immersion may reduce cognitive load because it allows for the direct perception of a VE (Psotka, 1995). Additionally, thanks to the use of VR, skills and procedures may be practiced in safe environments.

It should be noted that today, especially in a world that has faced the COVID-19 pandemic, various forms of virtual or on-line learning are more popular than ever. However, virtual training (VT), as defined for the purpose of the present paper, should be differentiated from other forms of distance learning. Virtual training does not create a VE *per se*, as it is defined in the previous paragraphs. Computers serve here simply as a medium of communication, and the whole learning process is inherently grounded in the real world. VT environments, on the other hand, create an *entirely new* virtual world in which trainees are immersed and with which they can interact and thus learn. In other words, VT should not be equated to training that uses VR technology as a way of displaying stimuli because the distinguishing feature of VT is not the stimulus presentation technology but the content of the tools that are used within VR. What is actually important is training performed with the use of digitally created environments which provide the opportunity to interact with and within them. Consequently, VT, which is the subject of this work, may belong to various categories distinguished by researchers; for example, VT can be considered as belonging to such thematic categories as context and collaborative learning, blended learning, or online/web-based learning, all of which are of great and growing scientific interest (Chen, Zou, et al., 2020; Chen, Zou, Cheng, & Xie, 2020).

To use a technology effectively, it should be first designed and tested in a strong, research-based theoretical framework (Mayer & Moreno, 1997). To date, there have been several reviews and meta-analyses on the subject of learning in VEs; however, they focus on specific domains (e.g., medical training, Alaker et al., 2016) or modes of training (e.g., serious games, Calderón & Ruiz, 2015). This article aims to be broader and more comprehensive in its scope of analysis, with the intention of providing an overview of the development of this field of research. We believe that enough empirical data has been gathered over the last two decades to take the next step, which means performing data synthesis and drawing conclusions from this data. Such a synthesis should contribute to the development of universal standards in the design of validation programs for individual didactic solutions. Therefore, it is high time to look at VR training from a domain-unlimited perspective with a focus on finding

optimal methods of verifying its effectiveness.

## 1.2. Theoretical framework

There are many theories that can be applied in the design of educational VEs. There are also studies that are not clearly embedded in any framework. In this section, the two most commonly used theoretical frameworks will be presented: Mayer's Cognitive Theory of Multimedia Learning (Mayer, 1997; Mayer & Moreno, 1998) and the Technology Acceptance Model (Davis, 1989).

### 1.2.1. Mayer's cognitive theory of multimedia learning

The cognitive theory of multimedia learning (Mayer, 1997) is based on several concepts and theories of cognitive psychology, such as dual-coding theory (Paivio, 1990; Clark & Paivio, 1991), the working memory model (Baddeley, 1992), cognitive load theory (Chandler & Sweller, 1991; Sweller et al., 1990), generative theory (Wittrock, 1989) and the selection-organization-integration (SOI) model of multimedia learning (Mayer, 1996).

Multimedia learning is defined as the reception of educational material in more than one mode (Mayer, 1997). The modes in which learning occurs are *delivery media*, *presentation modes*, and *sensory modalities*. Almost all computer-mediated learning is multimedia learning, as it contains pictures and words. However, more advanced VT tools, such as immersive VR, can also serve as examples of multimedia learning. In immersive VR, learners are placed in a multimodal environment, where they experience not only sounds and pictures but a whole animated and interactive environment. Therefore, the cognitive theory of multimedia learning, from its basic concept, appears to be applicable to all types of VT.

In teractivity is a core concept in multimedia learning; in Mayer's theory, the learner is perceived as a *knowledge constructor who actively selects and connects pieces of visual and verbal knowledge* (Mayer, 1997, p. 4). It should be noted that this concept is shared by Kolb's experiential learning theory (Kolb, 1984), although Mayer's theory focuses more on cognition than on the experience. Furthermore, the design of an instructional application strongly impacts learning because it affects the learner's level of engagement in processing the material. The learner has to *select* some stimuli or information, then *organize* it into a mental model, and then *integrate* it into a comprehensive representation (Mayer, 1997).

Mayer and Moreno (1998) draw five design principles based on Mayer's cognitive theory of multimedia learning:

- (1) The multiple representation principle: thanks to multiple representations, learners can build multiple mental models, build connections between them, and thus improve learning effectiveness.
- (2) The contiguity principle: because pieces of information have to be present in working memory at the same time in order to build connections between them, different modes of information delivery should be presented simultaneously to enhance learning.
- (3) The split attention principle: using auditory and visual stimuli engages both the visual and the verbal information processing systems, whereas presenting video with subtitles could overload the visual system.
- (4) The individual differences principle: learners with a high level of previous knowledge will already have some mental models, thus creating new models is easier for them.
- (5) The coherence principle: when using multiple modes of learning, it is better to provide a coherent summary of the topic rather than a long text or narration.

Due to their multimodal and interactive nature, all VR-based teaching tools have the potential to implement each of these five principles. The question is to what extent they are implemented in each of them.

### 1.2.2. Technology acceptance model

The technology acceptance model (Davis, 1989) aims to explain why people want to use a particular technological innovation. It is based, among others, on the theory of reasoned action (Fishbein & Ajzen, 1975), which uses preexisting attitudes and behavioral intention to predict actual human behavior. It also draws on Bandura's (1982) self-efficacy theory.

In the technology acceptance model, perceived ease of use and perceived usefulness serve as predictors of behavioral intention, which in turn should lead to actual behavior. Perceived usefulness is defined as the extent to which one believes that using an application will help one perform some tasks better, whereas perceived ease of use is the extent to which one believes that using an application will be effortless (Davis, 1989). The higher the perceived usefulness and perceived ease of use, the greater the chance that a person will use a system in the future.

The technology acceptance model, thanks to its simplicity, has been used in many studies on technological innovations (e.g., meta-analyses by King & He, 2006, and Schepers & Wetzels, 2007). Primarily, this theory is grounded in a vocational context, but it has also been used to explain user behaviors in other contexts, e.g., educational (Hou & Lin, 2017).

### 1.3. Measurement of training effectiveness

Validation of the effectiveness of education using VE is a condition for its widespread adaptation as a didactic method rather than a technological curiosity. The effectiveness of learning tools and programs can be measured in many different ways.

There are many different categories of the possible learning outcomes or the corresponding variables that can be used to measure training or learning effectiveness. The most popular are listed below. It should be noted that this categorization is not exhaustive as these measurements strongly depend on the specific content of the training.

First, knowledge concerning the learned topic can be measured. Declarative knowledge is intuitive and is a very convenient way of operationalizing outcomes since it may be measured and interpreted easily. Knowledge is most often measured using tests, and multiple-answer questions are the easiest way to assess and compare between individuals.

Second, skills can be measured. It is not surprising that such measurement is especially popular when the content of learning is practical (e.g., negotiation; Ding et al., 2020). A specific category of skill measurement is skill transfer. Skills can be taught and measured in various contexts. While the easiest-to-implement solution utilizes the learning context for measuring the learning outcome (in the present paper, this context is most probably some form of a VE), skills can also be measured in the target environment, which is referred to as skill transfer. Such measurement is particularly beneficial because it makes it possible to determine if the acquired skills can be *transferred* from an artificial environment to a real one that is similar or identical to the one in which the skills will be actually used in practice.

The measurement of attitudes can be applied to assess the effectiveness of a training method or materials. Attitudes are somewhat related to the subjective assessment of training, which is very frequently used. This measurement consists of various questions that aim to capture the characteristics of the subjective experience and can be applied in many different ways: closed-ended questions, open-ended questions, one-to-one interviews (Dalinger et al., 2020), focus groups (Adams et al., 2019), and more. The subjective experience is probably the most informal learning outcome and should not be used as the only indicator of the effectiveness of the training environment tested.

A common construct that is used in studies of training effectiveness is motivation, which is extremely useful as it indicates the willingness to participate in learning activities; without motivation, effective delivery of learning material is extremely difficult. Regular questionnaires (e.g., Sattar et al., 2019) are one way of measuring motivation, but due to the

immersive and continuous nature of the VR training experience, researchers are also looking for other methods of measuring engagement, such as noninvasive measurement of parameters that indicate the effort involved in tasks (e.g., Czarnek et al., 2021).

Of course, there are many other constructs that can be used as indicators of training effectiveness. These include emotions (Harley et al., 2020), locus of control (Nykänen et al., 2020), attention (Hart & Proctor, 2020), flow (Hou & Lin, 2018), or observation of in-training behavior (den Haan et al., 2020). In addition, physiological indices can be used to indirectly measure the effectiveness of a particular learning tool. Using such indices is not very common today, but some efforts have already been made in the past decade with the use of cardiovascular measures, electrodermal activity, eye tracking, functional near-infrared spectroscopy, and EEG (e.g., Legrand et al., 2011).

#### 1.3.1. Kirkpatrick's model of training evaluation

The above-mentioned methods can, to some extent, be mapped onto Kirkpatrick's model of training evaluation, which has been used over many decades and remains effective (Kirkpatrick, 1976, 1994). This framework is fairly universal and can be applied to many forms of training, but it was primarily designed to work in a vocational context. Kirkpatrick begins by defining what evaluation actually means and what its goals are. The effectiveness of training programs is determined by evaluating them in order to improve existing programs and to identify and further exclude ineffective ones from practice. In this model, four levels of training effectiveness assessment are defined:

- (1) Reaction. Operationalized as an affective response to training (engagement, satisfaction) and its relevance to everyday practice (Bates, 2004).
- (2) Learning. The achievement of learning outcomes is assessed. It is good practice to include a pre- and post-test of the learning outcomes.
- (3) Behavior. Here, the application of new knowledge and skills in daily practice is assessed. This is most easily done in a vocational setting, but it can be applied to some extent in other domains.
- (4) Results. This step of training effectiveness evaluation goes beyond the individual benefits. Tangible results are assessed from the perspective of an entire organization, such as quality improvement.

#### 1.3.2. The temporal aspect of training effectiveness measurement

When studying the effectiveness of any form of learning, there are a couple of specific points in time at which the learning outcome of interest can be measured. In a *pre-test*, the initial level of knowledge, skill, or other variable is measured. This is a very valuable measurement as it allows the baseline levels of learning outcomes to be established, it reduces the risk of a ceiling effect, and it allows the calculation of learning gain indices. An *in-training measurement* can also be applied in which the learning outcome is measured without interrupting the learning. Such non-intrusive measurement allows for insight into the learning process while it is actually happening. It can be performed, for example, by calculating scores in a learning application or by analyzing in-training behaviors. Another method of performing in-training measurements is the use of physiological indices to indirectly measure the learning progress. Such non-intrusive measurements are also beneficial as they do not interrupt the learning process. In a *mid-test*, learning outcome measurement is applied between separate sessions of the whole learning activity. This approach is particularly useful when training is divided into several stages. *Post-tests* occur just after the learning, while the *retention test* is done after some time has passed; this can be a couple of days or as much as a few months. By using a combination of the aforementioned measurements, secondary indices can be calculated, e.g., learning gain, or the difference between the post-test and pre-test scores. With multiple measurements, one can also observe the learning curve in more detail and implement Mayer and Moreno's (1998) principle of individual differences.

#### 1.4. The objectives of the study

The objectives of this review are threefold. First, its purpose is to summarize the current state of the art in the effectiveness of VT assessment. To achieve this goal, a broad literature review was conducted. Based on the analysis of the reviewed articles, this review also aims to point out some methodological shortcomings and propose further research directions in the field of VT effectiveness. Lastly, a framework will be proposed in the hope that it will be useful for researchers conducting studies in this area.

Specifically, the following research questions were posed:

- (1) How does research on VT effectiveness develop over time, taking into account the technologies used?
- (2) In which fields of education and human activity are virtual methods used for educational purposes?
- (3) What are the sample sizes in studies on the effectiveness of VT?
- (4) What study designs are used in terms of the number of test sessions and temporal organization?
- (5) What study designs are used in terms of experimental groups and comparisons?
- (6) What methods are used to evaluate the effectiveness of VT?
- (7) What conclusions can be drawn from the reviewed articles in terms of the effectiveness of VT?
- (8) What methodological shortcomings have been identified in research on the effectiveness of VT?

When searching for answers to Questions 1 and 2, we aimed to identify the state of interest of VT researchers in both temporal and domain terms. As previously mentioned, VT should not be equated with training in VR, so we also investigated the technologies used. In the case of Questions 3–5, our motivation was to look at the key parameters of research programs from the perspective of scientific reliability: sample size, research design, and temporal organization of the research. These parameters may affect the possibility of drawing conclusions. To generalize conclusions regarding the effectiveness of VT, operationalizations of dependent variables are of key importance. When looking for answers to Question 6, we tried to determine which methods of data collection dominate, which kinds of learning outcomes attract the greatest attention of researchers, and at which level, in Kirkpatrick's view, these outcomes are measured. Furthermore, we decided to check the basic parameters of VT effectiveness that were evaluated in the studies we analyzed (question 7) and to identify possible imperfections, if any (question 8).

## 2. Materials and methods

### 2.1. Search strategy

The search and selection process was conducted using guidelines from the PRISMA statement (Page et al., 2021). A literature search was performed in three electronic databases (Science Direct, Scopus, Taylor & Francis) with no publication date restriction. The following keywords and logic were used: (“virtual training” OR “virtual learning”) AND (effectiveness OR “learning outcomes”). The keywords and logic were based on the assumption that the topic of interest to us is research on VT, sometimes called ‘virtual learning’, but with the sole aim of assessing the effectiveness of VT, sometimes also expressed by the term ‘learning outcomes’. In other words, our intention was to exclude articles that did not report any attempts to verify the effectiveness of VT. The search was restricted to research articles (original empirical research). For the Science Direct and Scopus databases, the search was conducted on 6 November 2020; for the Taylor and Francis database, the search was conducted on 7 December 2020. From the Science Direct and Scopus databases, 1291 and 1253 records were retrieved, respectively. From the Taylor & Francis database, the 2000 most-relevant records were obtained (the search was limited to this number of records due to technical

constraints). In total, 4544 records were retrieved.

### 2.2. Study selection

We selected articles for analysis in a few steps, as detailed in Fig. 1. We assessed titles, abstracts, and possible duplicate studies. Subsequently, articles on VT tools for medical procedures were excluded as there are already many systematic reviews on this specific topic, and the number of records for evaluation was extremely large. The articles retrieved in full were evaluated using the following criteria:

Inclusion criteria:

- (1) The research focused on evaluating the effectiveness of training performed in an interactive VE;
- (2) Full text available;
- (3) Peer-reviewed, scholarly articles;
- (4) Empirical research conducted on a healthy population.

Exclusion criteria:

- (1) Studies focused on testing specific hardware;
- (2) Studies focused on testing entire curricula rather than a particular training method;
- (3) Language of publication other than English or Polish.

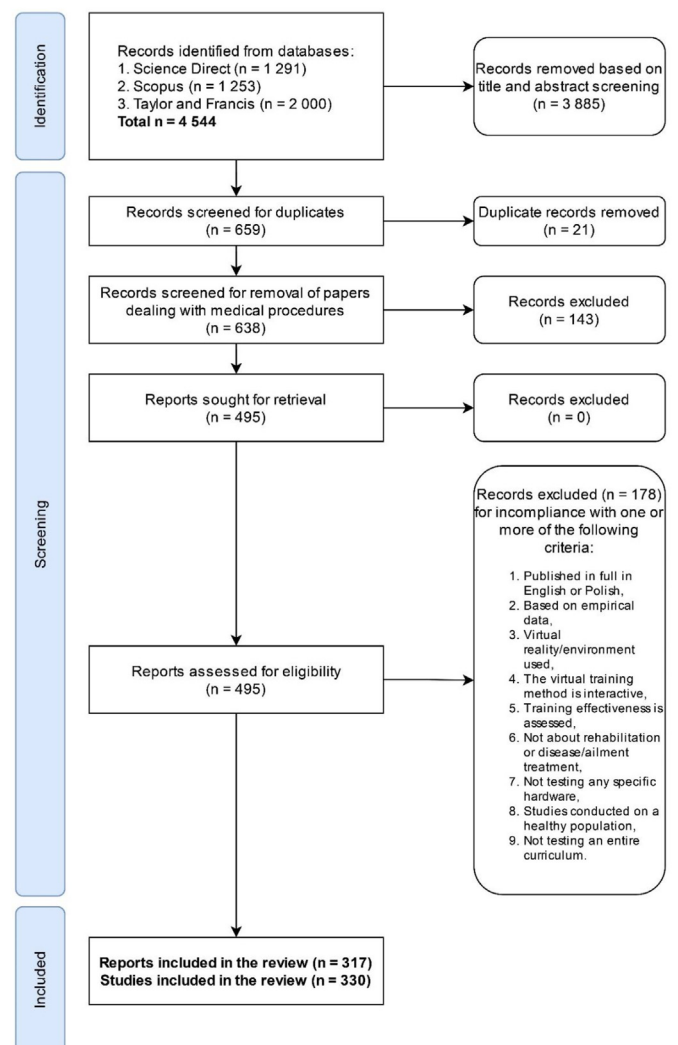


Fig. 1. Flow chart of the search and selection process for the relevant literature.



After this process, 317 articles were chosen for further review, 178 of which were rejected due to incompatibility with the criteria above. The final database consisted of a total of 317 articles (330 studies).

### 3. Results and discussion

#### 3.1. Question 1: how does research on VT effectiveness develop over time, taking into account the technologies used?

Various types of technology were used in the analyzed studies. Obviously, these differences are to some extent dependent on time as new technologies become available every year, but some regularities can still be observed. The technology types were categorized according to the reality-virtuality continuum (Milgram & Kishino, 1994). Desktop and mobile device applications, as they do not fall directly into any of the categories of the continuum, were treated separately. Tools that use real-world input to enhance the virtual experience (e.g., haptic devices) were mostly placed in the Augmented Virtuality category. When different types of technology were used in one study (e.g., for different experimental conditions), the leading technology, which was usually the most advanced, was chosen for purposes of this categorization.

As can be seen in Fig. 2, most of the studies were conducted with some kind of desktop software (166 studies, 50%). This is reasonable as desktop software is easily accessible and is not extremely expensive to develop today. Such tools and applications have a good chance of not only being used for scientific purposes but also being applied in real life. VR comes second in terms of frequency of use, with 110 studies (33%). Augmented reality (26 studies, 8%) and Augmented virtuality (23 studies, 7%) were used less frequently, and the rarest technology used in the reviewed studies was mobile device applications (5 studies, 2%).

The reviewed studies were carried out over almost three decades. The oldest study was published in 1994 and the newest ones were from 2021. As can be seen in Fig. 3, there is a strong increasing trend in the number of published studies exploring the effectiveness of VT. From 1994 to 2007, there were only a couple of studies published each year; however, from 2008, the number of published studies rose gradually, with a high spike from 2019 to 2020. This trend can be predicted to continue in the coming years. When the studies on the timeline are classified according to the technology used in them (using the reality-virtuality continuum; Milgram & Kishino, 1994), it can be seen that the number of studies using VR technology has increased over time as this technology has become more popular, advanced, and accessible. Especially in the decade from 2008 to 2018, desktop applications were the most popular, but in 2019 and 2020 the number of studies conducted with the use of VR exceeded the number of studies conducted with desktop applications. In the papers from 2021, there are twice as many studies using VR than desktop applications. It is worth noting that while VT can be effectively

implemented with 2D displays, some features of the more advanced displays that are used for VR and augmented reality/virtuality allow for higher levels of presence (Shu et al., 2019). Taking into account that presence turns out to be a predictor of learning outcomes (see, e.g., Dengel & Mägdefrau, 2020, June), it can be expected that, along with the development and increasing availability of these technologies, their share in both research and practical applications should increase. In the process of popularizing these technologies, not only objective factors (cost-benefit ratio) can play a key role, but also reliable information about the possible advantages of newer solutions; this should help overcome the habits of tool providers and educators. This is another reason why identifying and addressing the shortcomings of VT effectiveness research programs is important.

#### 3.2. Question 2: in which fields of education and human activity are virtual methods used for educational purposes?

The fields in which virtual technologies are used for educational purposes are wide and range from training simple motor skills (Peterson et al., 2018), through numerous tools for schooling (Zhang et al., 2020), to vocational training (see, e.g., Herrington & Tacy, 2020). The reviewed studies were classified according to the domain in which each one was conducted. As can be seen in Fig. 4, 35% of the studies were conducted in a university setting (117 studies). Training in practical and work-related skills accounted for the next 40% (130 studies). The remaining categories (83 studies) accounted for approximately 25% of the reviewed studies.

Interestingly, VT in areas such as sport and physical rehabilitation seems to have much greater potential than the number of published studies indicates. The dominance of the university context should not be surprising due to the early stage of development of these technologies. The probable reason for this is the ease of carrying out the tests and the availability of tools. However, it should be expected that, over time, the ratio of research in this context to other contexts (more directly related to possible areas of implementation) will change in favor of the latter. It cannot be expected that end users will be convinced to pay the costs of implementing such methods unless they have reliable data that confirm the legitimacy of the use of these methods in the specific domains of education in which they work.

#### 3.3. Question 3: what are the sample sizes and training times in studies on the effectiveness of VT?

On average, there were 91 participants in a single study; however, this number varies greatly ( $SD = 203.98$ ). The highest number of participants in the reviewed studies was 2,727, and the lowest was 4. This large variability is related to the variability in study designs and assessment methods. For review purposes, the number of participants in each reviewed study was classified into one of six categories: 1–10, 11–30, 31–50, 51–100, 101–200, and more than 200. As can be seen in Fig. 5, the most common number of participants in the reviewed studies was between 51 and 100 (108 studies, 33%), which is a considerably large sample for experimental studies. Smaller sample sizes, between 11 and 30 (71 studies, 21.5%) and between 31 and 50 (66 studies, 20%), were also quite common. More than 200 participants were reported by 18 studies (5%), most of which were large-scale evaluations of training effectiveness using a non-experimental design and self-assessment methods.

Despite the fact that researchers have emphasized for a very long time (see, e.g., Scheffé, 1959) that power analyses are necessary for rational statistical decisions, the implementation of these postulates used to be difficult and inconvenient, and therefore rarely practiced. The situation changed with the advent of easy-to-use tools such as G\*Power (Faul et al., 2007). As noted above, the great variability of sample sizes may be due to the variation in research goals and designs. However, if this is the only reason, it would be expected that the authors of particular research programs would choose their sample size according to their goals and

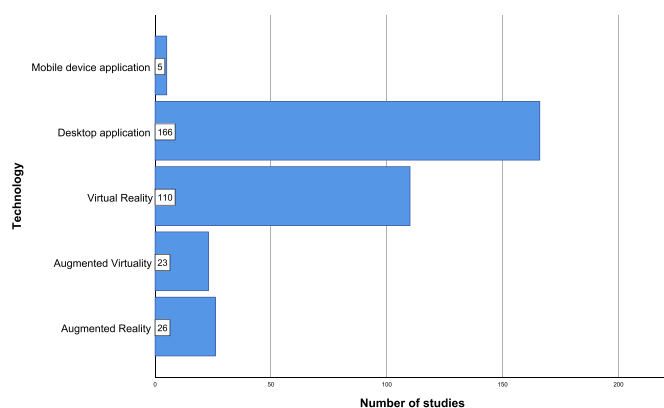


Fig. 2. The number of studies utilizing the types of technology categorized by the Reality-Virtuality continuum criteria (Milgram & Kishino, 1994), with additional categories for desktop and mobile tools.

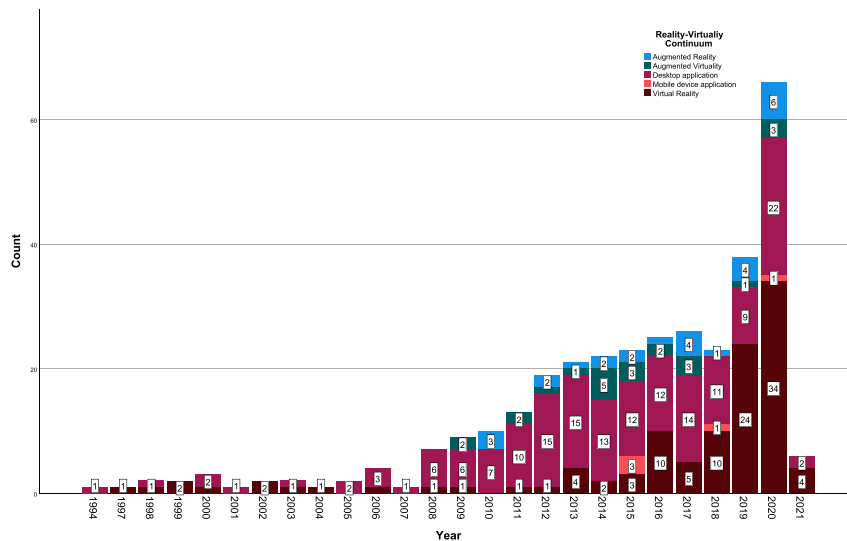


Fig. 3. The number of published studies on the effectiveness of virtual training, broken down by the technologies under investigation.

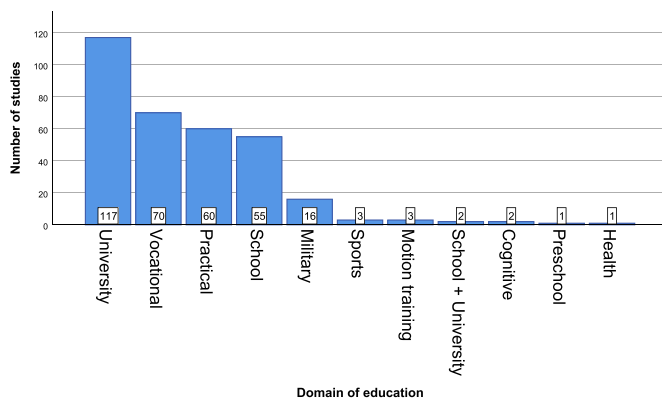


Fig. 4. The number of published studies on the effectiveness of virtual training, by education domain.

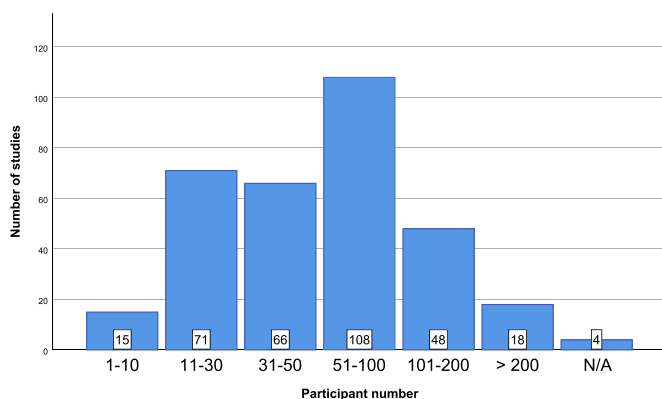


Fig. 5. The number of studies on the effectiveness of virtual training, depending on the number of participants.

design. To briefly check whether there are reasons for considering this speculation, we randomly selected 20 papers from the two previous years (10 from 2019 to 10 from 2020; 22 experiments in total) and scanned the method descriptions in search of substantive justifications for the decision regarding the sample size. The authors of only one of the scanned articles reported that they had used the G\*Power application to determine the sample size. Due to the assumption that awareness of the

existence and importance of using power analysis tools may gradually grow, we deliberately decided to draw papers from the last few years. Nevertheless, the obtained proportion seems unsatisfactory as it may result in erroneous research conclusions due to under- and overpowered designs. These studies may be misleading as they suggest the existence of effects that may not actually matter in pedagogical practice.

#### 3.4. Question 4: what study designs are used in terms of the number of test sessions and temporal organization?

The time spent learning itself may be a relevant factor when comparing studies on the effectiveness of VT. It varies between individual reports and to some extent depends on the field in which the study was conducted. Learning motor skills is undoubtedly easier and less time consuming than learning complex material, such as scientific concepts or vocational skills. For review purposes, training times were divided into eight groups (0–15 min, 16–30 min, 31–59 min, 1–2 h, >2 h, several sessions, self-paced, and N/A – information not available). A comparison of the number of studies that fall into each of these categories can be seen in Fig. 6. In some studies (9 studies, 3%), the learning was self-paced, so the learners could adjust the time spent on the learning activity according to their own needs. Quite often, the training phase spanned several sessions on separate days (67 papers, 20%). The longest training time in a single session was 10 h (Hatz, 1999), while some articles reported a training or learning phase lasting approximately 3 h (Annetta et al., 2014; Beaumont et al., 2011; Carenys et al., 2017; Chen, 2014; Chen & Huang,

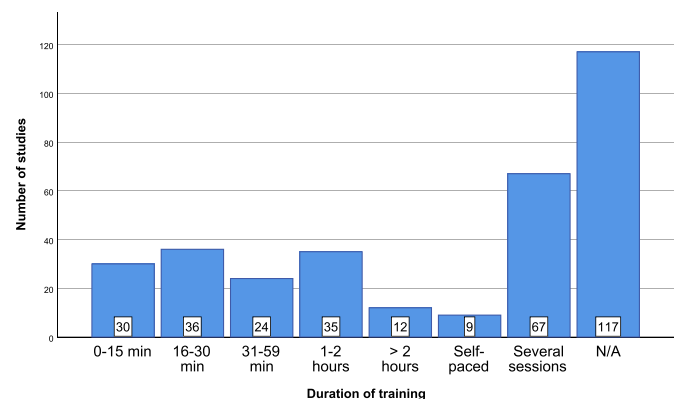


Fig. 6. The number of published studies on the effectiveness of virtual training, by the duration of the training.

2012; Hays & Vincenzi, 2000; Wener et al., 2015). The shortest training time reported was approximately 2 min (Burigat & Chittaro, 2016). It should also be noted that training time was not explicitly reported in many of the studies (117, 35%). However, this information should be provided so experimental procedures can be fully understood.

Because longitudinal designs are more complex but particularly informative, we decided to look at them in detail. In the next step, the 67 studies that were based on multiple VT sessions were analyzed. The following information was collected for each individual study which covered multiple sessions: the length of a single session, the number of sessions, the delay between sessions, and the length of the entire study program.

In terms of the length of a single session, there was no clear leading category. However, it is possible to distinguish two categories regarding the length of a single session in the reviewed studies: shorter (maximum 60 min, 21 studies) and longer (over 60 min, 20 studies). However, the most frequent observation was that data regarding the length of a single session were missing (see Fig. 7).

Regarding the number of sessions, the most frequent choice was to conduct between 2 and 6 sessions, and the highest number was 48 (Okutsu et al., 2013; see Fig. 8).

Most of the missing data concerned the delay between sessions (47 of 67 studies, 70%). However, in the subset of studies where these data were available, the most frequently applied delay duration was a week (see Fig. 9).

The duration of the entire study program in the reviewed studies ranged from one day to one year, with two weeks being the most frequent duration. The smallest amount of missing data was observed for this category (see Fig. 10).

Obviously, the high level of variation between the multisession studies in terms of their characteristics is understandable. The study design must correspond to the individual research questions posed by the authors; therefore, the presented results should not be a cause for criticism. However, what may be disappointing is the relatively large share of missing data, which prevents a more detailed analysis of individual programs and consequently makes it difficult to draw generalized conclusions about the dynamics of learning.

Regarding the length of the delay between sessions, there were 47 cases of missing data; therefore, this part of the analysis is the least reliable. It should be noted here that a large amount of missing information was observed in the reviewed articles in other cases, not only related to the delay length. Such a lack of attention to detail in describing study procedures makes it difficult to properly assess the methodology and results of studies, and it makes replication virtually impossible. Moreover, from the pedagogical perspective, these deficiencies may prevent (and certainly discourage) attempts to implement the studied pedagogical methods in practice. It is easy to imagine that an educator who is planning to attempt to implement a VT method that has been well

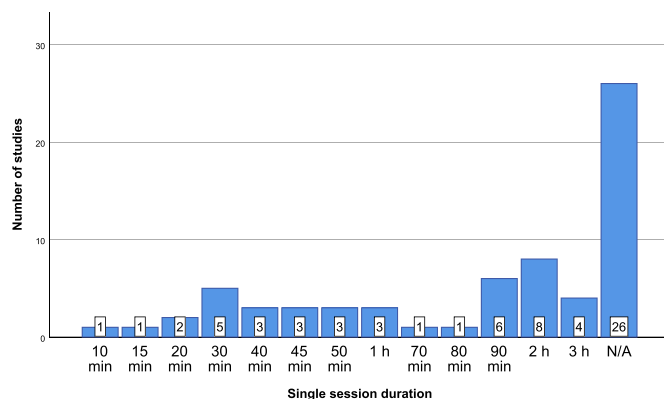


Fig. 7. The number of multisession studies on virtual training effectiveness, by the duration of the single session.

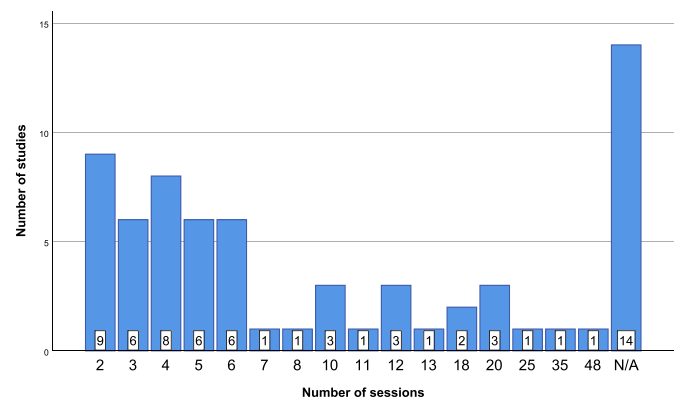


Fig. 8. The number of multisession studies on virtual training effectiveness, by the number of sessions that make up the entire program.

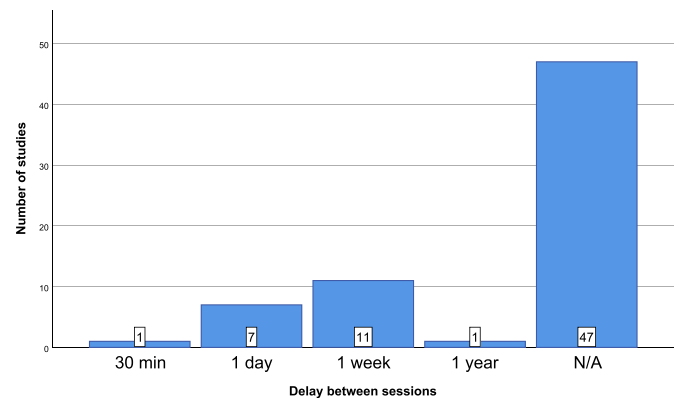


Fig. 9. The number of multisession studies on virtual training effectiveness, by the delay between sessions.

tested in empirical research may give up when they realize that the description of the study lacks basic information (e.g., the chosen approach was used in a certain number of sessions, but their length or time intervals are not known).

### 3.5. Question 5: what study designs are used in terms of experimental groups and comparisons?

The reviewed studies were analyzed in terms of study design. In terms of the types of comparisons, they were classified by means of between-subjects design, within-subjects design, or mixed-design studies, and in which stage of the training program the evaluation was conducted (pre-test, in-training measurement, mid-test, post-test, retention test).

It is not surprising that mixed-design studies were most commonly used (141 studies, 43%), in which comparisons are made between subjects and within subjects. In this way, one can test not only if a given VT tool is more effective than other modes of learning, but also how the learning effects change over time. A very common approach was to use a pre-test and a post-test (e.g., Dalim et al., 2020), or a pre-test, post-test and a retention test (e.g., Ding et al., 2020) of the learning outcome. This allows the initial knowledge or skill level to be taken into account in the analyses. A pure between-subjects design with one measurement point was also quite common (76 studies, 23%), followed by within-subjects designs (67 studies, 20%). Surprisingly, a large number of studies did not use any comparison at all (46 studies, 14%) as they focused on purely exploratory quantitative or qualitative analyses of learning outcomes (see Fig. 11).

The analyzed papers also differed in terms of the measurement point. In almost all studies (306, 93%) a post-test was applied. Pre-test

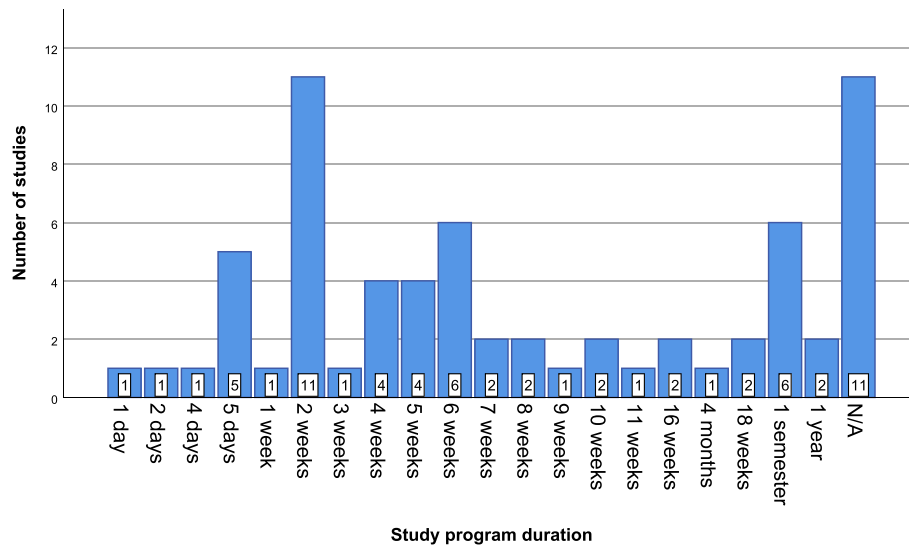


Fig. 10. The number of multisession studies on virtual training effectiveness, by the duration of the entire program.

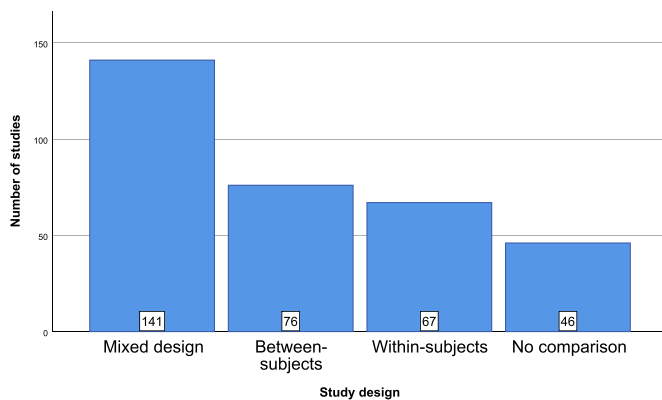


Fig. 11. The number of studies, depending on the study design.

measurements were also very common (187 studies, 57%). Other measurement points were not as common, with training measurement being used in 61 studies (18%), retention test in 25 studies (7.5%), and mid-test (a test conducted after a segment of training) in only 8 studies (2%) (see Fig. 12).

VT can be compared with many different modes of learning to assess its effectiveness. Among the reviewed studies, traditional learning, other

types of VT and no learning were the types of comparison groups observed. Traditional learning can take various forms, such as reading, watching a presentation, on-site training, or watching a video. VT effectiveness can also be evaluated in comparison to other forms of VT; here, different levels of fidelity, presence or absence of feedback, technology, types of stimuli, and difficulty level can serve as examples (see Fig. 13).

Differences between the nature of control groups will lead to differences in effectiveness assessment. Comparing a new learning tool to no learning at all is the easiest option and will probably give significant results in favor of the tested tool. However, comparing an innovation with a traditional form of learning is more valuable because it allows one to actually determine if the tested innovation is worth implementing in everyday practice. In addition, experimental comparisons between different types of VT may also provide valuable information. It should also be noted that, from a methodological perspective, a control group should be as close to the experimental group as possible, ideally differing in just one aspect so that other factors do not confound the results. Although this is very difficult to achieve in the evaluation of learning effectiveness since learning and learning interventions can be complex and difficult to disassemble into primary factors, efforts should be made in this direction whenever possible to minimize the risk of drawing conclusions from observed effects that are attributable to uncontrolled secondary variables.

### 3.6. Question 6: what methods are used to evaluate the effectiveness of VT?

Training effectiveness can be measured in many different ways. In many of the reviewed studies, more than one method was used, therefore the numbers presented below add up to more than 330. In the reviewed studies, the most common approach involved the use of some kind of objective method (e.g., knowledge test); such methods were applied in 271 (82%) of the reviewed studies. Subjective evaluation of the training tool, learning outcomes, motivation, or other psychological constructs was also used in 239 (72%) of the studies. Observation by an expert (21 studies, 6%) as a form of assessment of learning effectiveness was not used very frequently. It was used primarily where objective assessment methods could not be applied (for example, analyzing behavior; Bart et al., 2008), or for qualitative assessment of learners' reactions during learning (for example, Alves Fernandes et al., 2016). Physiological measurements (10 studies, 3%) were the least frequently applied. Such methods allow the indirect assessment of learning using physiological markers (see Fig. 14).

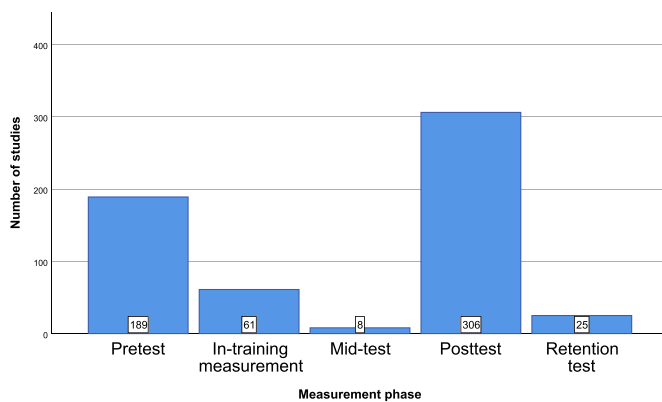


Fig. 12. The number of studies, depending on the measurement time. Most of the studies performed more than one measurement, so the results do not add up to 330.



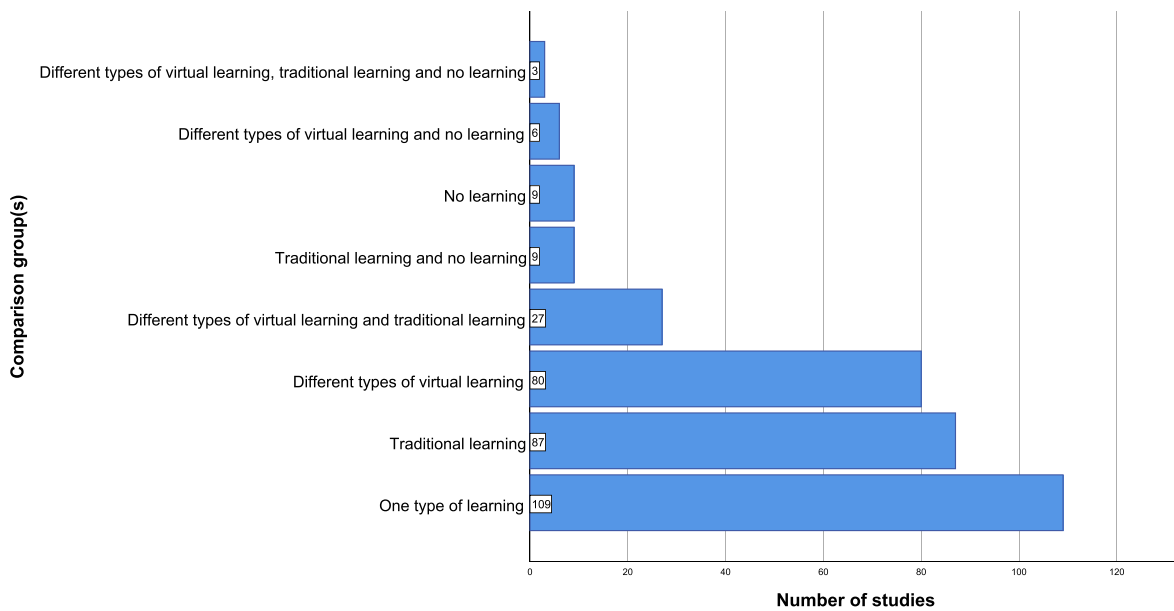


Fig. 13. Comparison groups used in the published studies on the effectiveness of virtual training.

As noted above, many studies used more than one method; in particular, using a combination of methods from two different categories was the most popular choice (181 studies, 55%). A large number of studies used only one method to evaluate training effectiveness (134 studies, 41%), while only 15 studies used a combination of three different methods (5%). Approaching the evaluation of training effectiveness from at least two perspectives appears to be a good choice as more unique characteristics of the studied learning method can be captured. Omitting subjective experience evaluation could lead to the creation of tools that are highly effective in teaching certain skills or knowledge, but are at the same time extremely frustrating or stressful for end users.

As learning outcomes are, in most cases, the intended direct effect of the use of pedagogical methods, it is not surprising that learning outcomes played a significant role as the dependent variable (irrespective of the measurement method, be it subjective, objective, observational, or physiological). Learning outcomes can be measured using different methods and indices. Sometimes simple objective methods (such as knowledge or skill tests) are used; other constructs are sometimes used to draw conclusions regarding the learning outcomes achieved. The reviewed articles were analyzed in terms of the learning outcomes measured. It is not surprising that the most commonly used learning outcome metrics were subjective evaluation of training (measured by

perceived usefulness, perceived ease of use, satisfaction level, etc., by questionnaires, interviews or focus groups; 169 studies, 51%) knowledge (149 studies, 45%) and skills (144 studies, 44%) as they are straightforward and easy to understand and interpret. Other metrics used in the articles include motivation (38 studies, 11.5%), skill transfer (to other tasks or to real-life tasks; 38 studies, 11.5%), self-efficacy (22 studies, 7%), attitudes (20 studies, 6%), and participation (13 studies, 4%). Other metrics were constructs such as emotions, anxiety, confidence, or enjoyment. All of these can be described as indirect indices of training effectiveness (see Fig. 15).

Learning outcomes can be classified according to the levels of training evaluation outlined by Kirkpatrick: Reaction, Learning, Behavior, and Results (Kirkpatrick, 1976; 1994). The most commonly measured level of training was Learning (271 studies, 82%). The Reaction level was also studied very frequently (231 studies, 70%). Forty-four studies (13%) evaluated the Behavior level. Skill transfer tests were classified as being related to the Behavior level as they require the implementation of newly learned knowledge or skills in other settings. None of the reviewed studies evaluated the effects of the Results level. The Results and Behavior levels are undoubtedly more difficult to investigate than the Reaction and Learning levels because they require longer evaluation; therefore, it is not surprising that the number of studies that evaluate the Behavior and Results levels is considerably lower than those that evaluate the Reaction and Learning levels of learning effectiveness. Considering that successive levels in Kirkpatrick's classification are more generalized but are also increasingly closer to reality, which is complex, the complete abandonment of research on educational impact at the highest level is worrying. Certainly, such extensive research is very demanding and carries a higher risk of failure, but it is precisely this research that has the potential to provide answers to guide educational policies. To facilitate research at the two most complex levels (i.e., Behavior and Results), collaboration between academia, industry, and NGOs may be helpful. Consortia composed in this way would have the best chance of implementing research projects that observe the impact of new educational methods at a high level of generality, while maintaining high scientific standards (see Fig. 16).

### 3.7. Question 7: what conclusions can be drawn from the reviewed articles in terms of the effectiveness of VT?

In addition to analyzing the papers in terms of methodology, their

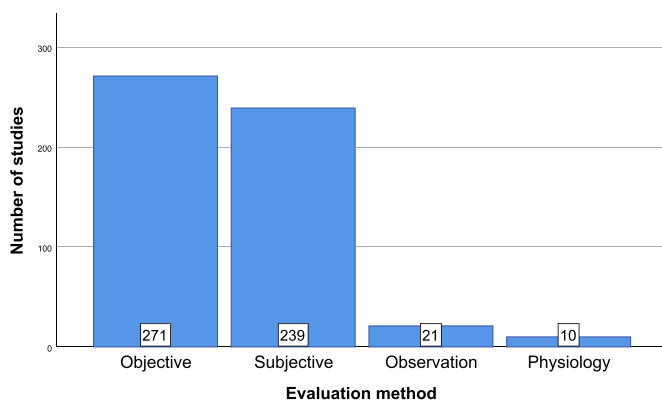
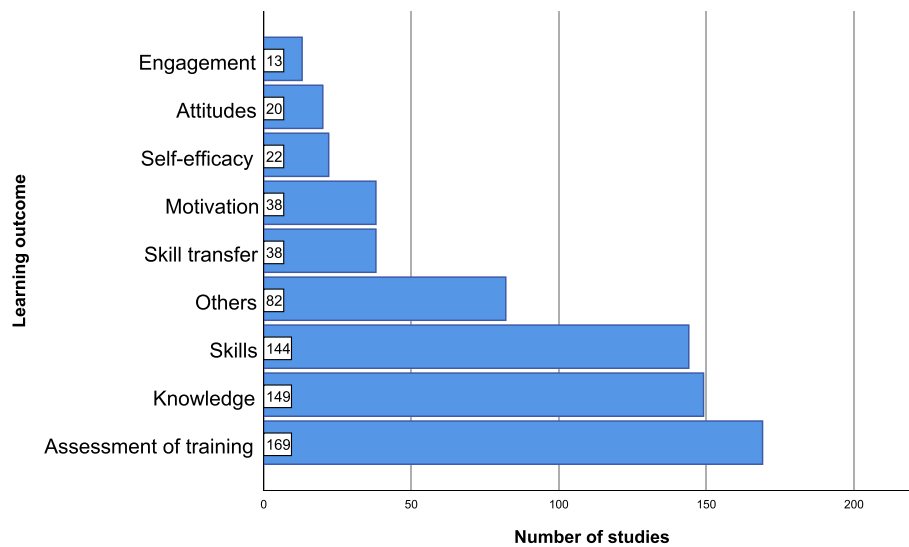
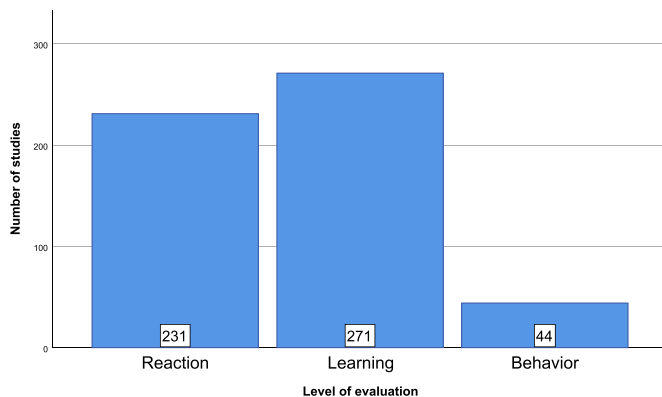


Fig. 14. The number of studies utilizing various learning effectiveness evaluation methods. Some of the studies used methods from more than one category, so the results do not add up to 330.



**Fig. 15.** The number of studies utilizing various learning outcome evaluation methods. Some of the studies used methods from more than one category, so the results do not add up to 330.



**Fig. 16.** The number of studies that tested learning outcomes, categorized according to Kirkpatrick's criteria (1976, 1994). Not a single study used methods belonging to the highest category – Results. Some of the studies used methods from more than one category, so the results do not add up to 330.

results were also reviewed. This was not the main objective of this work, but it was done to answer the question ‘Is VT effective?’ at a very basic level. The main findings of this study were classified into several possible effectiveness assessment outcomes. Note that because one study may have used more than one method, some of them might have been classified in more than one category; on the other hand, some studies did not report suitable comparisons. Hence, 251 comparisons have been taken into account. It should also be noted that the main conclusions of each individual study were as simplified as possible to allow quantitative analysis.

Most of the comparisons (213) reported a positive result using different comparisons: a significant gain in the learning outcome (117 studies, 47%), and the superiority of VT over both traditional learning (77 studies, 31%) and no learning at all (19 studies, 8%). Although the VT method was not the most effective of the methods studied, it turned out to be just as effective as traditional teaching methods in 26 cases (10%). Only a small number (12) of studies reported clearly negative effects: lower effectiveness of VT than traditional learning (9 studies, 3%) or no effectiveness at all (3 studies, 1%). Thirty-nine studies did not report directly on the measurement of learning effectiveness. We separately compared study participants’ subjective evaluations: 114 studies (93%) reported that the VT methods were positively assessed by the

participants, and in eight cases (7%) the evaluation was negative.

A study by [Jong \(2015\)](#), where the effectiveness of VT was moderated by the level of pre-learning achievement, is particularly interesting: for high-achieving students, VT was less effective than traditional learning; for moderately achieving students, the two modes of learning were equally effective; for low-achieving students, VT was more effective than traditional learning. This interesting example highlights the importance of controlling for initial levels of the learning outcome tested. Had this not been controlled, the conclusion would probably have been that VT is as effective as traditional learning. However, taking the initial levels of achievement into account allowed for an in-depth mode analysis and led to the conclusion that the effectiveness of VT differs between learners. The results obtained certainly suggest that VT is effective; however, given that the above cursory analysis was only the secondary goal of the current article, more work that focuses solely on the educational effectiveness of VT is definitely needed. Only such a systematic review would allow us to sufficiently identify and explore the complex dependencies in this field that were suggested by the above work by [Jong \(2015\)](#).

### 3.8. Question 8: what methodological shortcomings have been identified in the research on the effectiveness of VT?

In order to identify the potentially most important methodological shortcomings, we decided to conduct a qualitative analysis of the collected research. There are many issues that become visible when analyzing studies on the effectiveness of VT. First, the problem of the delay between the training and testing phases arises. In many studies (e.g., [Gonzalez-Franco et al., 2017](#)), the skill or knowledge test is conducted immediately after training or learning. This solution is obviously the easiest from the perspective of logistics, as one does not have to invite participants for several lab sessions, thus reducing participant dropout. However, by applying some form of delayed post-test or retention test that is conducted days, weeks (see, e.g., [Lovreglio et al., 2021](#)) or even months (see, e.g., [Diego-Mas et al., 2020](#)) later allows for a better understanding of knowledge and skills retention. Even when participants do dropout, some data can be collected and used in further analyses. However, during the delay phase, participants may (or may not) come into contact with the learned material, thus distorting the direct influence of VT. Despite these difficulties, including a retention test in research programs appears to be valuable.

Another issue is the trade-off between sample size and evaluation accuracy. In some of the reviewed studies (e.g., [Braghirolli et al., 2016](#)),

the sample sizes were considerably large (over 2000 in some cases; Rein et al., 2018), but such large studies make it impossible to conduct more in-depth analyses with the use of more-complex methods. On the other hand, the studies which had small sample sizes (the smallest number of participants was 4 in the study by Herrington & Tacy, 2020) allowed those researchers to perform in-depth analysis of learning effectiveness, but this comes at the expense of limited quantitative inference possibilities. Triangulation of methods, individual interviews, and even psychophysiological measurements are possible with smaller sample sizes. This trade-off is difficult to solve, and the best solution depends on the specific objectives of a particular study. However, there is a gap for well-designed experiments with sound methodology and sample sizes that are sufficient to perform proper statistical analyses.

What raises concerns is the significant percentage of reports with missing data on the temporal organization of the studies. Excluding studies consisting of more than one session, almost 50% (117/263) did not specify the duration of the subjects' exposure to VR. Also, frequent shortcomings have been identified in 67 studies that consisted of two or more sessions (see Table 1). As a consequence, drawing conclusions on the basis of such reports may be biased and their replication in practice is not possible.

How the experimental groups were defined may raise objections in relation to some studies. In the case of 46 studies, there was no control group; moreover, as a point of reference for comparisons, another 9 studies used a group of people deprived of learning. In practice, such a decision makes it impossible to draw conclusions about the effectiveness of the tested methods.

An interesting trend concerns how the dependent variable is operationalized. To conclusively assess the effectiveness of a given tool, one should look at the effects of its application at each of the four levels, namely Reaction, Learning, Behavior, and Results (Kirkpatrick, 1976; 1994). One cannot ignore the fact that while the first two levels are very well represented in the reviewed studies, the effectiveness of VT was examined at the Behavior level in only 44 studies, and not a single attempt to examine the effectiveness of VT at the Results level was made. A partial explanation for this avoidance of this level could certainly be the definition of the fourth step of the assessment of training effectiveness. This definition says that, at the Results level, effects that go beyond the individual level should be assessed, which obviously can be troublesome, especially in the case of small sample sizes. However, from the perspective of policy making or the implementation of new technologies, it is necessary to examine the effectiveness of learning in VR at this level. Perhaps studies on VT involving a significant proportion of a limited population (e.g., a large proportion of employees in one enterprise) could serve as a way to fill this gap in the measurement of VT effectiveness at the Results level. It is also encouraging that a relatively large amount of research has not been limited to examining effectiveness on a single level.

In many cases (see, e.g., Loch et al., 2018; Xue et al., 2018) where statistical tests could have been applied, the evaluation is performed on the basis of small sample sizes and a simple comparison of means. Without being able to support claims with scientific evidence, one cannot prove that a virtual (or any other) training method is indeed effective. Furthermore, sometimes in research on the effectiveness of VT, only simple and informal assessment methods are used. This 'informality' of evaluation leads to the improper assessment of VT tools. What may appear to be an effective learning tool turns out, after closer look, to be a software application evaluated by a simple means report or comparison without any attempt even to estimate the level of statistical significance, not to mention the effect size (e.g., Verner et al., 2019). That being said, the subjective assessment method is undoubtedly a useful tool for evaluating training effectiveness, especially on the Reaction level (Kirkpatrick, 1976). However, it should not be the only method employed as it gives only part of the picture. The same situation arises with objective methods of learning evaluation. Using only one method is never enough for a complete evaluation of training effectiveness.

The lack of gold-standard methodological standards poses a great

**Table 1**

Imperfections of the longitudinal studies' descriptions in terms of their temporal organization.

Shortcoming	Frequency (%)
The length of a single session is not indicated	26 (38)
The number of single sessions is not indicated	14 (21)
The interval between sessions is not indicated	47 (70)

challenge to researchers. Many different self-report methods are used, of which some more popular ones can be highlighted, such as the NASA-Task Load Index (Hart & Staveland, 1988) or the System Usability Scale (Bangor et al., 2008; Brooke, 1996). For performance evaluation, different indices (e.g. time, error rate) are used, but these must be content-specific to some extent. The evaluation part is just a small section of many of the reviewed papers, while the entire text focuses on describing the process of developing the educational application. While this may be very useful, the lack of systematic and methodologically sound evaluation leads to many different tools being developed which are used primarily for academic purposes but are not necessarily applied in practice. This variability in methods and evaluation techniques makes different studies on the topic incomparable or at least extremely difficult to compare. The development of and adherence to some kind of evaluation framework could be beneficial to researchers and users of VEs.

It is also important to account for publication bias. In many of the reviewed studies, virtual tools proved to be effective in some way. However, we cannot know how many studies produced nonsignificant results or proved the inferiority of VT and were thus never published. Therefore, all conclusions about the extent to which VEs are effective for learning purposes should be treated with caution.

#### 4. Practical implications for further research and counteracting methodological imperfections

A framework for measuring the effectiveness of VT is proposed below. Because the design of VT applications is inherently grounded in a specific context, a set of general suggestions is also proposed that can be tailored to individual research requirements.

First, we propose that all fundamental principles of experimentation should be adhered to, such as planning a control condition and using statistical methods to draw conclusions regarding a population on the basis of a limited sample.

Second, we propose that at least two measurement time points should be used, but ideally it would be three: a pretest, an immediate post-test, and a retention test. Using a pre-test and a post-test allows gain scores to be calculated, previous knowledge to be accounted for, and ceiling effects to be controlled; using a retention test allows the learning curve to be observed. Knowing that such delayed measurements may result in participant dropout, we suggest that a week-long delay is a reasonable choice as it is a good balance between minimizing the risk of participant dropout and maximizing the benefits of a knowledge retention measurement. For retention tests, we suggest emailing participants with a link to an online tool.

Third, the learning outcomes that are measured in a study should be carefully selected. We recommend that at least two types of outcome are used to ensure a broad understanding of the learning process. In the process of choosing the learning outcomes and the methods by which they will be evaluated, theoretical frameworks should be applied. For example, Kirkpatrick's approach can be used, and methods to measure the Reaction and Learning levels should be carefully chosen (Kirkpatrick, 1976; 1994). When applicable, measurement of the higher levels of Kirkpatrick's model (Behavior and Results) can also be used. However, these levels can be difficult to measure in an experimental model since the learning process is very often not an inherent part of a specific workplace or educational institution. Concepts from the technology acceptance model (Davis, 1989) can be used to measure reactions to

training. In this way, we are able to grasp the wider context in which VT operates: not only how much knowledge participants gain but also how they perceive a given virtual tool and learning process.

Fourth, as there are many existing operationalizations of investigated outcomes that can be used, there is usually no need to create new ones. In the authors' opinion, measures such as NASA's task load index (Hart & Staveland, 1988) or the system usability scale (Bangor et al., 2008; Brooke, 1996) are fairly universal and can be used in many contexts. The selection of specific methods should be determined by substantive considerations, taking into account their popularity in a given domain. If a specific method is used in a given domain, it is worth including it in the research plan as an additional method in order to allow previous research to be treated as reference points. In this way, different studies on the effectiveness of VT could be directly compared with each other. Since individual VTs' performance parameters obviously differ, the situation may be more complicated in the case of objective performance indicators. However, even in this case, it is worth looking for universal performance indices. This will increase the comparability of the results of different studies and even, to a limited extent, of different VTs. The possible performance indices that could be useful are task completion time, error rate, efficiency of task performance, and knowledge test score.

If possible, self-report methods should be combined with objective methods as this makes it possible to look at the entire learning process from different perspectives. It would also be ideal to consider the possibility of using psychophysiological metrics and behavior observation by experts. Regarding objective skills and knowledge tests, ideally they should be measured in the target environment because such measurement makes it possible to assess the skill/knowledge transfer.

Lastly, taking into account publication bias and the realities of publishing scientific papers, it is worth considering minimizing the risk of committing errors of the first and second type (sample size determination and power analysis). This issue is naturally related to the possibility of research preregistration, which should also be considered. Moreover, considering that much research is conducted on VEs that are the subject of development work in which researchers themselves participate, particular attention should be paid to identifying potential conflicts of interest.

## 5. Strengths and limitations

The main strength of the present paper is that it covers a very wide array of study reports, including older ones from the 1990s. Despite the large number of recovered studies, we have made every effort to analyze and present their results with maximum reliability.

On the other hand, limiting the languages of these papers to those known to the authors (Polish and English) may have resulted in the loss of articles published in other languages (especially Chinese) that are the mother tongues of a significant number of the main contributors in the field of educational technologies, according to Chen, Zou, et al. (2020). There is a risk that some of the work of authors representing languages other than English and Polish has been published in their native languages; however, it can be expected that those authors' methodological approach is usually similar to that used in English-language publications. Thus, our limited range of articles can be treated as an imperfect but sufficient approximation.

The key distinguishing feature of our work is the fact that we focused specifically on the methods used in research on the effectiveness of VR-based learning, but not on the results of the reviewed studies. A significant number of the reviewed studies turned out to have methodological shortcomings or did not include statistical analysis that were as thorough as expected, which can be considered a limitation of the present review; however, considering that we focused mainly on methodology, the impact of this defect is limited. The fact that our analysis involved numerous studies from very diverse domains is certainly a greater limitation. Obviously, we are unable to present a precise picture of the "gold standard" in research on the effectiveness of VR-based learning, but we

hope that such a broad perspective will prove to be a valuable benchmark for researchers. There are two limitations to this framework. First, some or all of the criteria we suggest may seem obvious for many readers. Second, our intention was to propose the most-general guidelines that could be implemented in many domains after appropriate adaptation. Despite these limitations, we believe the present review provides valuable insights into the methodology of research on VR-based learning effectiveness and could serve as a basis for future studies focused on this issue.

## 6. Conclusions

The purpose of this paper was to summarize the research on the effectiveness of VT. Our main goal was to provide a comprehensive review of the methodological diversity and trends in this domain. This goal came from our observation that research on the effectiveness of didactics in VR is very diverse, and studies in this area sometimes lack common theoretical and methodological foundations. As a consequence, it is difficult to compare the results obtained in different studies. As a result of our review, we decided to propose five general methodological guidelines that could help unify future research and increase its impact. A large literature review was conducted to better understand the actual state of research in this field. Despite the exclusion of medical studies from the review, 330 studies were collected and analyzed.

It is not surprising that the amount of research conducted on the effectiveness of learning using VR has increased rapidly in recent years. Although similar studies have been published since the 1990s, 2019 seems to be a breakthrough year as the number of studies increased by 50% compared to the previous year. Also, a key change can be observed in 2019 regarding the popularity of research on tools that can be classified on the mixed reality continuum, as defined by Speicher et al. (2019). It was in 2019 that, for the first time, research on VR was more numerous than research on tools based on traditional displays. Taking into account the tools classified as Augmented Virtuality & Reality, this difference in popularity has become even more significant. We expect the observed trend to continue. This is of great importance for quality of learning because teaching tools of this class enable trainees to interact with learning material more effectively due to the wider possibilities of focusing attention in a computer-generated environment. VR is also associated with specific challenges, such as the need to consider immersion, presence, social presence, and visually induced motion sickness when designing and validating training tools (e.g. Dużmańska et al., 2018; Lipp et al., 2021). The growing popularity of research on the effectiveness of VR-based training tools goes hand in hand with the increasing number of practical areas in which such training is applied. From the number of published studies, it can be concluded that while the majority of research is invariably conducted in university environments, the amount of research in vocational and practical contexts is also high. This is of paramount importance because testing didactic tools on samples and in the contexts for which they are designed increases their accuracy.

We have previously expressed reservations about some of the methodological choices we revealed, particularly in relation to the temporal organization of the studies, specifying of the control group, and operationalization of the level of the dependent variable. We believe that the framework we propose may help reduce these imperfections.

The research described above provides interesting information on the methodology that is used to test the effectiveness of VR-based training tools, and it appears that there are still issues that require further improvement. In the final parts of the review, we indicated a few specific methodological shortcomings that would be worth eliminating in the future. In our opinion, compliance with the consistent minimum quality criteria for research in our area could help solve some of these problems. We have proposed such a framework that consists of five points: cross-checking of the basic methodological guidelines; measuring variables at more than one moment; conscious selection of the effectiveness



operationalization; striving to use proven post-survey tools; counter-acting conflicts of interests; and abandoning the publication of statistically insignificant results.

## Funding

This research was supported by the Polish National Centre for Research and Development under the grant 'Bioadaptable training simulator for critical infrastructure operators – research and preparation for implementation' (POIR.01.01.01-00-1131/17; the Smart Growth Operational Program, submeasure 1.1.1, received by Nano Games sp. z o.o.), the Priority Research Area Digiworld under the program Excellence Initiative – Research University at the Jagiellonian University in Kraków (U1U/P06/NO/02.34), National Science Center grant SONATA BIS (2021/42/E/HS6/00068).

## CRediT authorship contribution statement

**Paweł Strojny:** Conceptualization, Methodology, Writing – original draft, Writing – review & editing, Funding acquisition, Project administration, Visualization, Supervision. **Natalia Dużmańska-Misiarczyk:** Conceptualization, Data curation, Investigation, Formal analysis, Methodology, Writing – original draft, Visualization.

## Declaration of competing interest

None of the authors identified a conflict of interest in the preparation of this article.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.cexr.2022.100006>.

## References

- Alaker, M., Wynn, G. R., & Arulampalam, T. (2016). Virtual reality training in laparoscopic surgery: A systematic review & meta-analysis. *International Journal of Surgery*, 29, 85–94.
- Baddeley, A. (1992). Working memory. *Science*, 255(5044), 556–559.
- Bandura, A. (1982). Self-efficacy mechanism in human agency. *American Psychologist*, 37(2), 122.
- Bangor, A., Kortum, P., & Miller, J. A. (2008). The system usability scale (SUS): An empirical evaluation. *International Journal of Human-Computer Interaction*, 24(6), 574–594.
- Bates, R. (2004). A critical analysis of evaluation practice: The Kirkpatrick model and the principle of beneficence. *Evaluation and Program Planning*, 27(3), 341–347.
- Beaumont, C., Savin-Baden, M., Conradi, E., & Poulton, T. (2011). Evaluating a second life PBL demonstrator project: What can we learn. *Interactive Learning Environments*, 20(1), 5–21.
- Brooke, J. (1996). SUS: A 'quick and dirty' usability scale. In P. W. Jordan, B. Thomas, B. A. Weerdmeester, & I. L. McClelland (Eds.), *Usability evaluation in industry* (pp. 189–195). Taylor and Francis.
- Calderón, A., & Ruiz, M. (2015). A systematic literature review on serious games evaluation: An application to software project management. *Computers & Education*, 87, 396–422.
- Chandler, P., & Sweller, J. (1991). Cognitive load theory and the format of instruction. *Cognition and Instruction*, 8(4), 293–332.
- Chen, X., Xie, H., Zou, D., & Hwang, G. J. (2020). Application and theory gaps during the rise of artificial intelligence in education. *Computers and Education: Artificial Intelligence*, 1, Article 100002.
- Chen, X., Zou, D., Cheng, G., & Xie, H. (2020). Detecting latent topics and trends in educational technologies over four decades using structural topic modeling: A retrospective of all volumes of computers & education. *Computers & Education*, 151, Article 103855.
- Clark, J. M., & Paivio, A. (1991). Dual coding theory and education. *Educational Psychology Review*, 3(3), 149–210.
- Cruz-Neira, C., Sandin, D. J., DeFanti, T. A., Kenyon, R. V., & Hart, J. C. (1992). The CAVE: Audio visual experience automatic virtual environment. *Communications of the ACM*, 35(6), 64–73.
- Czarnek, G., Richter, M., & Strojny, P. (2021). Cardiac sympathetic activity during recovery as an indicator of sympathetic activity during task performance. *Psychophysiology*, 58(2), Article e13724.
- Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, 13(3), 319.
- Dengel, A., & Mägdefrau, J. (2020). Immersive learning predicted: Presence, prior knowledge, and school performance influence learning outcomes in immersive educational virtual environments. In *2020 6th International Conference of the Immersive Learning Research Network (ILRN)* (pp. 163–170). IEEE.
- Dużmańska, N., Strojny, P., & Strojny, A. (2018). Can simulator sickness be avoided? A review on temporal aspects of simulator sickness. *Frontiers in Psychology*, 9, 2132.
- Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G\*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39, 175–191.
- Fishbein, M., & Ajzen, I. (1975). *Belief, attitude, intention and behavior: An introduction to theory and research*. Addison-Wesley.
- Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (task load index): Results of empirical and theoretical research. *Advances in Psychology*, 52, 139–183 (North-Holland).
- King, W. R., & He, J. (2006). A meta-analysis of the technology acceptance model. *Information & Management*, 43(6), 740–755.
- Kirkpatrick, D. L. (1976). Evaluation of training. In R. L. Craig (Ed.), *Training and development handbook: A guide to human resource development*. McGraw Hill.
- Kirkpatrick, D. L. (1994). Evaluating training programs: The four levels. *Berrett-Koehler*.
- Kolb, D. A. (1984). *Experiential learning: Experience as the source of learning and development*. Prentice Hall.
- Lipp, N., Dużmańska-Misiarczyk, N., Strojny, A., & Strojny, P. (2021). Evoking emotions in virtual reality: Schema activation via a freeze-frame stimulus. *Virtual Reality*, 25(2), 279–292.
- Mayer, R. E. (1996). Learning strategies for making sense out of expository text: The SOI model for guiding three cognitive processes in knowledge construction. *Educational Psychology Review*, 8(4), 357–371.
- Mayer, R. E. (1997). Multimedia learning: Are we asking the right questions? *Educational Psychologist*, 32(1), 1–19.
- Mayer, R. E., & Moreno, R. (1998). A cognitive theory of multimedia learning: Implications for design principles. *Journal of Educational Psychology*, 91(2), 358–368.
- Milgram, P., & Kishino, F. (1994). A taxonomy of mixed reality visual displays. *IEICE - Transactions on Info and Systems*, 77(12), 1321–1329.
- Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akk, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., McGuinness, L. A., & Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, 372.
- Paivio, A. (1990). *Mental representations: A dual coding approach*. Oxford University Press.
- Parsons, T. D., Gaggioli, A., & Riva, G. (2017). Virtual reality for research in social neuroscience. *Brain Sciences*, 7(4), 42.
- Psotka, J. (1995). Immersive training systems: Virtual reality and education and training. *Instructional Science*, 23(5–6), 405–431.
- Sattar, M. U., Palaniappan, S., Lokman, A., Hassan, A., Shah, N., & Riaz, Z. (2019). Effects of Virtual Reality training on medical students' learning motivation and competency. *Pakistan Journal of Medical Sciences*, 35(3), 852.
- Schepers, J., & Wetzels, M. (2007). A meta-analysis of the technology acceptance model: Investigating subjective norm and moderation effects. *Information & Management*, 44(1), 90–103.
- Schroeder, R. (2008). Defining virtual worlds and virtual environments. *Journal For Virtual Worlds Research*, 1(1).
- Seymour, N. E., Gallagher, A. G., Roman, S. A., O'Brien, M. K., Bansal, V. K., Andersen, D. K., & Satava, R. M. (2002). Virtual reality training improves operating room performance: Results of a randomized, double-blinded study. *Annals of Surgery*, 236(4), 458.
- Shibata, T. (2002). Head mounted display. *Displays*, 23(1–2), 57–64.
- Shu, Y., Huang, Y. Z., Chang, S. H., & Chen, M. Y. (2019). Do virtual reality head-mounted displays make a difference? A comparison of presence and self-efficacy between head-mounted displays and desktop computer-facilitated virtual environments. *Virtual Reality*, 23(4), 437–446.
- Skarbez, R., Smith, M., & Whitton, M. C. (2021). Revisiting Milgram and Kishino's reality-virtuality continuum. *Frontiers in Virtual Reality*, 2, 27.
- Slater, M., & Wilbur, S. (1997). A framework for immersive virtual environments (FIVE): Speculations on the role of presence in virtual environments. *Presence: Teleoperators and Virtual Environments*, 6(6), 603–616.
- Smith, R. (2010). The long history of gaming in military training. *Simulation & Gaming*, 41(1), 6–19.
- Speicher, M., Hall, B. D., & Nebeling, M. (2019). What is mixed reality?. In *Proceedings of the 2019 CHI Conference on human factors in computing systems* (pp. 1–15).
- Sweller, J., Chandler, P., Tierney, P., & Cooper, M. (1990). Cognitive load as a factor in the structuring of technical material. *Journal of Experimental Psychology: General*, 119(2), 176.
- Wittrock, M. C. (1989). Generative processes of comprehension. *Educational Psychologist*, 24(4), 345–376.
- Zhao, J., Xu, X., Jiang, H., & Ding, Y. (2020). The effectiveness of virtual reality-based technology on anatomy teaching: A meta-analysis of randomized controlled studies. *BMC Medical Education*, 20(1), 1–10.

## Coded papers

- Abidi, M. H., Al-Ahmari, A., Ahmad, A., Ameen, W., & Alkhalefah, H. (2019). Evaluation of virtual reality-based manufacturing assembly training system. *International Journal of Advanced Manufacturing Technology*, 105(9), 3743–3759.

- Adams, A., Hart, J., Iacovides, I., Beavers, S., Oliveira, M., & Magroudi, M. (2019). Co-created evaluation: Identifying how games support police learning. *International Journal of Human-Computer Studies*, 132, 34–44.
- Ahn, S., Kim, T., Park, Y. J., & Kim, J. M. (2020). Improving effectiveness of safety training at construction worksite using 3D BIM simulation. *Advances in Civil Engineering*, 1–12, 2020.
- Akçayır, M., Akçayır, G., Pektaş, H. M., & Ocak, M. A. (2016). Augmented reality in science laboratories: The effects of augmented reality on university students' laboratory skills and attitudes toward science laboratories. *Computers in Human Behavior*, 57, 334–342.
- Albert, A., Hallowell, M. R., Kleiner, B., Chen, A., & Golparvar-Fard, M. (2014). Enhancing construction hazard recognition with high-fidelity augmented virtuality. *Journal of Construction Engineering and Management*, 140(7), Article 04014024.
- Alfred, M., Neyens, D. M., & Gramopadhye, A. K. (2018). Comparing learning outcomes in physical and simulated learning environments. *International Journal of Industrial Ergonomics*, 68, 110–117.
- Alrehaili, E. A., & Al Osman, H. (2019). A virtual reality role-playing serious game for experiential learning. *Interactive Learning Environments*, 1–14.
- Alves Fernandes, L. M., Cruz Matos, G., Azevedo, D., Rodrigues Nunes, R., Paredes, H., Morgado, L., ... Cardoso, B. (2016). Exploring educational immersive videogames: An empirical study with a 3D multimodal interaction prototype. *Behaviour & Information Technology*, 35(11), 907–918.
- Amati, M., & McNeill, M. (2012). Learning from and through virtual worlds: A pilot study of second life. *Journal for Education in the Built Environment*, 7(1), 39–55.
- Annetta, L., Lamb, R., Minogue, J., Folta, E., Holmes, S., Vallett, D., & Cheng, R. (2014). Safe science classrooms: Teacher training through serious educational games. *Information Sciences*, 264, 61–74.
- Annetta, L., Mangrum, J., Holmes, S., Collazo, K., & Cheng, M. T. (2009). Bridging reality to virtual reality: Investigating gender effect and student engagement on learning through video game play in an elementary school classroom. *International Journal of Science Education*, 31(8), 1091–1113.
- Austin, M. B. (1994). A comparison of three interactive media formats in an RTV/DVI computer-mediated lesson. *Computers & Education*, 22(4), 319–333.
- Aziz, E. S., Chang, Y., Esche, S. K., & Chassapis, C. (2015). Virtual mechanical assembly training based on a 3D game engine. *Computer-Aided Design and Applications*, 12(2), 119–134.
- Aziz, E. S., Corter, J. E., Chang, Y., Esche, S. K., & Chassapis, C. (2012). Evaluation of the learning effectiveness of game-based and hands-on gear train laboratories. In *Frontiers in education conference proceedings* (pp. 1–6). IEEE, 2012.
- Babály, B., & Kárpáti, A. (2016). The impact of creative construction tasks on visuospatial information processing and problem solving. *Acta Polytechnica Hungarica*, 13(7), 159–180.
- Baker, J., Parks-Savage, A., & Rehffuss, M. (2009). Teaching social skills in a virtual environment: An exploratory study. *Journal for Specialists in Group Work*, 34(3), 209–226.
- Balci, M. A., Akgüller, M., Kaya, E., Rzhavska, N., Dobroskok, I., Basiuk, L., & Kosa, T. (2020). Potential of game «PLANT TISSUE CULTURE» in bioengineering education as distance laboratory classes. *Computer Applications in Engineering Education*, 29(4), 855–863.
- Balid, W., Abdulwahed, M., & Alrouh, I. (2014). Development of an educationally oriented open-source embedded systems laboratory kit: A hybrid hands-on and virtual experimentation approach. *International Journal of Electrical Engineering Education*, 51(4), 340–353.
- Barnett, B., Perrin, B., Curtin, J., & Helbing, K. (1998). Can computer-based assessment be used for motor skills learning? A training transfer study. *Proceedings of the Human Factors and Ergonomics Society - Annual Meeting*, 42(20), 1432–1436.
- Bart, O., Katz, N., Weiss, P. L., & Josman, N. (2008). Street crossing by typically developed children in real and virtual environments. *OTJR: Occupation, Participation and Health*, 28(2), 89–96.
- Basalp, E., Marchal-Crespo, L., Rauter, G., Riener, R., & Wolf, P. (2019). Rowing simulator modulates water density to foster motor learning. *Frontiers in Robotics and AI*, 6, 74.
- Beer, U. M., Neerincx, M. A., Morina, N., & Brinkman, W. P. (2017). Virtual agent-mediated appraisal training: A single case series among Dutch firefighters. *European Journal of Psychotraumatology*, 8(1), Article 1378053.
- Belanich, J., Orvis, K. L., & Sibley, D. E. (2013). PC-based game features that influence instruction and learner motivation. *Military Psychology*, 25(3), 206–217.
- Bhagat, K. K., Liou, W. K., & Chang, C. Y. (2016). A cost-effective interactive 3D virtual reality system applied to military live firing training. *Virtual Reality*, 20(2), 127–140.
- Bhargava, A., Bertrand, J. W., Gramopadhye, A. K., Madathil, K. C., & Babu, S. V. (2018). Evaluating multiple levels of an interaction fidelity continuum on performance and learning in near-field training simulations. *IEEE Transactions on Visualization and Computer Graphics*, 24(4), 1418–1427.
- Binsubai, A., Maddock, S., & Romano, D. (2006). A serious game for traffic accident investigators. *Interactive Technology and Smart Education*, 3(4), 329–346.
- Bliss, J. P., Tidwell, P. D., & Guest, M. A. (1997). The effectiveness of virtual reality for administering spatial navigation training to firefighters. *Presence: Teleoperators and Virtual Environments*, 6(1), 73–86.
- Borsci, S., Lawson, G., Jha, B., Burges, M., & Salanitri, D. (2016). Effectiveness of a multidevice 3D virtual environment application to train car service maintenance procedures. *Virtual Reality*, 20(1), 41–55.
- Borsci, S., Lawson, G., Salanitri, D., & Jha, B. (2016). When simulated environments make the difference: The effectiveness of different types of training of car service procedures. *Virtual Reality*, 20(2), 83–99.
- Braghirolli, L. F., Ribeiro, J. L. D., Weise, A. D., & Pizzolato, M. (2016). Benefits of educational games as an introductory activity in industrial engineering education. *Computers in Human Behavior*, 58, 315–324.
- Brom, C., Preuss, M., & Klement, D. (2011). Are educational computer micro-games engaging and effective for knowledge acquisition at high-schools? A quasi-experimental study. *Computers & Education*, 57(3), 1971–1988.
- Brown, B., Pérez, G., Ribay, K., Boda, P. A., & Wilsey, M. (2021). Teaching culturally relevant science in virtual reality: “When a problem comes, you can solve it with science”. *Journal of Science Teacher Education*, 32(1), 7–38.
- Buil, I., Catalán, S., & Martínez, E. (2019). Encouraging intrinsic motivation in management training: The use of business simulation games. *International Journal of Management in Education*, 17(2), 162–171.
- Burigat, S., & Chittaro, L. (2016). Passive and active navigation of virtual environments vs. traditional printed evacuation maps: A comparative evaluation in the aviation domain. *International Journal of Human-Computer Studies*, 87, 92–105.
- Burnley, S. (2007). The use of virtual reality technology in teaching environmental engineering. *Engineering Education*, 2(2), 2–15.
- Butavicius, M. A., Vozzo, A., Braithwaite, H., & Galanis, G. (2012). Evaluation of a virtual reality parachute training simulator: Assessing learning in an off-course augmented feedback training schedule. *The International Journal of Aviation Psychology*, 22(3), 282–298.
- Cagiltay, N. E., Ozcelik, E., & Ozcelik, N. S. (2015). The effect of competition on learning in games. *Computers & Education*, 87, 35–41.
- Cai, S., Liu, E., Shen, Y., Liu, C., Li, S., & Shen, Y. (2020). Probability learning in mathematics using augmented reality: Impact on student's learning gains and attitudes. *Interactive Learning Environments*, 28(5), 560–573.
- Çakiroğlu, Ü., & Gökoğlu, S. (2019). Development of fire safety behavioral skills via virtual reality. *Computers & Education*, 133, 56–68.
- Calderón, A., Ruiz, M., & O'Connor, R. V. (2018). A serious game to support the ISO 21500 standard education in the context of software project management. *Computer Standards & Interfaces*, 60, 80–92.
- Calvert, J., & Abadia, R. (2020). Impact of immersing university and high school students in educational linear narratives using virtual reality technology. *Computers & Education*, 159, Article 104005.
- Cao, L., Lin, J., & Li, N. (2019). A virtual reality based study of indoor fire evacuation after active or passive spatial exploration. *Computers in Human Behavior*, 90, 37–45.
- Carenys, J., Moya, S., & Perramon, J. (2017). Is it worth it to consider videogames in accounting education? A comparison of a simulation and a videogame in attributes, motivation and learning outcomes: Merece la pena considerar los videojuegos en la enseñanza de contabilidad? Comparación de una simulación y un videojuego respecto a atributos, motivación y resultados de aprendizaje. *Revista de Contabilidad - Spanish Accounting Review*, 20(2), 118–130.
- Carlson, P., Peters, A., Gilbert, S. B., Vance, J. M., & Luse, A. (2015). Virtual training: Learning transfer of assembly tasks. *IEEE Transactions on Visualization and Computer Graphics*, 21(6), 770–782.
- Castro-García, M., Perez-Romero, A. M., León-Bonillo, M. J., & Manzano-Agugliaro, F. (2017). Developing topographic surveying software to train civil engineers. *Journal of Professional Issues in Engineering Education and Practice*, 143(1), Article 04016013.
- Chakraborty, P. R., & Bise, C. J. (2000). A virtual-reality-based model for task-training of equipment operators in the mining industry. *Mineral Resources Engineering*, 9, 437–449, 04.
- Champney, R. K., Stanney, K. M., Milham, L., Carroll, M. B., & Cohn, J. V. (2017). An examination of virtual environment training fidelity on training effectiveness. *International Journal of Learning Technology*, 12(1), 42–65.
- Chan, C. S., Chan, Y. H., & Fong, T. H. A. (2020). Game-based e-learning for urban tourism education through an online scenario game. *International Research in Geographical & Environmental Education*, 29(4), 283–300.
- Chang, K. E., Zhang, J., Huang, Y. S., Liu, T. C., & Sung, Y. T. (2020). Applying augmented reality in physical education on motor skills learning. *Interactive Learning Environments*, 28(6), 685–697.
- Chang, Y. S., Chou, C. H., Chuang, M. J., Li, W. H., & Tsai, I. F. (2020). Effects of virtual reality on creative design performance and creative experiential learning. *Interactive Learning Environments*, 1–16.
- Chen, C. C., & Huang, T. C. (2012). Learning in a u-Museum: Developing a context-aware ubiquitous learning environment. *Computers & Education*, 59(3), 873–883.
- Chen, C. J., Toh, S. C., & Ismail, W. M. F. W. (2005). Are learning styles relevant to virtual reality? *Journal of Research on Technology in Education*, 38(2), 123–141.
- Chen, C. M., & Tsai, Y. N. (2012). Interactive augmented reality system for enhancing library instruction in elementary schools. *Computers & Education*, 59(2), 638–652.
- Chen, J. C. (2016). The crossroads of English language learners, task-based instruction, and 3D multi-user virtual learning in Second Life. *Computers & Education*, 102, 152–171.
- Chen, X., Chen, Z., Li, Y., He, T., Hou, J., Liu, S., & He, Y. (2019). ImmerTai: Immersive motion learning in VR environments. *Journal of Visual Communication and Image Representation*, 58, 416–427.
- Chen, X., Liu, H., Bai, J., & Wu, Z. (2015). Evaluation model of simulation based training for helicopter emergency rescue mission. *International Journal of Multimedia and Ubiquitous Engineering*, 10(9), 69–80.
- Chen, Y. F. (2014). Evaluation of strategic emergency response training on an OLIVE platform. *Simulation & Gaming*, 45(6), 732–751.
- Chen, Y. F., & Mo, H. E. (2013). Users' perspectives on and learning effectiveness of tour-guide training courses with 3D tourist sites. *Journal of Teaching in Travel & Tourism*, 13(4), 374–390.
- Cheng, K. H., & Tsai, C. C. (2019). A case study of immersive virtual field trips in an elementary classroom: Students' learning experience and teacher-student interaction behaviors. *Computers & Education*, 140, Article 103600.
- Cheng, M. T., & Annetta, L. (2012). Students' learning outcomes and learning experiences through playing a Serious Educational Game. *Journal of Biological Education*, 46(4), 203–213.

- Chien, S. Y., Hwang, G. J., & Jong, M. S. Y. (2020). Effects of peer assessment within the context of spherical video-based virtual reality on EFL students' English-Speaking performance and learning perceptions. *Computers & Education*, 146, Article 103751.
- Chin, K. Y., Wang, C. S., & Chen, Y. L. (2019). Effects of an augmented reality-based mobile system on students' learning achievements and motivation for a liberal arts course. *Interactive Learning Environments*, 27(7), 927–941.
- Chittaro, L., Corbett, C. L., McLean, G. A., & Zangrando, N. (2018). Safety knowledge transfer through mobile virtual reality: A study of aviation life preserver donning. *Safety Science*, 102, 159–168.
- Chung-Shing, C., Yat-Hang, C., & Agnes, F. T. H. (2020). The effectiveness of online scenario game for ecotourism education from knowledge-attitude-usability dimensions. *Journal of Hospitality, Leisure, Sport & Tourism Education*, 27, Article 100264.
- Coleman, B., Marion, S., Rizzo, A., Turnbull, J., & Nolty, A. (2019). Virtual reality assessment of classroom-related attention: An ecologically relevant approach to evaluating the effectiveness of working memory training. *Frontiers in Psychology*, 10, 1851.
- Cone, B. L., Levy, S. S., & Goble, D. J. (2015). Wii Fit exergame training improves sensory weighting and dynamic balance in healthy young adults. *Gait & Posture*, 41(2), 711–715.
- Dalgarno, B., Bishop, A. G., Adlong, W., & Bedgood, D. R., Jr. (2009). Effectiveness of a virtual laboratory as a preparatory resource for distance education chemistry students. *Computers & Education*, 53(3), 853–865.
- Dalim, C. S. C., Sunar, M. S., Dey, A., & Billingham, M. (2020). Using augmented reality with speech input for non-native children's language learning. *International Journal of Human-Computer Studies*, 134, 44–64.
- Dalinger, T., Thomas, K. B., Stansberry, S., & Xiu, Y. (2020). A mixed reality simulation offers strategic practice for pre-service teachers. *Computers & Education*, 144, Article 103696.
- Dashkina, A. I., Khaliyapina, L. P., Kobicheva, A. M., Odinolaya, M. A., & Tarkhov, D. A. (2020). Developing a model of increasing the learners' bilingual professional capacity in the virtual laboratory environment. *Applied Sciences*, 10(20), 7022.
- Davis, A., Linvill, D. L., Hodges, L. F., Da Costa, A. F., & Lee, A. (2020). Virtual reality versus face-to-face practice: A study into situational apprehension and performance. *Communication Education*, 69(1), 70–84.
- Degli Innocenti, E., Geronazzo, M., Vescovi, D., Nordahl, R., Serafin, S., Ludovico, L. A., & Avanzini, F. (2019). Mobile virtual reality for musical genre learning in primary education. *Computers & Education*, 139, 102–117.
- den Haan, R. J., van der Voort, M. C., Baart, F., Berends, K. D., Van Den Berg, M., Straatsma, M. W., ... Hulscher, S. J. M. H. (2020). The Virtual River Game: Gaming using models to collaboratively explore river management complexity. *Environmental Modelling & Software*, 134, Article 104855.
- Detyna, M., & Kadiri, M. (2020). Virtual reality in the HE classroom: Feasibility, and the potential to embed in the curriculum. *Journal of Geography in Higher Education*, 44(3), 474–485.
- Devlin, A. M., Lally, V., Sclater, M., & Parussel, K. (2015). Inter-life: A novel, three-dimensional, virtual learning environment for life transition skills learning. *Interactive Learning Environments*, 23(4), 405–424.
- Diego-Mas, J. A., Alcaide-Marzal, J., & Poveda-Bautista, R. (2020). Effects of using immersive media on the effectiveness of training to prevent ergonomics risks. *International Journal of Environmental Research and Public Health*, 17(7), 2592.
- Ding, D., Brinkman, W. P., & Neerincx, M. A. (2020). Simulated thoughts in virtual reality for negotiation training enhance self-efficacy and knowledge. *International Journal of Human-Computer Studies*, 139, Article 102400.
- Do, P. T., & Moreland, J. R. (2014). Facilitating role of 3D multimodal visualization and learning rehearsal in memory recall. *Psychological Reports*, 114(2), 541–556.
- Dong, W., Li, Z., Yan, J., Wu, X., Yang, Y., & Zheng, L. (2008). Adaptive interaction in a 3D product structure browsing system for maintenance training. *Human Factors and Ergonomics in Manufacturing & Service Industries*, 18(1), 14–29.
- Dorozhkin, D., Olasky, J., Jones, D. B., Schwaizberg, S. D., Jones, S. B., Cao, C. G., ... De, S. (2017). OR fire virtual training simulator: Design and face validity. *Surgical Endoscopy*, 31(9), 3527–3533.
- Duffy, V. G., Ng, P. P., & Ramakrishnan, A. (2004). Impact of a simulated accident in virtual training on decision-making performance. *International Journal of Industrial Ergonomics*, 34(4), 335–348.
- Ebnali, M., Hulme, K., Ebnali-Heidari, A., & Mazloumi, A. (2019). How does training effect users' attitudes and skills needed for highly automated driving? *Transportation Research Part F: Traffic Psychology and Behaviour*, 66, 184–195.
- Ebnali, M., Lamb, R., Fathi, R., & Hulme, K. (2021). Virtual reality tour for first-time users of highly automated cars: Comparing the effects of virtual environments with different levels of interaction fidelity. *Applied Ergonomics*, 90, Article 103226.
- Echeverría, A., Barrios, E., Nussbaum, M., Améstica, M., & Leclerc, S. (2012). The atomic intrinsic integration approach: A structured methodology for the design of games for the conceptual understanding of physics. *Computers & Education*, 59(2), 806–816.
- Echeverría, A., García-Campo, C., Nussbaum, M., Gil, F., Villalta, M., Améstica, M., & Echeverría, S. (2011). A framework for the design and integration of collaborative classroom games. *Computers & Education*, 57(1), 1127–1136.
- Eiris, R., Gheisari, M., & Esmacelli, B. (2020). Desktop-based safety training using 360-degree panorama and static virtual reality techniques: A comparative experimental study. *Automation in Construction*, 109, Article 102969.
- Eiris, R., Jain, A., Gheisari, M., & Wehle, A. (2020). Safety immersive storytelling using narrated 360-degree panoramas: A fall hazard training within the electrical trade context. *Safety Science*, 127, Article 104703.
- Colombo, S., & Golzio, L. (2016). The Plant Simulator as viable means to prevent and manage risk through competencies management: Experiment results. *Safety Science*, 84, 46–56.
- Farra, S., Miller, E., Timm, N., & Schafer, J. (2013). Improved training for disasters using 3-D virtual reality simulation. *Western Journal of Nursing Research*, 35(5), 655–671.
- Feng, Z., González, V. A., Amor, R., Spearpoint, M., Thomas, J., Sacks, R., ... Cabrera-Guerrero, G. (2020). An immersive virtual reality serious game to enhance earthquake behavioral responses and post-earthquake evacuation preparedness in buildings. *Advanced Engineering Informatics*, 45, Article 101118.
- Feng, Z., González, V. A., Mutch, C., Amor, R., Rahouti, A., Baghouz, A., ... Cabrera-Guerrero, G. (2020). Towards a customizable immersive virtual reality serious game for earthquake emergency training. *Advanced Engineering Informatics*, 46, Article 101134.
- Fidan, M., & Tuncel, M. (2019). Integrating augmented reality into problem based learning: The effects on learning achievement and attitude in physics education. *Computers & Education*, 142, Article 103635.
- Fletcher, S., & Dodds, W. (2003). The use of a virtual learning environment to enhance ICM capacity building. *Marine Policy*, 27(3), 241–247.
- Frank, J. A., & Kapila, V. (2017). Mixed-reality learning environments: Integrating mobile interfaces with laboratory test-beds. *Computers & Education*, 110, 88–104.
- Gallegos-Nieto, E., Medellín-Castillo, H. I., González-Badillo, G., Lim, T., & Ritchie, J. (2017). The analysis and evaluation of the influence of haptic-enabled virtual assembly training on real assembly performance. *International Journal of Advanced Manufacturing Technology*, 89(1–4), 581–598.
- García, A. A., Bobadilla, I. G., Figueroa, G. A., Ramírez, M. P., & Román, J. M. (2016). Virtual reality training system for maintenance and operation of high-voltage overhead power lines. *Virtual Reality*, 20(1), 27–40.
- García, I., Pacheco, C., Leon, A., & Calvo-Manzano, J. A. (2020). A serious game for teaching the fundamentals of ISO/IEC/IEEE 29148 systems and software engineering–Lifecycle processes–Requirements engineering at undergraduate level. *Computer Standards & Interfaces*, 67, Article 103377.
- Gavish, N., Gutiérrez, T., Webel, S., Rodríguez, J., Peveri, M., Bockholt, U., & Tecchia, F. (2015). Evaluating virtual reality and augmented reality training for industrial maintenance and assembly tasks. *Interactive Learning Environments*, 23(6), 778–798.
- Georgakopoulos, V. (2010). Food safety training: A model HACCP instructional technique. *Tourismos: An International Multidisciplinary Journal of Tourism*, 5(1), 55–72.
- Gibson, M. E., Kim, J. J., McManus, M., & Harris, L. R. (2020). The effect of training on the perceived approach angle in visual vertical heading judgements in a virtual environment. *Experimental Brain Research*, 238, 1861–1869.
- Goel, L., Johnson, N. A., Junglas, I., & Ives, B. (2013). How cues of what can be done in a virtual world influence learning: An affordance perspective. *Information & Management*, 50(5), 197–206.
- Gonzalez-Badillo, G., Medellín-Castillo, H., Lim, T., Ritchie, J., & Garbaya, S. (2014). The development of a physics and constraint-based haptic virtual assembly system. *Assembly Automation*, 34(1), 41–55.
- Gonzalez-Franco, M., Pizarro, R., Cermeron, J., Li, K., Thorn, J., Hutabarat, W., ... Bermell-García, P. (2017). Immersive mixed reality for manufacturing training. *Frontiers in Robotics and AI*, 4, 3.
- Graham, K., Anderson, J., Rife, C., Heitmeyer, B., Patel, P. R., Nykl, S., ... Merkle, L. D. (2020). Cyberspace odyssey: A competitive team-oriented serious game in computer networking. *IEEE Transactions on Learning Technologies*, 13(3), 502–515.
- Grant, B. L., Yelder, P. C., Patrick, T. A., Kapralos, B., Williams-Bell, M., & Murphy, B. A. (2020). Audiohaptic feedback enhances motor performance in a low-fidelity simulated drilling task. *Brain Sciences*, 10(1), 21.
- Guillén-Nieto, V., & Aleson-Carbonell, M. (2012). Serious games and learning effectiveness: The case of It's a Deal. *Computers & Education*, 58(1), 435–448.
- Haferkamp, N., Kraemer, N. C., Linehan, C., & Schembri, M. (2011). Training disaster communication by means of serious games in virtual environments. *Entertainment Computing*, 2(2), 81–88.
- Haginoya, S., Yamamoto, S., Pompèda, F., Naka, M., Antfolk, J., & Santtila, P. (2020). Online simulation training of child sexual abuse interviews with feedback improves interview quality in Japanese university students. *Frontiers in Psychology*, 11, 998.
- Hamza-Lup, F. G., & Goldbach, I. R. (2021). Multimodal, visuo-haptic games for abstract theory instruction: Grabbing charged particles. *Journal on Multimodal User Interfaces*, 15, 1–10.
- Harley, J. M., Lajoie, S. P., Tressel, T., & Jarrell, A. (2020). Fostering positive emotions and history knowledge with location-based augmented reality and tour-guide prompts. *Learning and Instruction*, 70, Article 101163.
- Hart, J. L., & Proctor, M. D. (2020). Behaving socially with a virtual human role-player in a simulated counseling session. *The Journal of Defense Modeling and Simulation*, 17(2), 175–189.
- Hassani, K., Nahvi, A., & Ahmadi, A. (2016). Design and implementation of an intelligent virtual environment for improving speaking and listening skills. *Interactive Learning Environments*, 24(1), 252–271.
- Hays, R. T., & Vincenzi, D. A. (2000). Fleet assessments of a virtual reality training system. *Military Psychology*, 12(3), 161–186.
- Hernandez-Pozas, O., & Carreon-Flores, H. (2019). Teaching international business using virtual reality. *Journal of Teaching in International Business*, 30(2), 196–212.
- Herrington, A., & Tacy, J. (2020). Crossing the power line: Using virtual simulation to prepare the first responders of utility Linemen. *Informatics*, 7(3), 26.
- Ho, C. M. (2010). What's in a question? The case of students' enactments in the second life virtual world. *Innovation in Language Learning and Teaching*, 4(2), 151–176.
- Hoareau, C., Querec, R., Buche, C., & Ganier, F. (2017). Evaluation of internal and external validity of a virtual environment for learning a long procedure. *International Journal of Human-Computer Interaction*, 33(10), 786–798.
- Hodges, N., Watchravesringkan, K., Min, S., Lee, Y., & Seo, S. (2020). Teaching virtual apparel technology through industry collaboration: An assessment of pedagogical



- process and outcomes. *International Journal of Fashion Design, Technology and Education*, 13(2), 120–130.
- Hoedt, S., Claeys, A., Van Landeghem, H., & Cottyn, J. (2017). The evaluation of an elementary virtual training system for manual assembly. *International Journal of Production Research*, 55(24), 7496–7508.
- Holbrook, H. A., & Cennamo, K. S. (2014). Effects of high-fidelity virtual training simulators on learners' self-efficacy. *International Journal of Gaming and Computer-Mediated Simulations*, 6(2), 38–52.
- Hou, H. T., & Lin, Y. C. (2017). The development and evaluation of an educational game integrated with augmented reality and virtual laboratory for chemistry experiment learning. In *2017 6th IIAI international congress on advanced applied informatics (IIAI-AAI)* (pp. 1005–1006). IEEE.
- Howard, M. C. (2017). Investigating the simulation elements of environment and control: Extending the Uncanny Valley Theory to simulations. *Computers & Education*, 109, 216–232.
- Hsu, L. (2012). Web 3D simulation-based application in tourism education: A case study with second life. *Journal of Hospitality, Leisure, Sports and Tourism Education*, 11(2), 113–124.
- Huang, W. (2019). Examining the impact of head-mounted display virtual reality on the science self-efficacy of high schoolers. *Interactive Learning Environments*, 30(1), 100–112.
- Huang, Y. C., Backman, S. J., & Backman, K. F. (2010). Student attitude toward virtual learning in second life: A flow theory approach. *Journal of Teaching in Travel & Tourism*, 10(4), 312–334.
- Huang, Z., Javaid, A., Devabhaktuni, V. K., Li, Y., & Yang, X. (2019). Development of cognitive training program with EEG headset. *IEEE Access*, 7, 126191–126200.
- Hubal, R. C., & Day, R. S. (2006). Informed consent procedures: An experimental test using a virtual character in a dialog systems training application. *Journal of Biomedical Informatics*, 39(5), 532–540.
- Hudlicka, E. (2013). Virtual training and coaching of health behavior: Example from mindfulness meditation training. *Patient Education and Counseling*, 92(2), 160–166.
- Ibáñez, M. B., Di Serio, A., Villarán, D., & Kloos, C. D. (2014). Experimenting with electromagnetism using augmented reality: Impact on flow student experience and educational effectiveness. *Computers & Education*, 71, 1–13.
- Ijaz, K., Bogdanovych, A., & Trescak, T. (2017). Virtual worlds vs books and videos in history education. *Interactive Learning Environments*, 25(7), 904–929.
- Jacobson, M. J., Taylor, C. E., & Richards, D. (2016). Computational scientific inquiry with virtual worlds and agent-based models: New ways of doing science to learn science. *Interactive Learning Environments*, 24(8), 2080–2108.
- Jayaram, S., Connacher, H. I., & Lyons, K. W. (1997). Virtual assembly using virtual reality techniques. *Computer-aided design*, 29(8), 575–584.
- Jeelani, I., Han, K., & Albert, A. (2020). Development of virtual reality and stereo-panoramic environments for construction safety training. *Engineering Construction and Architectural Management*, 27(8), 1853–1876.
- Jia, D., Bhatti, A., & Nahavandi, S. (2014). The impact of self-efficacy and perceived system efficacy on effectiveness of virtual training systems. *Behaviour & Information Technology*, 33(1), 16–35.
- Jin, S. A. A. (2011). Leveraging avatars in 3D virtual environments (Second Life) for interactive learning: The moderating role of the behavioral activation system vs. behavioral inhibition system and the mediating role of enjoyment. *Interactive Learning Environments*, 19(5), 467–486.
- Johnson, D. M., & Stewart, J. E. (1999). Use of virtual environments for the acquisition of spatial knowledge: Comparison among different visual displays. *Military Psychology*, 11(2), 129–148.
- Johnson, D., Damian, D., & Tzanetakis, G. (2020). Evaluating the effectiveness of mixed reality music instrument learning with the theremin. *Virtual Reality*, 24(2), 303–317.
- Johnson-Glenberg, M. C., Birchfield, D., & Usyal, S. (2009). SMALLab: Virtual geology studies using embodied learning with motion, sound, and graphics. *Educational Media International*, 46(4), 267–280.
- Jong, M. S. (2015). Does online game-based learning work in formal education at school? A case study of VISOLE. *Curriculum Journal*, 26(2), 249–267.
- Jost, P., Cobb, S., & Hämmerle, I. (2020). Reality-based interaction affecting mental workload in virtual reality mental arithmetic training. *Behaviour & Information Technology*, 39(10), 1062–1078.
- Jou, M., & Wang, J. (2013). Investigation of effects of virtual reality environments on learning performance of technical skills. *Computers in Human Behavior*, 29(2), 433–438.
- Juang, J. R., Hung, W. H., & Kang, S. C. (2013). SimCrane 3D+: A crane simulator with kinesthetic and stereoscopic vision. *Advanced Engineering Informatics*, 27(4), 506–518.
- Jung, T., Ineson, E. M., Hains, C., & Kim, M. (2013). Contributors to hospitality students' knowledge enhancement. *Journal of Hospitality, Leisure, Sports and Tourism Education*, 13, 97–106.
- Kaber, D. B., Riley, J. M., Endsley, M. R., Sheik-Nainar, M., Zhang, T., & Lampton, D. R. (2013). Measuring situation awareness in virtual environment-based training. *Military Psychology*, 25(4), 330–344.
- Kamińska, D., Zwoliński, G., Wiak, S., Petkovska, L., Cvetkovski, G., Barba, P. D., Mognaschi, M. E., Haamer, R. E., & Anbarjafari, G. (2020). Virtual reality-based training: Case study in mechatronics. *Technology, Knowledge and Learning*, 26(4), 1043–1059.
- Kao, C. W. (2020). The effect of a digital game-based learning task on the acquisition of the English Article System. *System*, 95, Article 102373.
- Kaphingst, K. A., Persky, S., McCall, C., Lachance, C., Loewenstein, J., Beall, A. C., & Blasovich, J. (2009). Testing the effects of educational strategies on comprehension of a genomic concept using virtual reality technology. *Patient Education and Counseling*, 77(2), 224–230.
- Katz, Y. J. (1999). Kindergarten teacher training through virtual reality: Three-dimensional simulation methodology. *Educational Media International*, 36(2), 151–156.
- Ke, F. (2008). A case study of computer gaming for math: Engaged learning from gameplay? *Computers & Education*, 51(4), 1609–1620.
- Kebritchi, M., Hirumi, A., & Bai, H. (2010). The effects of modern mathematics computer games on mathematics achievement and class motivation. *Computers & Education*, 55(2), 427–443.
- Kim, H., & Ke, F. (2017). Effects of game-based learning in an OpenSim-supported virtual environment on mathematical performance. *Interactive Learning Environments*, 25(4), 543–557.
- Kim, H., Ke, F., & Paek, I. (2017). Game-based learning in an OpenSim-supported virtual environment on perceived motivational quality of learning. *Technology, Pedagogy and Education*, 26(5), 617–631.
- Kim, K. G., Oertel, C., Dobricki, M., Olsen, J. K., Coppi, A. E., Cattaneo, A., & Dillenbourg, P. (2020). Using immersive virtual reality to support designing skills in vocational education. *British Journal of Educational Technology*, 51(6), 2199–2213.
- King, S. A., Dzenga, C., Burch, T., & Kennedy, K. (2021). Teaching partial-interval recording of problem behavior with virtual reality. *Journal of Behavioral Education*, 30(2), 202–225.
- Koo, G., Lee, N., & Kwon, O. (2019). Combining object detection and causality mining for efficient development of augmented reality-based on-the-job training systems in hotel management. *New Review in Hypermedia and Multimedia*, 25(3), 112–136.
- Koumaditis, K., Chinello, F., Mitkidis, P., & Karg, S. (2020). Effectiveness of virtual versus physical training: The case of assembly tasks, trainer's verbal assistance, and task complexity. *IEEE Computer Graphics and Applications*, 40(5), 41–56.
- Lackey, S. J., Salcedo, J. N., Szalma, J. L., & Hancock, P. A. (2016). The stress and workload of virtual reality training: The effects of presence, immersion and flow. *Ergonomics*, 59(8), 1060–1072.
- Langley, A., Lawson, G., Hermawati, S., D'cruz, M., Apold, J., Arlt, F., & Mura, K. (2016). Establishing the usability of a virtual training system for assembly operations within the automotive industry. *Human Factors and Ergonomics in Manufacturing & Service Industries*, 26(6), 667–679.
- Larrue, F., Sauzeon, H., Wallet, G., Foloppe, D., Cazalets, J. R., Gross, C., & N'Kaoua, B. (2014). Influence of body-centered information on the transfer of spatial learning from a virtual to a real environment. *Journal of Cognitive Psychology*, 26(8), 906–918.
- Leder, J., Horlitz, T., Puschmann, P., Wittstock, V., & Schütz, A. (2019). Comparing immersive virtual reality and powerpoint as methods for delivering safety training: Impacts on risk perception, learning, and decision making. *Safety Science*, 111, 271–286.
- Lee, E. A. L., & Wong, K. W. (2014). Learning with desktop virtual reality: Low spatial ability learners are more positively affected. *Computers & Education*, 79, 49–58.
- Lee, E. K. O. (2014). Use of avatars and a virtual community to increase cultural competence. *Journal of Technology in Human Services*, 32(1–2), 93–107.
- Lee, K. M., Jeong, E. J., Park, N., & Ryu, S. (2011). Effects of interactivity in educational games: A mediating role of social presence on learning outcomes. *International Journal of Human-Computer Interaction*, 27(7), 620–633.
- Legrand, F. D., Joly, P. M., Bertucci, W. M., Soudain-Pineau, M. A., & Marcel, J. (2011). Interactive-virtual reality (IVR) exercise: An examination of in-task and pre-to-post exercise affective changes. *Journal of Applied Sport Psychology*, 23(1), 65–75.
- Levine, J., & Adams, R. H. (2013). Introducing case management to students in a virtual world: An exploratory study. *Journal of Teaching in Social Work*, 33(4–5), 552–565.
- Li, C., Ip, H. H., Wong, Y. M., & Lam, W. S. (2020). An empirical study on using virtual reality for enhancing the youth's intercultural sensitivity in Hong Kong. *Journal of Computer Assisted Learning*, 36(5), 625–635.
- Li, C., Liang, W., Quigley, C., Zhao, Y., & Yu, L. F. (2017). Earthquake safety training through virtual drills. *IEEE Transactions on Visualization and Computer Graphics*, 23(4), 1275–1284.
- Li, J. C., Tang, J. K., & Jia, C. X. (2019). Virtual offender: A pilot project on nurturing social work students' capacity to work with offenders. *China Journal of Social Work*, 12(1), 56–69.
- Li, J., Li, Y., Bao, H., Yu, C., & Guo, H. (2020). An empirical study of the virtual simulation system teaching method in NC machining. *International Journal of Technology and Human Interaction*, 16(3), 109–123.
- Liang, J. S. (2008). A study on virtual reality-based EDM learning framework and effectiveness analysis. *Computer-Aided Design and Applications*, 5(1–4), 391–400.
- Lin, F., Ye, L., Duffy, V. G., & Su, C. J. (2002). Developing virtual environments for industrial training. *Information Sciences*, 140(1–2), 153–170.
- Lin, H. H., Yen, W. C., & Wang, Y. S. (2018). Investigating the effect of learning method and motivation on learning performance in a business simulation system context: An experimental study. *Computers & Education*, 127, 30–40.
- Lin, M. C., Tutwiler, M. S., & Chang, C. Y. (2011). Exploring the relationship between virtual learning environment preference, use, and learning outcomes in 10th grade earth science students. *Learning, Media and Technology*, 36(4), 399–417.
- Lin, P. H., & Yeh, S. C. (2019). How motion-control influences a VR-supported technology for mental rotation learning: From the perspectives of playfulness, gender difference and technology acceptance model. *International Journal of Human-Computer Interaction*, 35(18), 1736–1746.
- Link, M. W., Armsby, P. P., Hubal, R. C., & Guinn, C. I. (2006). Accessibility and acceptance of responsive virtual human technology as a survey interviewer training tool. *Computers in Human Behavior*, 22(3), 412–426.
- Liou, H. C. (2012). The roles of Second Life in a college computer-assisted language learning (CALL) course in Taiwan, ROC. *Computer Assisted Language Learning*, 25(4), 365–382.



- Liu, G., Zhang, Y., Lu, K., & Liu, L. (2014). A hybrid haptic guidance model for tank gunners in high precision and high speed motor skill training. *Multimedia Tools and Applications*, 69(3), 1111–1130.
- Liu, T. Y., & Chu, Y. L. (2010). Using ubiquitous games in an English listening and speaking course: Impact on learning outcomes and motivation. *Computers & Education*, 55(2), 630–643.
- Loch, F., Koltun, G., Karaseva, V., Pantförder, D., & Vogel-Heuser, B. (2018). Model-based training of manual procedures in automated production systems. *Mechatronics*, 55, 212–223.
- Lok, B., Naik, S., Whitton, M., & Brooks, F. P. (2003). Effects of handling real objects and self-avatar fidelity on cognitive task performance and sense of presence in virtual environments. *Presence: Teleoperators and Virtual Environments*, 12(6), 615–628.
- Loup-Escande, E., Jamet, E., Ragot, M., Erhel, S., & Michinov, N. (2017). Effects of stereoscopic display on learning and user experience in an educational virtual environment. *International Journal of Human-Computer Interaction*, 33(2), 115–122.
- Lovreglio, R., Duan, X., Rahouti, A., Phipps, R., & Nilsson, D. (2021). Comparing the effectiveness of fire extinguisher virtual reality and video training. *Virtual Reality*, 25, 133–145.
- Lugrin, J. L., Latoschik, M. E., Habel, M., Roth, D., Seufert, C., & Grafe, S. (2016). Breaking bad behaviors: A new tool for learning classroom management using virtual reality. *Frontiers in ICT*, 3, 26.
- Luo, H., Yang, T., Kwon, S., Zuo, M., Li, W., & Choi, I. (2020). Using virtual reality to identify and modify risky pedestrian behaviors amongst Chinese children. *Traffic Injury Prevention*, 21(1), 108–113.
- Makransky, G., Borre-Gude, S., & Mayer, R. E. (2019). Motivational and cognitive benefits of training in immersive virtual reality based on multiple assessments. *Journal of Computer Assisted Learning*, 35(6), 691–707.
- Makransky, G., Mayer, R. E., Veitch, N., Hood, M., Christensen, K. B., & Gadegaard, H. (2019). Equivalence of using a desktop virtual reality science simulation at home and in class. *PLoS One*, 14(4), Article e0214944.
- Makransky, G., Terkildsen, T. S., & Mayer, R. E. (2019). Adding immersive virtual reality to a science lab simulation causes more presence but less learning. *Learning and Instruction*, 60, 225–236.
- Maratou, V., Chatzidakis, E., & Xenos, M. (2016). Enhance learning on software project management through a role-play game in a virtual world. *Interactive Learning Environments*, 24(4), 897–915.
- Martín-Gutiérrez, J., Saorín, J. L., Contero, M., Alcañiz, M., Pérez-López, D. C., & Ortega, M. (2010). Design and validation of an augmented book for spatial abilities development in engineering students. *Computers & Graphics*, 34(1), 77–91.
- Martín-SanJosé, J. F., Juan, M. C., Gil-Gómez, J. A., & Rando, N. (2014). Flexible learning itinerary vs. linear learning itinerary. *Science of Computer Programming*, 88, 3–21.
- Mathews, S., Andrews, L., & Luck, E. (2012). Developing a second life virtual field trip for university students: An action research approach. *Educational Research*, 54(1), 17–38.
- Mayne, R., & Green, H. (2020). Virtual reality for teaching and learning in crime scene investigation. *Science & Justice*, 60(5), 466–472.
- Mead, C., Buxner, S., Bruce, G., Taylor, W., Semken, S., & Anbar, A. D. (2019). Immersive, interactive virtual field trips promote science learning. *Journal of Geoscience Education*, 67(2), 131–142.
- Merchant, Z., Goetz, E. T., Keeney-Kennicutt, W., Kwok, O. M., Cifuentes, L., & Davis, T. J. (2012). The learner characteristics, features of desktop 3D virtual reality environments, and college chemistry instruction: A structural equation modeling analysis. *Computers & Education*, 59(2), 551–568.
- Moreno, R. (2009). Learning from animated classroom exemplars: The case for guiding student teachers' observations with metacognitive prompts. *Educational Research and Evaluation*, 15(5), 487–501.
- Morrongioello, B. A., Corbett, M., Beer, J., & Koutsoulianos, S. (2018). A pilot randomized controlled trial testing the effectiveness of a pedestrian training program that teaches children where and how to cross the street safely. *Journal of Pediatric Psychology*, 43(10), 1147–1159.
- Moskaliuk, J., Bertram, J., & Cress, U. (2013). Impact of virtual training environments on the acquisition and transfer of knowledge. *Cyberpsychology, Behavior, and Social Networking*, 16(3), 210–214.
- Münzer, S., & Zadeh, M. V. (2016). Acquisition of spatial knowledge through self-directed interaction with a virtual model of a multi-level building: Effects of training and individual differences. *Computers in Human Behavior*, 64, 191–205.
- Murcia-López, M., & Steed, A. (2018). A comparison of virtual and physical training transfer of bimanual assembly tasks. *IEEE Transactions on Visualization and Computer Graphics*, 24(4), 1574–1583.
- Narciso, D., Melo, M., Raposo, J. V., Cunha, J., & Bessa, M. (2020). Virtual reality in training: An experimental study with firefighters. *Multimedia Tools and Applications*, 79(9), 6227–6245.
- Nathanael, D., Mosialos, S., & Vosniakos, G. C. (2016). Development and evaluation of a virtual training environment for on-line robot programming. *International Journal of Industrial Ergonomics*, 53, 274–283.
- Nathanael, D., Mosialos, S., Vosniakos, G. C., & Tsagkas, V. (2016). Development and evaluation of a virtual reality training system based on cognitive task analysis: The case of CNC tool length offsetting. *Human Factors and Ergonomics in Manufacturing & Service Industries*, 26(1), 52–67.
- Neri, L., Noguez, J., & Robledo-Rella, V. (2010). Improving problem-solving skills using adaptive on-line training and learning environments. *International Journal of Engineering Education*, 26(6), 1316.
- Noteborn, G., Carbonell, K. B., Dailey-Hebert, A., & Gijssels, W. (2012). The role of emotions and task significance in virtual education. *The Internet and Higher Education*, 15(3), 176–183.
- Nykänen, M., Puro, V., Tiikkaja, M., Kannisto, H., Lantto, E., Simpura, F., Uusitalo, J., Lukander, K., Räsänen, T., Heikkilä, T., & Teperi, A. M. (2020). Implementing and evaluating novel safety training methods for construction sector workers: Results of a randomized controlled trial. *Journal of Safety Research*, 75, 205–221.
- Oberdörfer, S., & Latoschik, M. E. (2019). Predicting learning effects of computer games using the Gamified Knowledge Encoding Model. *Entertainment Computing*, 32, Article 100315.
- Okutsu, M., DeLaurentis, D., Brophy, S., & Lambert, J. (2013). Teaching an aerospace engineering design course via virtual worlds: A comparative assessment of learning outcomes. *Computers & Education*, 60(1), 288–298.
- Orvis, K. A., Horn, D. B., & Belanich, J. (2008). The roles of task difficulty and prior videogame experience on performance and motivation in instructional videogames. *Computers in Human Behavior*, 24(5), 2415–2433.
- Osti, F., de Amicis, R., Sanchez, C. A., Tilt, A. B., Prather, E., & Liverani, A. (2020). A VR training system for learning and skills development for construction workers. *Virtual Reality*, 25(2), 523–538.
- Pagé, C., Bernier, P. M., & Trempe, M. (2019). Using video simulations and virtual reality to improve decision-making skills in basketball. *Journal of Sports Sciences*, 37(21), 2403–2410.
- Papachristos, N. M., Vrellis, I., Natsis, A., & Mikropoulos, T. A. (2014). The role of environment design in an educational Multi-User Virtual Environment. *British Journal of Educational Technology*, 45(4), 636–646.
- Papastergiou, M. (2009). Digital game-based learning in high school computer science education: Impact on educational effectiveness and student motivation. *Computers & Education*, 52(1), 1–12.
- Park, J., Liu, D., Mun, Y. Y., & Santhanam, R. (2019). GAMESIT: A gamified system for information technology training. *Computers & Education*, 142, Article 103643.
- Pasin, F., & Giroux, H. (2011). The impact of a simulation game on operations management education. *Computers & Education*, 57(1), 1240–1254.
- Pataty, T. C., & Lamb, P. F. (2018). Effects of physical randomness training on virtual and laboratory golf putting performance in novices. *Journal of Sports Sciences*, 36(12), 1355–1362.
- Patera, M., Draper, S., & Naef, M. (2008). Exploring magic cottage: A virtual reality environment for stimulating children's imaginative writing. *Interactive Learning Environments*, 16(3), 245–263.
- Patiar, A., Kensbock, S., Benckendorff, P., Robinson, R., Richardson, S., Wang, Y., & Lee, A. (2020). Hospitality students' acquisition of knowledge and skills through a virtual field trip experience. *Journal of Hospitality and Tourism Education*, 33(1), 14–28.
- Patiar, A., Kensbock, S., Ma, E., & Cox, R. (2017). Information and communication technology-Enabled innovation: Application of the virtual field trip in hospitality education. *Journal of Hospitality and Tourism Education*, 29(3), 129–140.
- Patiar, A., Ma, E., Kensbock, S., & Cox, R. (2017). Hospitality management students' expectation and perception of a virtual field trip web site: An Australian case study using importance–performance analysis. *Journal of Hospitality and Tourism Education*, 29(1), 1–12.
- Paxinou, E., Georgiou, M., Kakkos, V., Kalles, D., & Galani, L. (2020). Achieving educational goals in microscopy education by adopting virtual reality labs on top of face-to-face tutorials. *Research in Science & Technological Education*, 1–20.
- Pedram, S., Palmisano, S., Skarbez, R., Perez, P., & Farrelly, M. (2020). Investigating the process of mine rescuers' safety training with immersive virtual reality: A structural equation modelling approach. *Computers & Education*, 153, Article 103891.
- Peng, C., Cao, L., & Timalsena, S. (2017). Gamification of Apollo lunar exploration missions for learning engagement. *Entertainment Computing*, 19, 53–64.
- Peralbo-Uzquiano, M., Fernández-Abella, R., Durán-Bouza, M., Brenlla-Blanco, J. C., & Cotos-Yáñez, J. M. (2020). Evaluation of the effects of a virtual intervention programme on cognitive flexibility, inhibitory control and basic math skills in childhood education. *Computers & Education*, 159, Article 104006.
- Perera, I., Miller, A., & Allison, C. (2016). A case study in user support for managing OpenSim based multi user learning environments. *IEEE Transactions on Learning Technologies*, 10(3), 342–354.
- Pérez, L., Diez, E., Usamentiaga, R., & García, D. F. (2019). Industrial robot control and operator training using virtual reality interfaces. *Computers in Industry*, 109, 114–120.
- Perez-Romero, A. M., Castro-Garcia, M., Leon-Bonillo, M. J., & Manzano-Agugliaro, F. (2017). Learning effectiveness of virtual environments for 3D terrain interpretation and data acquisition. *Survey Review*, 49(355), 302–311.
- Petersen, G. B., Klingenberg, S., Mayer, R. E., & Makransky, G. (2020). The virtual field trip: Investigating how to optimize immersive virtual learning in climate change education. *British Journal of Educational Technology*, 51(6), 2099–2115.
- Peterson, S. M., Rios, E., & Ferris, D. P. (2018). Transient visual perturbations boost short-term balance learning in virtual reality by modulating electrocortical activity. *Journal of Neurophysiology*, 120(4), 1998–2010.
- Pickering, C. E., Ridenour, K., Salaysay, Z., Reyes-Gastelum, D., & Pierce, S. J. (2018). EATI Island-A virtual-reality-based elder abuse and neglect educational intervention. *Gerontology & Geriatrics Education*, 39(4), 445–463.
- Pinheiro, A., Fernandes, P., Maia, A., Cruz, G., Pedrosa, D., Fonseca, B., ... Rafael, J. (2012). Development of a mechanical maintenance training simulator in Open Simulator for F-16 aircraft engines. *Procedia Computer Science*, 15, 248–255.
- Pompedda, F., Palu, A., Kask, K., Schiff, K., Soveri, A., Antfolk, J., & Santtila, P. (2020). Transfer of simulated interview training effects into interviews with children exposed to a mock event. *Nordic Psychology*, 73(1), 43–67.
- Pringle, J. K. (2013). Educational environmental geoscience e-gaming to provide stimulating and effective learning. *Planet*, 27(1), 21–28.
- Ragan, E. D., Bowman, D. A., Kopper, R., Stinson, C., Scerbo, S., & McMahan, R. P. (2015). Effects of field of view and visual complexity on virtual reality training effectiveness for a visual scanning task. *IEEE Transactions on Visualization and Computer Graphics*, 21(7), 794–807.

- Ranchhod, A., Gurău, C., Loukis, E., & Trivedi, R. (2014). Evaluating the educational effectiveness of simulation games: A value generation model. *Information Sciences*, 264, 75–90.
- Rein, B. A., McNeil, D. W., Hayes, A. R., Hawkins, T. A., Ng, H. M., & Yura, C. A. (2018). Evaluation of an avatar-based training program to promote suicide prevention awareness in a college setting. *Journal of American College Health*, 66(5), 401–411.
- Reinsmith-Jones, K., Kibbe, S., Crayton, T., & Campbell, E. (2015). Use of second life in social work education: Virtual world experiences and their effect on students. *Journal of Social Work Education*, 51(1), 90–108.
- Richards, D., & Taylor, M. (2015). A Comparison of learning gains when using a 2D simulation tool versus a 3D virtual world: An experiment to find the right representation involving the Marginal Value Theorem. *Computers & Education*, 86, 157–171.
- Rogers, C. B., El-Mounayri, H., Wasfy, T., & Satterwhite, J. (2018). Assessment of STEM e-learning in an immersive virtual reality (VR) environment. *ASEE Computers in Education Journal*, 8(4), 1–9.
- Rossler, K. L., Ganesh Sankaranarayanan, D., & Duvall, A. (2019). Acquisition of fire safety knowledge and skills with virtual reality simulation. *Nurse Educator*, 44(2), 88.
- Sacks, R., Perlman, A., & Barak, R. (2013). Construction safety training using immersive virtual reality. *Construction Management & Economics*, 31(9), 1005–1017.
- Salleh, S. M., Tasir, Z., & Shukor, N. A. (2012). Web-based simulation learning framework to enhance students' critical thinking skills. *Procedia-Social and Behavioral Sciences*, 64, 372–381.
- Samsudin, K., Rafi, A., & Mohamad Ali, A. Z. (2014). Enhancing a low-cost virtual reality application through constructivist approach: The case of spatial training of Middle graders. *Turkish Online Journal of Educational Technology-TOJET*, 13(3), 50–57.
- Sarabando, C., Cravino, J. P., & Soares, A. A. (2014). Contribution of a computer simulation to students' learning of the physics concepts of weight and mass. *Procedia Technology*, 13, 112–121.
- Saus, E. R., Johnsen, B. H., Eid, J., Riisem, P. K., Andersen, R., & Thayer, J. F. (2006). The effect of brief situational awareness training in a police shooting simulator: An experimental study. *Military Psychology*, 18(sup1), 3–21.
- Sánchez, A., Redondo, E., Fonseca, D., & Navarro, I. (2013). Hand-held augmented reality: Usability and academic performance assessment in educational environments. Case study of an engineering degree course. *Information*, 16, 8621–8634.
- Schiller, S. Z., Goodrich, K., & Gupta, P. B. (2013). Let them play! Active learning in a virtual world. *Information Systems Management*, 30(1), 50–62.
- Schmitz, B., Klemke, R., Walhout, J., & Specht, M. (2015). Attuning a mobile simulation game for school children using a design-based research approach. *Computers & Education*, 81, 35–48.
- Schrader, C., & Bastiaens, T. J. (2012). The influence of virtual presence: Effects on experienced cognitive load and learning outcomes in educational computer games. *Computers in Human Behavior*, 28(2), 648–658.
- Serge, S. R., Priest, H. A., Durlach, P. J., & Johnson, C. I. (2013). The effects of static and adaptive performance feedback in game-based training. *Computers in Human Behavior*, 29(3), 1150–1158.
- Severengiz, M., Seliger, G., & Krüger, J. (2020). Serious game on factory planning for higher education. *Procedia Manufacturing*, 43, 239–246.
- Shen, H., Zhang, J., Yang, B., & Jia, B. (2019). Development of an educational virtual reality training system for marine engineers. *Computer Applications in Engineering Education*, 27(3), 580–602.
- Shi, Y., Du, J., & Worthy, D. A. (2020). The impact of engineering information formats on learning and execution of construction operations: A virtual reality pipe maintenance experiment. *Automation in Construction*, 119, Article 103367.
- Shi, Y., Du, J., Ahn, C. R., & Ragan, E. (2019). Impact assessment of reinforced learning methods on construction workers' fall risk behavior using virtual reality. *Automation in Construction*, 104, 197–214.
- Shi, Y., Zhu, Y., Mehta, R. K., & Du, J. (2020). A neurophysiological approach to assess training outcome under stress: A virtual reality experiment of industrial shutdown maintenance using functional near-infrared spectroscopy (fNIRS). *Advanced Engineering Informatics*, 46, Article 101153.
- Sholihin, M., Sari, R. C., Yuniarti, N., & Ilyana, S. (2020). A new way of teaching business ethics: The evaluation of virtual reality-based learning media. *The International Journal of Management Education*, 18(3), Article 100428.
- Sidhu, M. S. (2015). The effects of using learning-aided cues in an augmented reality environment for a multi-body mechanism. *International Journal of Computer Applications in Technology*, 52(4), 220–227.
- Singh, G., Mantri, A., Sharma, O., & Kaur, R. (2021). Virtual reality learning environment for enhancing electronics engineering laboratory experience. *Computer Applications in Engineering Education*, 29(1), 229–243.
- Smith, M. J., Bornheimer, L. A., Li, J., Blajeski, S., Hiltz, B., Fischer, D. J., ... Ruffolo, M. (2021). Computerized clinical training simulations with virtual clients abusing alcohol: Initial feasibility, acceptability, and effectiveness. *Clinical Social Work Journal*, 49(2), 184–196.
- Sommerauer, P., & Müller, O. (2014). Augmented reality in informal learning environments: A field experiment in a mathematics exhibition. *Computers & Education*, 79, 59–68.
- Sportillo, D., Paljic, A., & Ojeda, L. (2018). Get ready for automated driving using virtual reality. *Accident Analysis & Prevention*, 118, 102–113.
- Squelch, A. P. (2001). Virtual reality for mine safety training in South Africa. *Journal of the Southern African Institute of Mining and Metallurgy*, 101(4), 209–216.
- Srinivasa, A. R., Jha, R., Ozkan, T., & Wang, Z. (2020). Virtual reality and its role in improving student knowledge, self-efficacy, and attitude in the materials testing laboratory. *International Journal of Mechanical Engineering Education*, 49(4), 382–409.
- Stone, R. T., Watts, K. P., & Zhong, P. (2011). Virtual reality integrated welder training. *Welding Journal*, 90(7), 136s.
- Sutcliffe, A., & Alrayes, A. (2012). Investigating user experience in Second Life for collaborative learning. *International Journal of Human-Computer Studies*, 70(7), 508–525.
- Szalma, J. L., Daly, T. N., Teo, G. W. L., Hancock, G. M., & Hancock, P. A. (2018). Training for vigilance on the move: A video game-based paradigm for sustained attention. *Ergonomics*, 61(4), 482–505.
- Tagliabue, M., & Sarlo, M. (2015). Affective components in training to ride safely using a moped simulator. *Transportation Research Part F: Traffic Psychology and Behaviour*, 35, 132–138.
- Tagliabue, M., Gianfranchi, E., & Sarlo, M. (2017). A first step toward the understanding of implicit learning of hazard anticipation in inexperienced road users through a moped-riding simulator. *Frontiers in Psychology*, 8, 768.
- Tai, T. Y., Chen, H. H. J., & Todd, G. (2020). The impact of a virtual reality app on adolescent EFL learners' vocabulary learning. *Computer Assisted Language Learning*, 35(4), 892–917.
- Tan, J. L., Goh, D. H. L., Ang, R. P., & Huan, V. S. (2016). Learning efficacy and user acceptance of a game-based social skills learning environment. *International Journal of Cild-computer Interaction*, 9, 1–19.
- Tang, Y. M., Au, K. M., & Leung, Y. (2018). Comprehending products with mixed reality: Geometric relationships and creativity. *International Journal of Engineering Business Management*, 10, Article 1847979018809599.
- Tang, Y. M., Au, K. M., Lau, H. C., Ho, G. T., & Wu, C. H. (2020). Evaluating the effectiveness of learning design with mixed reality (MR) in higher education. *Virtual Reality*, 24(4), 797–807.
- Tegos, S., Demetriadis, S., & Karakostas, A. (2015). Promoting academically productive talk with conversational agent interventions in collaborative learning settings. *Computers & Education*, 87, 309–325.
- ter Vrugte, J., de Jong, T., Vandercruysse, S., Wouters, P., van Oostendorp, H., & Elen, J. (2015). How competition and heterogeneous collaboration interact in prevocational game-based mathematics education. *Computers & Education*, 89, 42–52.
- Tiernan, P. (2010). Enhancing the learning experience of undergraduate technology students with LabVIEW™ software. *Computers & Education*, 55(4), 1579–1588.
- Toh, C., Miller, S., & Simpson, T. (2015). The impact of virtual product dissection environments on student design learning and self-efficacy. *Journal of Engineering Design*, 26(1–3), 48–73.
- Treves, R., Viterbo, P., & Haklay, M. (2015). Footprints in the sky: Using student track logs from a “bird's eye view” virtual field trip to enhance learning. *Journal of Geography in Higher Education*, 39(1), 97–110.
- Trundle, K. C., & Bell, R. L. (2010). The use of a computer simulation to promote conceptual change: A quasi-experimental study. *Computers & Education*, 54(4), 1078–1088.
- Tsai, M. H., Chang, Y. L., Shiau, J. S., & Wang, S. M. (2020). Exploring the effects of a serious game-based learning package for disaster prevention education: The case of Battle of Flooding Protection. *International Journal of Disaster Risk Reduction*, 43, Article 101393.
- Turkan, Y., Radkowski, R., Karabulut-Ilgu, A., Behzadan, A. H., & Chen, A. (2017). Mobile augmented reality for teaching structural analysis. *Advanced Engineering Informatics*, 34, 90–100.
- van der Meij, H. (2013). Do pedagogical agents enhance software training? *Human-Computer Interaction*, 28(6), 518–547.
- van Ginkel, S., Gulikers, J., Biemans, H., Noroozi, O., Roozen, M., Bos, T., ... Mulder, M. (2019). Fostering oral presentation competence through a virtual reality-based task for delivering feedback. *Computers & Education*, 134, 78–97.
- Vando, S., Unim, B., Cassarino, S., Paulo, J., & Masala, D. (2013). Effectiveness of perceptual training-proprioceptive feedback in a virtual visual diverse group of healthy subjects: A pilot study. *Epidemiology Biostatistics and Public Health*, 10(2), e-8844.
- Vargas González, A. N., Kapalo, K., Koh, S. L., & LaViola, J. J. (2017). Exploring the virtuality continuum for complex rule-set education in the context of soccer rule comprehension. *Multimodal Technologies and Interaction*, 1(4), 30.
- Venkatesh, V., & Speier, C. (2000). Creating an effective training environment for enhancing telework. *International Journal of Human-Computer Studies*, 52(6), 991–1005.
- Vélaz, Y., Rodríguez Arce, J., Gutiérrez, T., Lozano-Rodero, A., & Suescun, A. (2014). The influence of interaction technology on the learning of assembly tasks using virtual reality. *Journal of Computing and Information Science in Engineering*, 14(4).
- Verner, I. M., Cuperman, D., Gamer, S., & Polishuk, A. (2019). Training robot manipulation skills through practice with digital twin of baxter. *International Journal of Online and Biomedical Engineering (iJOE)*, 15, 58, 09.
- Vidotto, G., Tagliabue, M., & Tira, M. D. (2015). Long-lasting virtual motorcycle-riding trainer effectiveness. *Frontiers in Psychology*, 6, 1653.
- Villena Taranilla, R., Cózar-Gutiérrez, R., González-Calero, J. A., & López Cirugeda, I. (2019). Strolling through a city of the Roman empire: An analysis of the potential of virtual reality to teach history in primary education. *Interactive Learning Environments*, 30(4), 608–618.
- Virvou, M., Katsionis, G., & Manos, K. (2005). Combining software games with education: Evaluation of its educational effectiveness. *Journal of Educational Technology & Society*, 8(2), 54–65.
- Waller, D., Hunt, E., & Knapp, D. (1998). The transfer of spatial knowledge in virtual environment training. *Presence*, 7(2), 129–143.
- Wallis, G., & Tichon, J. (2013). Predicting the efficacy of simulator-based training using a perceptual judgment task versus questionnaire-based measures of presence. *Presence*, 22(1), 67–85.

- Wang, J., Guo, D., & Jou, M. (2015). A study on the effects of model-based inquiry pedagogy on students' inquiry skills in a virtual physics lab. *Computers in Human Behavior*, 49, 658–669.
- Wang, P. Y., Vaughn, B. K., & Liu, M. (2011). The impact of animation interactivity on novices' learning of introductory statistics. *Computers & Education*, 56(1), 300–311.
- Wang, P., Wu, P., Chi, H. L., & Li, X. (2020). Adopting lean thinking in virtual reality-based personalized operation training using value stream mapping. *Automation in Construction*, 119, Article 103355.
- Watson, P., & Livingstone, D. (2018). Using mixed reality displays for observational learning of motor skills: A design research approach enhancing memory recall and usability. *Research in Learning Technology*, 26, 0.
- Webster, R. (2016). Declarative knowledge acquisition in immersive virtual learning environments. *Interactive Learning Environments*, 24(6), 1319–1333.
- Wener, R., Panindre, P., Kumar, S., Feygina, I., Smith, E., Dalton, J., & Seal, U. (2015). Assessment of web-based interactive game system methodology for dissemination and diffusion to improve firefighter safety and wellness. *Fire Safety Journal*, 72, 59–67.
- Williamon, A., Aufegger, L., & Eiholzer, H. (2014). Simulating and stimulating performance: Introducing distributed simulation to enhance musical learning and performance. *Frontiers in Psychology*, 5, 25.
- Winkelmann, K., Keeney-Kennicutt, W., Fowler, D., Lazo Macik, M., Perez Guarda, P., & Joan Ahlborn, C. (2020). Learning gains and attitudes of students performing chemistry experiments in an immersive virtual world. *Interactive Learning Environments*, 28(5), 620–634.
- Witmer, B. G., Sadowski, W. J., & Finkelstein, N. M. (2002). VE-based training strategies for acquiring survey knowledge. *Presence: Teleoperators and Virtual Environments*, 11(1), 1–18.
- Wrzesien, M., & Raya, M. A. (2010). Learning in serious virtual worlds: Evaluation of learning effectiveness and appeal to students in the E-Junior project. *Computers & Education*, 55(1), 178–187.
- Wu, A., Zhang, W., & Zhang, X. (2009). Evaluation of wayfinding aids in virtual environment. *Intl. Journal of Human-Computer Interaction*, 25(1), 1–21.
- Wu, W., Hartless, J., Tesi, A., Gunji, V., Ayer, S., & London, J. (2019). Design assessment in virtual and mixed reality environments: Comparison of novices and experts. *Journal of Construction Engineering and Management*, 145(9), Article 04019049.
- Xu, D., Huang, W. W., Wang, H., & Heales, J. (2014). Enhancing e-learning effectiveness using an intelligent agent-supported personalized virtual learning environment: An empirical investigation. *Information & Management*, 51(4), 430–440.
- Xu, J., Tang, Z., Zhao, H., & Zhang, J. (2019). Hand gesture-based virtual reality training simulator for collaboration rescue of a railway accident. *Interacting with Computers*, 31(6), 577–588.
- Xue, S. U. N., Hu, L. I. U., Guanghui, W. U., & Yaoming, Z. H. O. U. (2018). Training effectiveness evaluation of helicopter emergency relief based on virtual simulation. *Chinese Journal of Aeronautics*, 31(10), 2000–2012.
- Xue, S. U. N., Hu, L. I. U., Yongliang, T. I. A. N., Guanghui, W. U., & Yuan, G. A. O. (2020). Team effectiveness evaluation and virtual reality scenario mapping model for helicopter emergency rescue. *Chinese Journal of Aeronautics*, 33(12), 3306–3317.
- Yamani, Y., Samuel, S., Knodler, M. A., & Fisher, D. L. (2016). Evaluation of the effectiveness of a multi-skill program for training younger drivers on higher cognitive skills. *Applied Ergonomics*, 52, 135–141.
- Yang, Q. F., Chang, S. C., Hwang, G. J., & Zou, D. (2020). Balancing cognitive complexity and gaming level: Effects of a cognitive complexity-based competition game on EFL students' English vocabulary learning performance, anxiety and behaviors. *Computers & Education*, 148, Article 103808.
- Yasin, A., Liu, L., Li, T., Wang, J., & Zowghi, D. (2018). Design and preliminary evaluation of a cyber security requirements education game (SREG). *Information and Software Technology*, 95, 179–200.
- Zawadzki, P., Żywicki, K., Buń, P., & Górski, F. (2020). Employee training in an intelligent factory using virtual reality. *IEEE Access*, 8, 135110–135117.
- Zhang, H., Yu, L., Ji, M., Cui, Y., Liu, D., Li, Y., Liu, H., & Wang, Y. (2020). Investigating high school students' perceptions and presences under VR learning environment. *Interactive Learning Environments*, 28(5), 635–655.
- Zhong, B., Zheng, J., & Zhan, Z. (2020). An exploration of combining virtual and physical robots in robotics education. *Interactive Learning Environments*, 1–13.
- Zhou, S. P., Ting, S. P., Shen, Z. Q., & Luo, L. B. (2008). Twilight City—A virtual environment for MOUT. *International Journal of Computers and Applications*, 30(2), 117–123.
- Zinchenko, Y. P., Khoroshikh, P. P., Sergievich, A. A., Smirnov, A. S., Tumyalis, A. V., Kovalev, A. I., Gutnikov, S. A., & Golokhvast, K. S. (2020). Virtual reality is more efficient in learning human heart anatomy especially for subjects with low baseline knowledge. *New Ideas in Psychology*, 59, Article 100786.