



**HAL**  
open science

## An analysis of VR technology used in immersive simulations with a serious game perspective

Aline Menin, Rafael Torchelsen, Luciana Nedel

► **To cite this version:**

Aline Menin, Rafael Torchelsen, Luciana Nedel. An analysis of VR technology used in immersive simulations with a serious game perspective. *IEEE Computer Graphics and Applications*, 2018, 10.1109/mcg.2018.021951633 . hal-03404738

**HAL Id: hal-03404738**

**<https://hal.science/hal-03404738v1>**

Submitted on 26 Oct 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# An analysis of VR technology used in immersive simulations with a serious game perspective

Aline Menin, Rafael Torchelsen, and Luciana Nedel, *Member, IEEE*

## Abstract

The use of Virtual Environments (VEs) is a safer and cost-effective alternative for executing dangerous tasks, such as training firefighters and industrial operators. Immersive VR combined with game aspects have the potential to improve the user experience in VEs by increasing realism, engagement, and motivation. This work investigates the impact of VR technology on 46 immersive “gamified” simulations with serious purposes and classifies it towards a taxonomy. Our findings suggest that immersive VR improves simulations outcomes, increasing learning gain, knowledge retention, and improving clinical outcomes for rehabilitation. However, it also presents limitations such as motion sickness and restricted access to VR hardware by people. Our contributions are to provide a better understanding of the benefits and limitations of using VR in immersive simulations with serious purposes, to propose a taxonomy that classifies them, and to discuss their assessment investigating whether methods and participants profile influence on results.

## Index Terms

Virtual Reality, Virtual and Augmented Reality, Computers and Society, Computer Applications, Computer and Education, Serious Games

## I. INTRODUCTION

Virtual Reality (VR) is defined as a medium of communication that requires physical immersion, provides synthetic sensory stimulation, can mentally immerse the user, and is interactive[s01]. Immersion, according to Slater [1] “*stands for what the technology delivers from an objective point of view. The more a system delivers displays (in all sensory modalities) and tracking that preserves fidelity in relation to their equivalent real-world sensory modalities, the more it is 'immersive'*”. Muhanna [2] classifies VR systems toward a taxonomy with three levels of immersion: low, partially, and fully immersive systems. The lowest level of immersion is seen in basic virtual systems, such as hand and monitor-based; partially immersive VR systems use a single projector to display a virtual world on a large screen, and fully immersive systems provide the participant with three-dimensional virtual scenes in a large field of view.

VR systems providing low or partial levels of immersion have been widely used to train professionals in domains as diverse as medicine[s02], firefighting[s03] and industry[s04], as well as for education[s05] and health[s06]. These applications are presented as games with a serious purpose (*serious games*). According to Djaouti et al. [3], the oxymoron “Serious Games (SGs)” emerged in 2002 carrying a definition later refined by Sawyer as “*any meaningful use of computerized game/game industry resources whose chief mission is not entertainment*”[s07]. Zyda[s08] describes it as being “*a mental contest, played with a computer in accordance with specific rules, that uses entertainment to further government or corporate training, education, health, public policy, and strategic communication objectives*”. Finally, according to Chen & Michael[s09] they are “*games that do not have entertainment, enjoyment or fun as their primary purpose*”.

Equivalently to video games, SGs consist of gameplay, challenge, interaction, and objective[s6], while “gamified” applications merely incorporate elements of games. Therefore, gamification is defined by Deterding et al. [4] as “*the use of game design elements in non-game contexts*”. These elements could be related to the interaction design components and design solutions for a known problem in a context, such as badge, leaderboard and level; gameplay components, such as time constraint, limited resources, and turns; evaluative guidelines, such as enduring play and clear goals; conceptual models of components of games (e.g., challenge, fantasy, curiosity); and practices and processes, such as play-testing and play-centric design.

Furthermore, they also differ by the fact that SGs are constructed over pedagogical and educational frameworks, which define the relationship between learning and game mechanisms, ensuring a successful combination of such factors to reach its serious purpose. The framework presented by Ibáñez et al.[s10] covers six facets of the SG development: learning objectives, which defines a frame of reference of the domain to teach; domain simulation, which defines a formal model settling the basements of the simulation; interaction with the domain simulation, which is the heart of the metaphor; problems and progression, which defines the level of difficulty and skills progression; decoration, which describes how to entertain and engage the player; and deployment, which describes the use conditions to preserve the learning qualities of the game. The theoretical framework proposed by Suttie et al.[s11] examines the individual learning and play components of successful SGs to identify the SG

A. Menin is with Federal University of Rio Grande do Sul, Institute of Informatics, 91509-900 Brazil, and Univ. Grenoble Alpes, CNRS, Grenoble INP, LIG, F-38000 Grenoble, France, e-mail: aline.menin@univ-grenoble-alpes.fr

R. Torchelsen is with Federal University of Pelotas, CDTEC, 96010-610 Brazil, e-mail: rafael.torchelsen@inf.ufpel.edu.br

L. Nedel is with Federal University of Rio Grande do Sul, Institute of Informatics, 91509-900 Brazil, e-mail: nedel@inf.ufrgs.br.

mechanics that have proven successful and advance towards a practical conceptual design tool for effectively implementing educational mechanisms for future SGs. De Freitas and Oliver’s framework helps tutors to evaluate the potential of using games and simulation-based learning in their practice[12].

Game design techniques can be useful to create more engaging VR applications, which could make them more attractive to the public. These engaging applications combined with the increasing availability of consumer’s Head-Mounted Displays (HMDs) could bring a future in which people train and learn by using immersive VR. Higher levels of immersion have presented positive effects in small-scale spatial judgments [5] and memorization [6]. The combination of these factors leads to memorization of complex procedures [7], which allows the Virtual Environment (VE) to produce training and learning outcomes. Bowman et al. [8] provide empirical evidence that high levels of immersion can also produce a measurable improvement in the performance of an abstract mental activity. Chalmers et al. [9] claim that high level of realism is needed to ensure that training and learning in VEs are equivalent to the real world. They should also be able to simulate all human senses simultaneously.

Immersive VR combined with game techniques leads then to the development of immersive simulations with game aspects that enable engagement in learning activities otherwise expensive or too dangerous, difficult or impractical to implement in the classroom. Their use helps to change learners’ relationships to information by encouraging visualization, experimentation, and creativity. They are flexible and complex enough to cater different learning styles, and they broaden learners’ exposure to different people and perspectives, encourage collaboration and support meaningful post-game discussion [10].

This paper presents an analysis of VR technology used in immersive “gamified” simulations with serious purposes, such as training, learning, and treatment, pointing out its advantages and limitations. We propose a taxonomy to classify these applications according to VE’s features, serious purpose, target public profile, and methods used in user studies, considering the evaluation of effectiveness, usability, and immersion. Further, we identify the profile of the experiments’ volunteers to investigate whether it influences the results.

We structured this document as follows. Section II presents the research questions, method, and scope that structured this review. Section III presents the taxonomy developed to classify the immersive “gamified” simulations. Section IV presents the results classifying them towards our taxonomy. In Section V we discuss our results pointing to benefits and limitations of using immersive VR in simulations designed for serious purposes. Finally, we present our conclusions.

## II. METHODOLOGY

An extensive, but not exhaustive, search of the literature on using immersive VR technology on simulations with game aspects designed for serious purposes was undertaken. To define the scope of the study, we structured our analysis through a set of research questions (described below). Then, the research papers were chosen based on five inclusion criteria: it should present at least one immersive simulation, being or not evaluated through user studies; the simulation should contain at least one of the game elements previously described and a serious purpose; the game should provide an enhanced VR system following Muhanna’s taxonomy; the authors should properly describe the game/user study to support its classification and analysis; and the paper should not be a review of other papers.

We analyzed every article’s abstract and dropped or accepted it for a further analysis. The accepted ones were read to identify whether their content matches the scope of the study and they were then cataloged. In the next loop, we began the search with the reference list of the relevant papers. The last publications search date was December, 2016.

### A. Research Questions

#### 1) Does VR technology influences immersive simulations outcomes?

The surveyed simulations usually contain physical interaction or feedback, which grant to the user higher levels of immersion and, consequently, may improve their performance during the game. However, VR displays still have limitations, such as connection wires and low resolution, which might cause motion sickness and discomfort. Thereby, these drawbacks on usability could reduce user experience and affect the simulation outcomes, which would make it less effective towards its serious purpose (e.g., training and learning). We identified the current technological setting employed to provide immersion and investigated possible impacts on usability and effectiveness.

#### 2) How is the validation of immersive simulations conducted?

The surveyed simulations contain the serious purpose factor, which intends to train, learn or treat people. Thereby, the system must be correctly validated in order to be used in real world. We analyzed every paper to understand how this validation is executed, emphasizing procedures, experimental design, participants, and metrics used to measure the variables, which changes according to each simulation objective and/or the experimental hypotheses. In particular, we considered effectiveness, usability, and immersion aspects. The simulation must be effective towards its purpose, being usable to every user, and the immersive VR aspects should be considered to grant that it would be beneficial to the “gamified” application.

#### 3) What is the profile of subjects taking part in the experiments and does this profile affects the outcomes?

Since every immersive simulation presented in this paper has a specific purpose, their users should have a particular profile, which would be consistent with the serious purpose and application area. For instance, a simulation designed to treat a

disease should have users described as patients carrying the particular illness treated by the application. Nevertheless, to design and implement user experiments may be a hard-working task, and, not rarely, it can be difficult to find volunteers matching the simulation's target public to participate in the study. Then, students might be recruited, particularly because the research is normally conducted in Universities research laboratories where, consequently, undergraduate students are easily available. However, their profile would be different from the real user, because of their proximity to the research environment. Therefore, we correlate the target public profile, the participants' profile and experiments results to investigate whether replacing the target public influences the study results, affecting the simulation validity.

### III. TOWARDS A TAXONOMY FOR IMMERSIVE "GAMIFIED" SIMULATIONS

Based on the literature reviewed we identify the following main criteria as important for the classification of immersive "gamified" simulation:

- **Display Device:** The VR device employed to display the VE. We identified the use of large-screens, three-walls, CAVEs and HMDs.
- **Level of Immersion:** We identified the level of immersion provided by the VE following the taxonomy for VR systems presented by Muhanna, in which devices such as HMDs and CAVEs are included into fully immersive systems, while large-screens and three-walls are considered partially immersive systems.
- **Interaction:** Every surveyed simulation provides interaction by direct tracking the user's body parts: the head, the hands and fingers, the eyes, the torso, the feet and other body parts, or by indirect tracking it using props and platforms. A few games provide interaction by touch or voice, but we did not consider them in the classification because it was a small sample.
- **Feedback:** We observed the types of sensory feedback supported by the VR system and whether the simulation provided performance feedback. Sensory feedback is an essential element of VR and depends on participants' physical position. Though it is mainly provided by audiovisual channels, haptic feedback has been applied in VR systems enabling the user to "touch" and "feel" virtual objects. In our sample, we identified the use of haptic, visual, and aural feedback. Performance feedback guides the user through the simulation, and thus it might help to improve user performance and the simulator effectiveness.
- **Serious purpose:** The surveyed simulations were designed for five purposes: learning, training, evaluation, treatment, and self-caring. Learning simulations support knowledge acquisition and content understanding. In simulations designed for training, users are introduced into a VE to which they already possess the required knowledge to perform the proposed training, so they can improve their skills on this matter. Eventually, it could as well improve the content understanding through its practice. Evaluation means to verify users' reaction when facing a specific situation or to assess their knowledge about a given content. The simulations for treatment of patients are commonly employed for rehabilitation and therapy. Finally, VEs designed for self-caring frequently aims to encourage people to be more physically active.
- **Target public:** The surveyed simulations are applied in diverse areas, such as medicine, aviation, sports, driving, military, firefighting, industry, education, and safety. Therefore, based on serious purposes and application areas we determined a set of profiles to classify the simulation target public: the professional category includes people such as machines and industrial operators, pilots, and surgeons; patient refers to the people that have the specific disease treated by the simulation, such as children, elderly, and others, which includes people with a non-specific profile, such as the ones that play the games for self-caring.

Furthermore, we are going to survey the user studies by classifying them into the following categories:

- **Features:** The variables addressed in the user studies depend on the simulation purpose and the study hypotheses. For the purposes of this review, we considered the evaluation of effectiveness, usability, and immersion aspects. Effectiveness refers to the user ability to complete the tasks using the system, and their outcomes. Usability concerns to efficiency, i.e. the level of resources consumed to perform tasks, and satisfaction, i.e. the user subjective reactions while using the system. Immersion refers to engagement, i.e. the quality of user experience, a multi-dimensional construction characterized by aesthetic appeal, novelty, perceived challenge, feedback and control, attention, motivation, and affection [11].
- **Design:** There are three common experimental designs for conducting user studies: the experimental design which contains comparisons among groups randomly assigned; the quasi-experimental design which includes comparisons among non-equivalent groups, i.e. there is no random assignment; and non-experiment design, which has no comparison between groups and is usually employed for studying the effects of some intervention [12]. The studies conducted by experimental and quasi-experimental designs usually compare technological components, different participants' profiles, and different approaches used for the addressed serious purpose (e.g., a traditional method for training vs. the proposed simulation).
- **Participants:** We identified the participants' profile and crossed this information with the target public description. In particular, when the simulation does not require a specific target public and the experiment participants were students, we considered them as being part of the target public. Therefore, participants could be classified as being *target*, *non-target* or *partial* target public. Moreover, the sample size was used for helping to justify the experiments results significance.

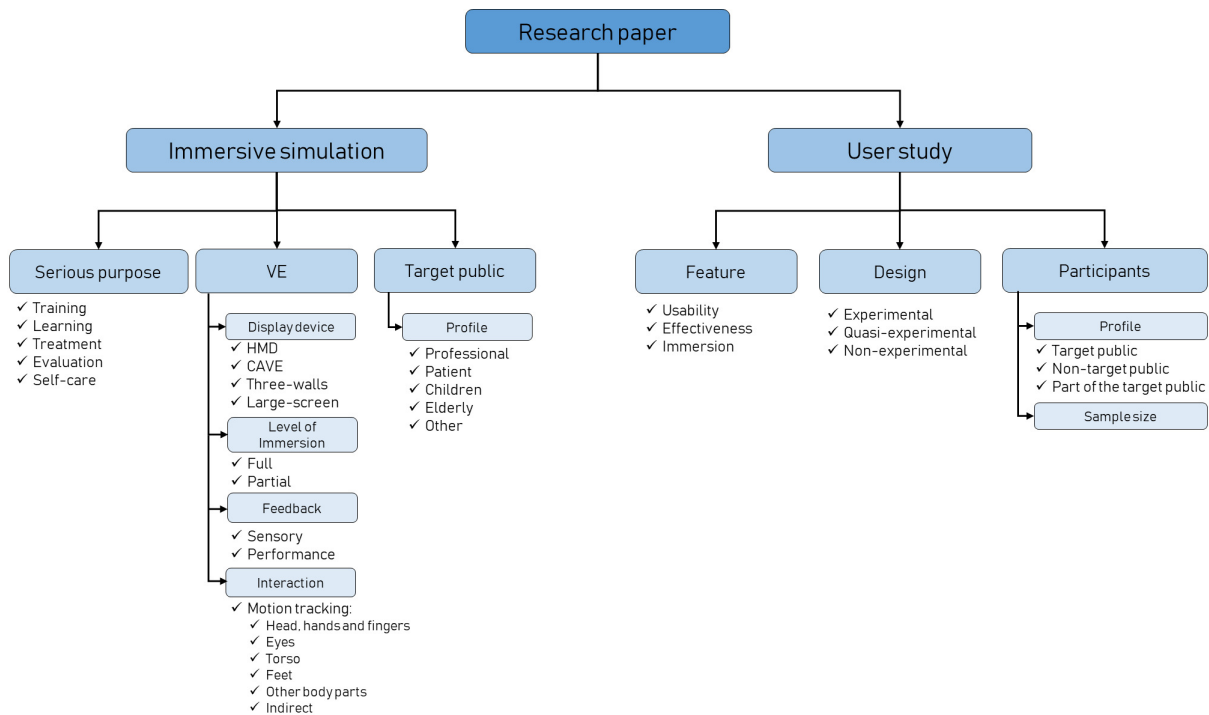


Fig. 1: Classification of the surveyed papers according to the immersive simulation and the user study characteristics.

#### IV. RESULTS

This section presents the surveyed papers with respect to the proposed classifications. Table I categorizes the simulations and user studies according to the features addressed in the taxonomy. From left to right, column (1) contains the article reference, column (2) presents the level of immersion provided by the VR system, column (3) presents the display device and column (4) informs whether it has 3D stereo. The following four columns relate to the feedback given to the user, i.e. haptic, visual, aural and performance. The next two columns describe the tracked body parts. In the column (11) the simulation purpose is specified. Column (12) classifies the target public into the categories earlier presented. Regarding the user study, column (13) presents the experimental design and column (14) the features evaluated. The final two columns introduce the participants' profile and the sample size. We based our analysis on 46 immersive “gamified” simulations, collected from 51 papers.

##### A. Immersive “gamified” simulations

From the surveyed papers, it is possible to observe a great interest in designing simulations for training, treatment, and learning (82% of the simulations). The remaining (12%) were designed for evaluation and self-care (Figure 2a). In learning and training, the greater part of VEs is fully immersive (about 80% and 60%, respectively), while fully and partially immersive systems are equally employed for treatment purposes, and self-care and evaluation systems are displayed in partially immersive VEs (Figure 2b). More than half of the systems (54%) provide stereoscopic vision. Most VEs provide full immersion (56%), and consequently, they are displayed in CAVEs and HMDs. Large-screens are largely used for treatment. For treatment and training purposes, the simulations frequently employ HMDs, while CAVEs are most applied for learning purposes. The few papers that use three-wall devices had employed them for learning, evaluation, and training.

The interaction, regardless its type, has the power of engaging users in the VE. Therefore, when it is done by motion tracking, which supports the VE to respond according to users movements, the feeling of being “there” (in the VE) is significantly increased. A great part of the surveyed simulations (41%) provides indirect body tracking using tools such as wands, treadmills, and walking platforms. Though it provides interaction through movements, the drawback is to keep the users attached to devices during the simulation. A better strategy would be to direct track the user's body, supporting free interaction, such as it is done in most simulations designed for treatment, training, and learning. These VEs often track the torso and head to get the user position and full body movements (Figure 2c). The importance of providing free interaction is accentuated in simulations designed for treatment, especially the ones used in rehabilitation because it requires users to freely move the body part to be recovered (e.g., the patient's arms).

Few VEs provide the three described sensory feedbacks, though it is more frequent the combination of only visual and auditory feedback. The haptic feedback is mainly employed in simulations for treatment, such as the ones for rehabilitation, in which the system directly tracks the body, focusing on the body parts that must be recovered. Performance feedback is widely



Fig. 2: Distribution of the immersive simulations according to its serious purpose, the VE level of immersion, the tracked users' body parts, and the experiment participants' profile.

used to help users to enhance their performance during the simulation (71% of the applications), and consequently, to improve the simulation outcomes. Certain simulations require someone to assist the user and analyze its results afterward, such as the treatment systems, in which a medical doctor receives the patient's log and analyze it to identify his/her progress. In this cases, the user does not always have a performance feedback during the simulation.

Due to the engagement and performance improvement provided by these immersive "gamified" simulations, they have been employed in various areas aiming to improve or replace the traditional approaches used for the same purpose. In health, they are applied to treat diseases such as epilepsy [s50], sclerosis [s47, s48], autism [s49], anxiety[s43], child obesity[s55], as well as for rehabilitation of stroke patients [s37, s45], and for training surgeons[s21]. They are also employed to teach safety procedures to children[s16] and to aircrafts' passengers[s13]. In industry, they are used to train industrial operators[s18, s19] and firefighters[s14]. In sports, their use intend to improve players' skills[s61, s62].

## B. User Studies

These surveyed immersive simulations were designed to be used in real life to train, treat and teach people, which makes its validation highly relevant before applying them in real scenarios. Indeed, the majority of papers (91%) considers the relevance of this factor and conduct user studies to validate the system. As mentioned earlier, there are three possible designs to carry out experiments: non-experimental, quasi-experimental, and experimental. Further, their format can be pretest-posttest or posttest only. The pretest stage includes the application of socio-demographics questionnaires and the achievement of baselines for eventual physiological measures. The posttest stage consists of obtaining the users' impressions about the simulation/experiment outcomes. Some systems could require sessions of hardware calibration and training before the intervention. A few papers report the adoption of consent forms, which gets the participants' agreement for anonymously using their data in the study.

The metrics used to measure the dependent variables could be subjective and objective. The subjective ones are frequently questionnaires and interviews that collect user impressions about the experiment and/or the system. The objective measures do not rely on the user's judgment, such as time to complete tasks, physiological data and output variables from the simulation.

Reference	Level of Immersion	Display Device	Virtual Environment Feedback						Interaction		Serious Purpose	Target Public	User Study		Participants	
			3D	Haptic	Visual	Auditory	Performance	Head Tracking	Body Tracking	Design			Feature	Sample size	Profile	
																yes
[s13]	full	HMD	yes	no	yes	yes	yes	yes	no	indirect	learning and training	others	experimental	effectiveness and immersion	48	partial
[s14]	full	CAVE	yes	no	yes	yes	yes	no	no	indirect, body movements and orientation	training and learning	professional	experimental	effectiveness, immersion and usability	31	target
[s15]	full	CAVE	no	yes	yes	no	no	no	no	indirect	training and learning	others children	experimental	effectiveness	70	target
[s16]	full	CAVE	yes	no	yes	yes	yes	no	no	indirect	training and learning	children	experimental	effectiveness and Usability	21	target
[s17]	full	CAVE	yes	no	yes	yes	yes	no	no	indirect	training and learning	professional	non-experimental	effectiveness, immersion and usability	52	target
[s018]	full	CAVE	yes	no	yes	yes	yes	no	no	indirect	training and learning	professional	experimental	effectiveness	24	non-target
[s19]	partial	large-screen three-wall	no	no	yes	yes	no	no	no	body position, hand gestures	training and learning	professional	non-experimental	effectiveness	24	non-target
[s20]	partial	large-screen three-wall	yes	no	yes	yes	no	no	no	indirect	evaluation and treatment	professional	non-experimental	effectiveness and immersion	15	target
[s21]	partial	large-screen three-wall	yes	no	yes	yes	no	no	no	indirect	evaluation and treatment	professional	non-experimental	effectiveness and immersion	15	target
[s22]	partial	large-screen three-wall	yes	no	yes	yes	no	no	no	indirect	evaluation and treatment	professional	non-experimental	effectiveness and immersion	15	target
[s23]	partial	large-screen HMD	yes	no	yes	yes	yes	no	no	arm's movements	evaluation and treatment	patient	experimental	effectiveness	6	target
[s24]	full	large-screen HMD	yes	yes	yes	yes	yes	yes	yes	finger flex, hand postures, hand, arms and upper body movements	treatment	patient	experimental	effectiveness	8	target
[s25]	partial	large-screen	no	no	yes	yes	yes	no	no	body movements	treatment	elderly	non-experimental	effectiveness and usability	10	non-target
[s26]	partial	large-screen	no	yes	yes	yes	yes	no	no	hand and body position and movements, and indirect	training	professional	non-experimental	effectiveness and usability	10	non-target
[s27]	partial	large-screen	no	yes	yes	yes	yes	no	no	body movements	treatment	elderly	non-experimental	effectiveness and usability	10	non-target
[s28]	partial	large-screen	no	yes	yes	yes	yes	no	no	body movements	treatment	elderly/patient	non-experimental	effectiveness, immersion and usability	24	partial
[s29]	partial	large-screen three-wall	no	yes	yes	yes	yes	no	no	indirect	self-care treatment	others	experimental	effectiveness, immersion and usability	5	target
[s30]	partial	large-screen three-wall	no	yes	yes	yes	yes	no	no	indirect	treatment and learning	children	non-experimental	usability	10	partial
[s31]	partial	large-screen	no	no	yes	yes	yes	yes	yes	body movements and indirect	treatment	patient	non-experimental	effectiveness, immersion and usability	6	partial
[s32]	full	HMD	yes	no	yes	yes	yes	yes	yes	hands, fingers	training	professional	non-experimental	effectiveness, immersion and usability	6	partial
[s33]	full	CAVE	yes	no	yes	yes	yes	yes	yes	body orientation	learning	children	experimental	effectiveness and Usability	45	target
[s34]	full	CAVE	yes	no	yes	yes	yes	yes	yes	steps, body position, hands motion, eyes position	learning	children	experimental	effectiveness and usability	50	target
[s35]	full	CAVE	yes	no	yes	no	no	no	no	body motion	learning	children	non-experimental	usability	12	partial
[s36]	partial	large-screen	yes	no	yes	yes	yes	yes	no	arms movements and position	training and treatment	professional	experimental	usability	15	non-target
[s37]	partial	large-screen	no	no	yes	yes	yes	yes	no	indirect	treatment	patient	experimental	effectiveness and usability	17	target
[s38]	partial	three-wall	no	no	yes	yes	yes	yes	no	indirect	learning	children	quasi-experimental	effectiveness and immersion	102	target
[s39]	partial	three-wall	no	no	yes	yes	yes	yes	no	indirect	learning	children	quasi-experimental	effectiveness and immersion	72	target
[s40]	full	HMD	yes	no	yes	yes	no	no	no	thumb and index finger	evaluation and treatment	professional	experimental	usability	21	target
[s41]	full	CAVE	yes	no	yes	yes	no	no	no	pose	treatment	patient	experimental	effectiveness and usability	20	target
[s42]	full	CAVE	yes	no	yes	yes	no	no	no	indirect	evaluation and treatment	children/patient	experimental	effectiveness	40	target
[s43]	full	HMD	yes	no	yes	yes	no	no	no	indirect	treatment	others	experimental	effectiveness and immersion	40	target
[s44]	full	HMD	yes	no	yes	yes	no	no	no	indirect	treatment	others	experimental	effectiveness and immersion	18	partial
[s45]	partial	large-screen	yes	no	yes	no	no	no	no	hands	training	professional	experimental	effectiveness, immersion and usability	21	target
[s46]	full	HMD	yes	no	yes	yes	yes	yes	yes	upper trunk joints, i.e., the shoulder, the elbow and the hand	treatment	elderly/patient	non-experimental	effectiveness and usability	15	target
[s47]	full	HMD	yes	no	yes	yes	yes	yes	yes	feet	treatment	patient	non-experimental	usability	9	target
[s48]	full	CAVE	yes	no	yes	no	no	no	no	hand gestures, body position and orientation	treatment	children/patient	experimental	effectiveness and usability	8	target
[s49]	full	CAVE	yes	no	yes	no	no	no	no	hand gestures, body position and orientation	treatment	children/patient	non-experimental	effectiveness and usability	15	target
[s50]	full	CAVE	no	no	yes	no	no	no	no	body orientation	treatment	patient	quasi-experimental	effectiveness, immersion and usability	19	non-target
[s51]	partial	large-screen	no	no	yes	no	no	no	no	arms movements and position	treatment	patient	experimental	effectiveness	59	target
[s52]	full	CAVE	no	no	yes	no	no	no	no	indirect	treatment	patient	experimental	effectiveness and immersion	38	target
[s53]	partial	large-screen	yes	no	yes	yes	yes	yes	yes	body movement and orientation	training	others	experimental	effectiveness and immersion	8	partial
[s54]	full	CAVE	yes	no	yes	yes	yes	yes	yes	forehead, wrists, and waist	training	others	non-experimental	effectiveness and immersion	30	target
[s55]	partial	large-screen	no	no	yes	yes	yes	yes	no	speech and gesture	self-care	children	experimental	effectiveness	61	target
[s56]	partial	large-screen	no	no	yes	no	no	no	no	indirect	treatment	patient	quasi-experimental	effectiveness	13	target
[s57]	partial	large-screen	no	no	yes	yes	yes	yes	no	indirect	training	professional	experimental	effectiveness	7	non-target
[s58]	full	HMD	yes	no	yes	yes	yes	yes	yes	indirect	training	professional	experimental	effectiveness	45	non-target
[s59]	full	HMD	yes	no	yes	yes	yes	yes	yes	indirect	treatment	patient	experimental	effectiveness	23	target
[s60]	full	HMD	yes	no	yes	no	no	no	no	hands, back	training	professional	experimental	effectiveness, immersion and usability	25	non-target
[s61]	full	CAVE	yes	no	yes	yes	yes	yes	yes	body movement and orientation	training	professional	experimental	effectiveness	17	non-target
[s62]	partial	large-screen	no	no	yes	yes	yes	yes	no	indirect	training	professional	experimental	effectiveness	25	non-target
[s63]	full	HMD	yes	yes	yes	no	no	yes	yes	hands and indirect	training	professional	experimental	effectiveness	17	non-target

TABLE I: Classification of the immersive “gamified” simulations toward our taxonomy.

Twenty-seven studies adopted an experimental design, from which 75% were arranged as pretest-posttest. Eleven studies were non-experimental (26%), of which nine were pretest-posttest and two were posttest only. Finally, three user studies were pretest-posttest nonequivalent quasi-experimental.

Most studies assessed effectiveness through comparisons between the simulator and a traditional approach used for the same purpose (59%). For instance, Chitarro et al. [s13] compared the simulator developed to educate passengers about safety procedures to the safety card traditionally used in aircrafts. Different levels of immersion were compared to assess the influences of the immersion factor, such as in Stach et al. [s27], in which they investigated the benefits of using haptic feedback on three exergames (i.e. games developed to encourage people to be more physically active) by using versions of the games with and without feedback. The studies that implemented a quasi-experimental design compared groups of participants with different profiles and investigated the simulation outcomes on them. In the work of Crescentini et al., this kind of comparison was implemented by recruiting a group of mindfulness-oriented meditation participants and another group formed by people not involved with meditation at all aiming to explore how they react to stressful situations [s43].

Ninety percent of the studies measures user performance to assess effectiveness, in which the expected outcome would be attached to the simulation purpose. For instance, in the simulations for learning, the dependent variable would be knowledge acquisition, while for treatment it would be the improvement of patients' health. Usability is assessed by 76% of the studies through factors such as user experience, satisfaction, workload, and feasibility. The techniques used to measure them are well-known and have been presented and discussed in the literature review on evaluation components of SGs conducted by Bellotti et al. [13].

Nearly half of the studies investigated the impacts of using immersive VR technology in the simulation outcomes by measuring engagement, physiologic arousal, sense of presence, simulator sickness, anxiety, and simulation realism.

Motion sickness can be determined through the Simulator Sickness Questionnaire (SSQ) [14], which measures nausea, disorientation, and oculomotor distress. The subjects' tolerance to the VE could be measured by the Subjective Units of Discomfort (SUD) questionnaire [15] combined with vital signs, such as blood pressure and pulse rate.

The sense of immersion can be measured through flow, cognitive absorption, and presence, which is addressed in the questionnaire presented by Jennet et al. [16]. The sense of presence, which refers to the psychological sense of being in the VE [1], could also be measured by the SUS Questionnaire of Slater, Usoh, and Steed [17], the Presence Questionnaire (PQ) [18] and the Immersive Tendencies Questionnaire (ITQ) [18], which evaluates the tendency of subjects to be involved (or immersed) with the VE.

The assessment of the interaction between players serves to validate the multi-player approach in simulations supporting collaborative environments and can be addressed by the Social Presence in Gaming Questionnaire (SPGQ) [19], which measures aspects of co-presence, psychological involvement, behavioral interdependence, enjoyment of social context, connectivity, empathy, sympathy, admiration, jealousy, and revenge. This should prove users' awareness and involvement with their co-players.

In addition to these metrics, the participants are an important element of the assessment. As discussed before, each simulation requires a specific public depending on its serious purpose and application area. Then, it would be appropriate to recruit this public to take part in the experiment. However, conducting studies with real users, particularly if it involves an uncontrolled environment, is challenging, because it is difficult to ensure the control of conditions, and the data gathering is not always available, neither are real users [20]. Most surveyed simulations employ large VR setups, which constrains the experiment to a controlled environment. However, the assessment of immersive environments, even the controlled ones, still faces safety and ethical issues.

Frequently participants are University students because of the difficulty to recruit volunteers fitting the target public profile, or because the simulation does not require a specific public. In this case, volunteers would belong to the target public. Though their profile would be homogeneous, which is good for the experiment, it is normally different from people outside the University environment, which might include bias. Since a great part of the surveyed simulations is designed for treatment, which requires a public carrying the illness it is treating, the majority of the studies were successful in recruiting people matching the target public profile (62%). Indeed, the user studies assessing simulations for treatment and learning frequently involved the target public in the experiment. About 20% of the simulations for treatment were evaluated using a group consisting of a non-target public (Figure 2d). These are the quasi-experimental designs, which investigated the simulation outcomes over groups with different profiles. These results are discussed in the following section. In general, 13% recruited part of the public and 25% of experiments' volunteers did not match the target public, which most simulations were designed for training.

## V. DISCUSSION

The use of immersive "gamified" simulations for serious purposes have been widespread within areas as diverse as aviation, firefighting, health, and education. In addition to their interaction, motivation, and engagement, the use of simulators supports a safer and cost-effective environment for training, learning, treatment, evaluation, and self-care.

These simulations can provide similar gain of knowledge to the traditional learning methods, with the advantage of being more effective for knowledge retention [s13]. For assessment and training of cognitive functions, the task performance is



significantly correlated to participants' figural-spatial memory abilities and subjective level of immersion in the VE [s50]. Immersive VEs help to reach a situational awareness and process understanding better than the ones achieved by conventional training methods[s18].

On clinical change, there is a positive trend of decreasing pain intensity and disability scores, and increasing walking distance after a minimum training period of 4 weeks [s31]. In Parkinson's disease rehabilitation, the simulation increased the stimulus and motivation for exercising, and despite being intensive at times, participants were motivated to complete the game and improve their score at each level [s28]. Immersive rehabilitation simulations could also contribute to patients' upper limb motor function improvement [s37]. In vestibular rehabilitation, the simulation produced equivalent functional outcomes than the clinically accepted physical therapy [s52]. Simulations are also effective for increasing mindfulness and decreasing trait and anxiety levels, providing psychological well-being [s43]. In firefighting, a simulation designed to develop systematic and thorough search behavior in the presence of physical tension and other stress factors provided successful outcomes for the exposition to physical stress, considering that all players mentioned the physical load in some way [s14].

Lorenzo et al. [s41] evaluated a VE designed to train, develop, and improve deficits in emotional skills shown in students with autism spectrum disorder. They find the children emotional behavior improved in the real school environment, which indicates that the simulation outcomes can be transferred to the real world. Collaborative VEs presented benefits for interpersonal coordination, which is also improved when the player receives performance feedback [s62].

As indicated by these studies, immersive VEs can be effective while being engaging and motivating. In some cases, they are even capable to replace traditional methods. Nevertheless, some studies point to the benefits of using it combined with the regular approaches. As part of the training experience, they can lead to great learning gains over time for young children [s16]. For unfamiliar tasks, the simulation could use guidance by a qualified trainer [s19]. In rehabilitation, Ma et al. showed the use of the simulation combined with functional training could be more effective than using functional training solely [s24]. Simulations can be also an effective supplementary intervention in stroke rehabilitation for functional recovery, and visual feedback motivates participants and induces cortical reorganization [s45]. A learning environment that combines guided activity with an enhanced prompting mechanism on behalf of the system may be more effective in fostering a reflective process that can lead to conceptual change [s34].

Haptic feedback showed to be a useful tool for balancing exergames for people of disparate physical abilities while increasing presence and realism [s27]. Ragan et al. [s58] demonstrated visual fidelity does not have a significant effect on assessing target detection or strategy usage, but training systems for visual scanning and similar tasks should, when possible, use a level of visual complexity that is as close to the real environment as possible in order to ensure good transfer. Vision-based tracking systems support natural movement in space, as well as intuitive grasping and manipulation of objects. In training, using these systems can produce positive effects on participants own sense of confidence and knowledge on the reactions of the supervisors and the environment [s44]. Performance feedback presented through scores stimulates users to achieve better, improving their skills in the learning process [s53].

In general, the studies reported that players enjoyed using the simulation. It could be a consequence of either using immersion or adopting game aspects, which keeps the user motivated to reach the goals and increase its score. For an aircraft experience, the game showed to be more engaging and more fearful than the safety card reading, and it produced consistent and higher emotional arousal. This experience also showed the repeated exposure to virtual experiences of aircraft emergencies combined with performing the right actions to survive in crash situations could make passengers less fearful of flying in general, as well as less anxious and stressed in the event of an emergency during their real flights. Backlund et al. [s15] demonstrated that between two simulations designed for the same purpose, the participants performed better while playing the "gamified" version, which also enhanced the learning objectives.

This enjoyment and preference for the "gamified" simulation could be a consequence of previous gaming experience, seeing that the game aspects provide a similar path to reach the simulation objective, i.e. following the rules increases score. To exemplify it, the work of Ordaz et al. [s26] showed the gaming experience influenced positively on training completion time. However, participants made a considerable amount of mistakes, which indicates that gaming experience does not influence the understanding of manufacturing operation (the area where the simulation was employed). Though the completion time could take benefit from the use of simulation, its serious purpose cannot be neglected. Likewise, the acceptability of such immersive simulations tends to be higher for people experienced in video games and VR than for those without such previous experience [s36].

Most surveyed simulations are displayed in HMDs and CAVEs. It restricts their use to people who have access to such (not yet widespread) technology, and it might have side-effects, such as motion sickness. A few studies reported that participants felt nausea or another kind of discomfort after performing the simulation. The intervention time was acceptable varying between 15 to 90 minutes – in one single case people were exposed to the immersive VE up to 3 hours. Some studies reported that a few participants could not finish the experiment due to motion sickness [s15, s16, s17, s22, s35, s39, s49, s50, s52, s58]. Though most of them claim the sickness did not significantly affect the results and that after a number of interventions people got used to the system, reducing the symptoms, this remains a great limitation in the field.

Moreover, most studies validated the efficacy of the simulation without taking into account the practicality and usability of immersive VR technologies. The comparisons were mainly between the simulation and the traditional methods, verifying its

effectiveness. These validation designs could add a bias in the results due to technological issues, so a better approach would be to compare the simulation in different immersive setups varying aspects such as haptic, aural and visual feedback, FOV, FOR, and level of realism. This way, it is possible to explore lower levels of immersion that might be engaging to the user without producing much discomfort. While a desktop version, for instance, might be less interesting than the immersive ones, it would have the advantage of being accessible to a much larger user population, while being more comfortable. However, the field should be able to benefit from the advantages introduced by technological innovations, thus the relevance of designing effective simulations that benefit from immersion without causing discomfort and being accessible to a larger public.

The sample selected for the user studies is an important component because its profile could influence the results. As discussed earlier, in cases which the participants were experienced in video games the results were positive for completion time and for user preference. However, the same sample presented negative outcomes regarding the simulation objective. Schwebel et al. [s38] evaluated a simulation for understanding and preventing children’s injuries. The experiment had compared the performance of a group of children against a group of adults, in which the latter one presented better performance. There were typical differences: adults had safer behavior for nearly all measures of interest, while the 7-year-old children had significantly more hits than adults; they also waited longer to cross the street. Grewe et al. [s50] present and evaluate a simulation for cognitive neurorehabilitation. They compared a sample of healthy people against a small sample of patients with focal epilepsy, in which the healthy group executed the tasks better and faster than the target public. Killane et al. [s56] conducted a user study to validate a simulation for Freezing of Gait (FOG), comparing a group of people with FOG against a non-FOG group of people, which the first group outperforming the second one. These findings reveal relevant differences in performance between groups, which implies that evaluating the simulation using the data collected by the non-target public solely could mislead the conclusions about its effectiveness. Moreover, the sample size in these experiments could be considered small ( $M=28.7$  participants,  $SD=23.72$ ) taking into account the purpose of these simulations (e.g., to train surgeons, treat diseases, learn content to children, and so on), and consequently the importance of having a significant validation.

## VI. CONCLUSION

The majority of these simulations were developed for training, treatment and learning, followed by evaluation and self-care. Although the VR technology still might be limited by introducing motion sickness, using it in “gamified” simulations presented great benefits for knowledge gain and retention, firefighting training, and clinical changes in physical rehabilitation and anxiety. User studies were conducted by almost every reviewed work, following a well-known protocol. Most of them evaluated effectiveness by comparing the VE with a traditional approach in the specific application area. Moreover, the subjects can influence the experiment results by recruiting people with previous gaming experience or people not fitting the target public profile.

We pointed to the influences of using immersive VR technology combined with game aspects on simulations for serious purposes. However, our analysis is limited by the description of user studies and results provided in the surveyed papers, which sometimes lack in technical details, such as the ones related to sensory feedback support. Although they provide enough details about user studies to allow the reader to extract subjects, methods and design information, as well as an adequate presentation of their findings, some of them do not provide every result, such as when they mention the use of a certain questionnaire without presenting its outcomes. In a future work, the sample of papers could be expanded to include immersive “gamified” simulations with other serious purposes besides the ones we covered, and to update the employed VR technology as it advances. Moreover, it could include the frameworks and educational approaches used to give a serious game perspective to these simulations.

## VII. ACKNOWLEDGMENT

We gratefully acknowledge the partial financial support from CNPq through grant 449555/2014-3 and grant 131603/2015-9, as well as NVIDIA Corporation for hardware donation.

## REFERENCES

- [1] M. Slater, “A note on presence terminology,” *Presence connect*, vol. 3, no. 3, pp. 1–5, 2003.
- [2] M. A. Muhanna, “Virtual reality and the cave: Taxonomy, interaction challenges and research directions,” *Journal of King Saud University-Computer and Information Sciences*, vol. 27, no. 3, pp. 344–361, 2015.
- [3] D. Djaouti, J. Alvarez, J.-P. Jessel, and O. Rampnoux, “Origins of Serious Games,” *Serious Games and Edutainment Applications*, pp. 25–43, 2011.
- [4] S. Deterding, D. Dixon, R. Khaled, and L. Nacke, “From game design elements to gamefulness: defining gamification,” in *Proceedings of the 15th international academic MindTrek conference: Envisioning future media environments*, pp. 9–15, ACM, 2011.
- [5] E. D. Ragan, R. Kopper, P. Schuchardt, and D. A. Bowman, “Studying the effects of stereo, head tracking, and field of regard on a small-scale spatial judgment task,” *Visualization and Computer Graphics, IEEE Transactions on*, vol. 19, no. 5, pp. 886–896, 2013.

- [6] E. D. Ragan, A. Sowndararajan, R. Kopper, and D. A. Bowman, "The effects of higher levels of immersion on procedure memorization performance and implications for educational virtual environments," *Presence: Teleoperators and Virtual Environments*, vol. 19, no. 6, pp. 527–543, 2010.
- [7] A. Sowndararajan, R. Wang, and D. A. Bowman, "Quantifying the benefits of immersion for procedural training," in *Proceedings of the 2008 workshop on Immersive projection technologies/Emerging display technologies*, p. 2, ACM, 2008.
- [8] D. Bowman, A. Sowndararajan, E. Ragan, and R. Kopper, "Higher levels of immersion improve procedure memorization performance," in *Proceedings of the 15th Joint virtual reality Eurographics conference on Virtual Environments*, pp. 121–128, Eurographics Association, 2009.
- [9] A. Chalmers and K. Debattista, "Level of realism for serious games," in *Games and Virtual Worlds for Serious Applications, 2009. VS-GAMES'09. Conference in*, pp. 225–232, IEEE, 2009.
- [10] A. Mitchell and C. Savill-Smith, "The use of computer and video games for learning: A review of the literature," 2004.
- [11] H. L. O'Brien, E. G. Toms, E. K. Kelloway, and E. Kelley, "Developing and evaluating a reliable measure of user engagement," *Proceedings of the American Society for Information Science and Technology*, vol. 45, no. 1, pp. 1–10, 2008.
- [12] W. Trochim, J. P. Donnelly, and K. Arora, *Research methods: The essential knowledge base*. Nelson Education, 2015.
- [13] F. Bellotti, B. Kapralos, K. Lee, P. Moreno-Ger, and R. Berta, "Assessment in and of serious games: an overview," *Advances in Human-Computer Interaction*, vol. 2013, p. 1, 2013.
- [14] R. S. Kennedy, N. E. Lane, K. S. Berbaum, and M. G. Lilienthal, "Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness," *The international journal of aviation psychology*, vol. 3, no. 3, pp. 203–220, 1993.
- [15] D. M. Kaplan, T. Smith, and J. Coons, "A validity study of the subjective unit of discomfort (sud) score.," *Measurement and Evaluation in Counseling and Development*, 1995.
- [16] T. Tijds and A. Waltond, "Measuring and defining the experience of immersion in games," *Int. J. Human-Computer Studies*, vol. 66, pp. 641–661, 2008.
- [17] M. Slater, J. McCarthy, and F. Maringelli, "The influence of body movement on subjective presence in virtual environments," *Human Factors: The Journal of the Human Factors and Ergonomics Society*, vol. 40, no. 3, pp. 469–477, 1998.
- [18] B. G. Witmer and M. J. Singer, "Measuring presence in virtual environments: A presence questionnaire," *Presence: Teleoperators and virtual environments*, vol. 7, no. 3, pp. 225–240, 1998.
- [19] Y. A. De Kort, W. A. IJsselsteijn, and K. Poels, "Digital games as social presence technology: Development of the social presence in gaming questionnaire (spgq)," *Proceedings of PRESENCE*, vol. 195203, 2007.
- [20] A. Steed, S. Frlston, M. M. Lopez, J. Drummond, Y. Pan, and D. Swapp, "An in the wild experiment on presence and embodiment using consumer virtual reality equipment," *IEEE transactions on visualization and computer graphics*, vol. 22, no. 4, pp. 1406–1414, 2016.