



# Virtual team training with Mixed Reality and Virtual Reality – benefits and limitations illustrated on the example of two paramedic classes

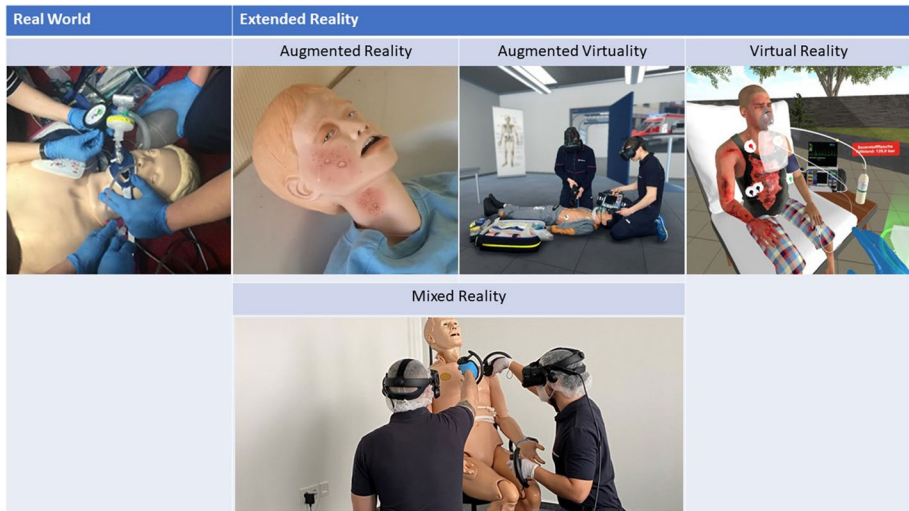
Christian Elsenbast<sup>1</sup> · Philipp Dahlmann<sup>2</sup> · Dominik Schnier<sup>1</sup>

Received: 26 June 2023 / Revised: 21 November 2023 / Accepted: 13 December 2023 /  
Published online: 11 January 2024  
© The Author(s) 2024

## Abstract

In the field of medical case simulations, strong technologization has been observed for years, for example through Extended Reality. This study examined the impact of Virtual and Mixed Reality (VR, resp. MR) on the team training of paramedic trainees. In the quasi-experimental cross-sectional controlled comparison study, participants were assigned to an experimental group (=EG, mixed reality with manikin) and a comparison group (=CG, virtual reality without manikin). After a virtual case simulation, “VR sickness”, sense of presence, motivation, and sociodemographic variables were elicited. Parametric tests were used for mean comparison and correlation analysis. A total of 20 participants were evaluated in the EG and 16 in the CG. In both groups, high intrinsic motivation (mean 5.32/7), an acceptable sense of spatial presence (mean 4.24/6), and advanced usability (mean 68.54/100), were observed. VR sickness symptoms were present (SSQ total score: 11.43). In both groups, correlation was found between Identified Regulation and Usability (EG:  $r = .74$ ,  $p < .001$ , CG:  $r = .76$ ,  $p < .001$ ). The same pattern could be observed for Intrinsic Motivation and Usability (EG:  $r = .83$ ,  $p < .001$ , CG:  $r = .70$ ,  $p = .003$ ). Spatial Presence and Usability were also correlated in both groups (EG:  $r = .71$ ,  $p < .001$ , CG:  $r = .61$ ,  $p = .01$ ). A significant mean score difference ( $p = .021$ ) was found for Amotivation between the EG (Mean = 1.92/6) and the CG (Mean = 1.56/6). VR and MR are potential learning methods if they are implemented in a planned manner by competent faculty. Our recommendation is to use setup checklists, but also a competence-oriented approach, considering fiction contract and structured debriefing. The use of wireless head-mounted displays and “data gloves” is also recommended.

**Keywords** Virtual Reality · Augmented Reality · Mixed Reality · Extended Reality · Team training · Team resource management · Interprofessional training · Paramedic education · Virtual learning environment · Digital education



**Fig. 1** Extended Reality and Mixed Reality continuum

## 1 Background

### 1.1 Overview extended reality

Case simulations as an imitation of clinical experience are accepted educational resources for training technical skills [1] and non-technical skills such as teamwork, communication, and situation awareness [2, 3]. Case simulations can take place in conventional settings (Fig. 1, left picture, “Real World”), but also using highly immersive technologies such as Virtual Reality (VR).

VR has been researched for many decades and has also gained popularity in medical education as well as, to some extent, in Emergency Medical Services (EMS) [4]. VR plays an important role in the professional training of healthcare professionals and is comparable to traditional media in learning gains [5]. Let us start by distinguishing the terms Extended Reality (XR), Virtual Reality (VR), Augmented Reality (AR), Augmented Virtuality (AV), and Mixed Reality (MR). XR is an umbrella term for technologies that create computer-generated environments or objects. VR describes a completely computer-generated world as diametrically opposed to the real environment [6]. AR means that the user still perceives the physical reality, but it is additionally enriched or extended with digital information. MR combines real and virtual objects. AV consists predominantly of virtual spaces that incorporate real physical objects and people that can interact with the virtual world in real time. Figure 1 shows the relationship between these technologies. The left image in Fig. 1 shows a case simulation for EMS without virtual elements. On the right, the virtual spectrum extends. On the far right, a complete treatment situation is virtualized (VR). The two middle images are variants of MR. In the center left image, a skin phenomenon (e.g., an allergic reaction) is projected onto the real existing manikin. In the center right image, the environment of the physically present trainees is virtualized with physically existing objects.

This article focuses on the results of the ViTAWiN project [7]. ViTAWiN is an acronym for the German equivalent of “Virtually augmented training for education and training in

interprofessional emergency care" (German: "Virtuell-augmentiertes Training für die Aus- und Weiterbildung in der interprofessionellen Notfallversorgung"). The project developed a multi-user Virtual Learning Environment (VLE) to enable interprofessional training of paramedics and emergency nurses in a virtualized, highly immersive scenario, independent of location. This means that paramedics, emergency nurses, and physicians can train at the same time in the same case scenario, but from different locations [8]. At the time of the study, ViTAWiN was the only research project that addressed both structured diagnostic and treatment processes and interprofessional aspects in the context of EMS and emergency care. An interprofessional aspect is, for example, the cooperation between paramedics and emergency care in the shock room. This is a special treatment unit for severely injured or seriously ill people. The interface between EMS and hospital is characterized by complexity resulting from the large number of decision variables, unknown information, the large treatment team from various disciplines with different mental models of care, and the imbalance between the time needed to structure the situation and the dynamic course of the illness or injury. The ViTAWiN project, which was based on the experiences of the predecessor project EPICSAVE [9, 10], was launched to enable simulating structured courses of action on the one hand and, on the other hand, to raise awareness of this critical phase of patient transfer among students of emergency medical services and emergency care during their school education. The common human resources of the treatment team are important aspects of the treatment phase. "Virtual team training" in this study also addressed the team dynamics within the emergency medical services and emergency care teams, as well as the joint behavior and targeted synergetic actions of the two ad-hoc teams. These aspects are explored in the presentation of the relevant literature. In addition to the team training component, the patient's manikin was integrated into the VR environment as a haptic element to enable examination steps such as palpating the pulse. A higher degree of realism was expected. During the course of the project, aspects of facial expressions and gestures were implemented on the virtual patient to visualize pain expression and fear, in particular (Fig. 2).



**Fig. 2** Virtualized settings (Source: ViTAWiN project)

In the project, an evaluation concept was developed first, which included an evaluation at the different milestones. In this article, we report the results of the final evaluation.

## 1.2 Review of relevant research

This narrative review focuses primarily on VR and MR in EMS and paramedic education.

In their "Trauma bay VR study", Colonna et al. [11] describe the learning successes of their convenience sample in clinical decision-making using critical decision points, i.e., intubation, cricothyroidotomy, chest tube, and intravenous access. Chabaane et al. [12] showed that a single VR application can increase competence perception and significantly reduce anxiety in paramedics. How long this effect lasts is unknown. Effect sizes were small (anxiety reduction:  $d=0.33$ , increase in competence perception:  $d=0.35$ ). When training triage processes during mass casualty incidents, Mills et al. [13] showed that VR simulation was nearly identical in simulation efficiency. They concluded that there were no measurable differences in the mental demand, temporal demand, performance, effort, or frustration domains between the VR group and control groups. Equal levels of satisfaction and triage results could be measured in both groups. Harrington et al. [14] used the example of participants in an Advanced Trauma Life Support course to show that participants enjoyed using VR, rated it as a platform of choice, and considered VR to be a cost-effective tool. Regarding the intensity of cognitive-emotional stress, it can be assumed that case simulations in VR can reach high intensity [15]. Birtill et al. [4] describe three VR studies in their scoping review (up to February 2020) and describe the technology as promising in terms of teaching methods.

With regard to the use of tactile elements, the first promising approaches already exist in the field of emergency medicine, as Schild et al. point out [8]. Girao et al. [16] also report on the implementation of a manikin in first aid training and document stable integration of the manikin into the VLE and positive feedback from the user experience. Furthermore, Melo et al. [17] show in their systematic review that haptic integration can have a positive impact on user performance and enhance the sense of actually being "there" (sense of presence) in the virtualized environment. Melo et al. also note that performance studies were found to be significantly more frequent with haptic integration as opposed to user perception. They further note that the existing literature has a distinctly technical focus and underemphasizes human-centered aspects. Dinh et al. [18] describe the strong positive correlation between multisensory stimuli and presence, while Biocca specifically highlights the potential of haptics to improve presence [19].

The reason for focusing the project and this study on patient transfer and trauma care is also supported by the literature regarding the chosen scenario. Interprofessional handovers in ad-hoc teams pose high risks due to the lack of clear agreements, different mental models of the treatment situation, and/or difficult team briefings, making the process less straightforward [16–18]. Boosting team performance in dynamic settings requires a mix of strategies involving individual readiness, teamwork, and environmental optimization [19].

Apart from ViTAWiN [7], no other XR simulations were found at the time of the study that addressed patient handover in interprofessional teams. Furthermore, there are no known XR applications that specifically address the shock room phase. The combination of EMS treatment and patient transfer with tactile elements was also new at the time of the study, although multisensory input, e.g., through tactile augmentation, seems promising.

### 1.3 Objectives

**Aim:** As there is still a lack of systematic data, especially in the area of haptic augmented training, team training, and location-independent settings, our study aims to contribute to the evaluation of MR and VR in the context of paramedic education.

**Objectives:** In our study, we analyze the media use factors associated with the use of a highly immersive VLE for paramedic students. In order to extrapolate the relevance for highly immersive training in paramedic education, we analyzed the differences that occur in terms of media perception and motivation when a haptic tactile patient manikin is incorporated into a highly immersive simulation. The question of the usefulness of haptic augmentation is important because multidimensional approaches have been shown to enhance students' perception of realism [4, 18, 19].

**RQ 1:** To what extent do situational motivation, VR sickness, presence, and usability impact paramedic trainees after a VR and MR training sequence as measured by a questionnaire survey?

**RQ 2:** How do paramedic trainees in a VR training sequence differ from a group of paramedic trainees using an integrated haptic, palpable manikin in an MR setting in terms of situational motivation, VR sickness, presence, and usability?

## 2 Methods

### 2.1 Research design and organizational background

The cross-sectional quasi-experimental study with controlled comparison is based on the examination of a VR and MR training sequence in paramedic trainees, with a subsequent survey on media effect factors, motivation, and experience of the immersive environment conducted via an online questionnaire.

### 2.2 Data collection & eligibility criteria

The data was collected from the two educational EMS project partners in the ViTAWiN project. The two educational partners are part of the vocational training scene for EMS in Germany. To be included in the study, participants had to be trainees in the German emergency medical services profession and be physically and mentally healthy according to their own subjective assessment. In addition, the requirements of the Covid-19 regulations at the time had to be met in order to identify asymptomatic patients. In addition to the absence of symptoms, this also included a negative rapid Covid-19 test. These criteria applied to both the control group and the experimental group.

### 2.3 Inclusion and exclusion criteria

The participants were interviewed about their health status on the day of the simulation. Only participants who did not report being acutely ill or in any way acutely impaired

were included. Participants were given verbal and written information about the study and had the option of not participating or dropping out, as well as withdrawing consent.

## 2.4 Participants' characteristics

The participants were students in the German three-year paramedic training program. Unfortunately, due to staffing difficulties during the pandemic, we were not able to recruit emergency nurses from our educational partner. Prior to MR/VR exposure, experienced instructors gave the participants the opportunity to familiarize themselves with the hardware and software as well as the elements of the virtual environment. The participants had little or no prior experience with VR/MR.

## 2.5 Technical setup & hygiene measures

All simulations were performed in a play area setup, allowing the participants to roam freely. During the entire simulation, the participants wore hair nets to improve hygienic conditions regarding the head-mounted displays (HMD). In addition, disposable gloves and FFP2 masks (comparable to N95 and KN95) were used due to the Covid-19 pandemic.

While a 3 × 4 m play area was used, the size can be adapted to local conditions. In this study, we used the VR set Valve Index® (Valve Cooperation, Bellevue, Washington/USA). The position of the HMD and the controller is detected by laser sensors. This position information is transmitted to and processed by a computer, which reflects the updated VLE back to the HMD. In the classroom, the VLE can be additionally displayed on a screen from different perspectives so that other learners can observe the virtualized patient care.

## 2.6 Experimental manipulation

In the comparison group (which used VR; hereinafter referred to as VR Group), prehospital patient care was completed without a manikin (Fig. 3); in the experimental group (which used MR; hereinafter referred to as MR Group), a patient manikin was integrated as a haptic element (Fig. 4). In the MR Group, the patient manikin was integrated



Fig. 3 VR setting without manikin

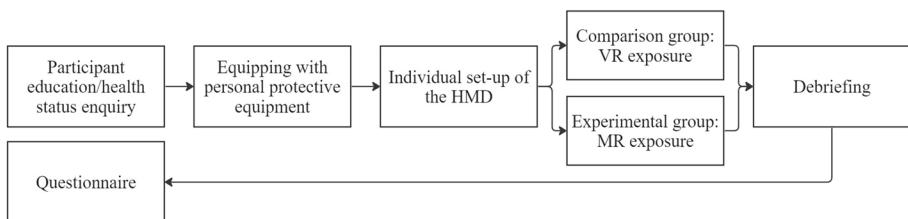


**Fig. 4** MR setting with manikin

into the virtualized environment by a simulation technician. A simulation technician was also available for the VR Group. As a consequence of the experimental design, the participants of the MR Group could "feel" the patient on a rudimentary level. This enabled examination steps such as palpating the pulse. The VR Group did not have this possibility and could only rely on the optically virtualized patient. The lack of a haptic reference could lead to less realistic perception as, for example, when auscultating the lungs, there is no corresponding resistance on the patient’s body surface. Other important examination steps, such as palpation of the pulse, can be simulated by haptic feedback. Both groups received an individual debriefing in the form of an after-action review by an educational professional from the education partners. This involved discussing the objectives and achievement of the medical treatment with the participants.

Figure 5 shows the main steps of the study implementation on site.

The participants had to assess the scene in terms of safety and had to provide an initial assessment and structured care to a burn patient in the MR/VR environment using medical guidelines and regional protocols. The initial assessment referred to may indicate the need for some type of action. For example, an assessment of the respiratory tract may reveal that a foreign body has been inhaled and needs to be removed. As an additional resource, we provided a detailed description of the assessment and treatment steps. This also served as a basis for the debriefing.



**Fig. 5** Main steps of the study implementation on site

## 2.7 Assignment method

Assignment to an experimental condition was made in a non-randomized manner on a class-by-class basis due to the extensive technical setting. Assignment at the individual level was not feasible from a research economics perspective as it would have involved a lot more time and personnel. The reasons for the increased resource requirements were the technical setup, the room arrangement, and especially the still very complex integration and calibration of the tracking of the manikin.

## 2.8 Instrumentation

After the VR/MR exposure, the participants were asked to complete an online questionnaire. This comprised the collection of sociodemographic data, the Situational Motivation Scale – SIMS [20], the System Usability Scale – SUS [21, 22], and the IGroup Presence Questionnaire – IPQ [23] (Table 1). The Simulator Sickness Questionnaire – SSQ [24] was completed after the evaluation. Approximately three to five minutes elapsed between the end of the exposure and the start of the questionnaire. The instruments used are described below.

Motivation and engagement are predictors of academic performance and may indicate planning problems (e.g., integrating the new learning method just before an exam) [25]. Based on the self-determination theory [26], the SIMS measures Intrinsic Motivation, Identified Regulation, External Regulation, and Amotivation. The higher the score, the higher the level. The instrument was tested with moderate to excellent reliability

**Table 1** Constructs, scale, and sample items

Construct	Scale (weight*)	Sample Item
<b>SIMS</b>		
Intrinsic Motivation	1–7	Because I think that this activity is interesting
Identified Regulation	1–7	Because I am doing it for my own good
External Regulation	1–7	Because I am supposed to do it
Amotivation	1–7	I do this activity, but I am not sure if it is worth it
<b>SSQ</b>		
Nausea**	0–3 (9.45)	Nausea
Disorientation**	0–3 (7.58)	Vertigo
Oculomotor**	0–3 (13.92)	Blurred vision
Total Score***		
<b>IPQ</b>		
Involvement	0–6	I was completely captivated by the virtual world
Realism	0–6	How real did the virtual world seem to you?
Spatial Presence	0–6	I did not feel present in the virtual space
General Factor	0–6	In the computer-generated world I had a sense of "being there"
<b>SUS</b>	1–5 (2.5)	I felt very confident using the system

\*if appropriate. \*\* The sum of the symptom scores is weighted in each case. \*\*\* Total Score =  $([1] + [2] + [3]) * 3.74$



(Cronbach's alpha: intrinsic motivation=0.95, identified regulation=0.85, external regulation=0.62, amotivation=0.83) and factor validity. The results of the confirmatory factor analysis showed  $\chi^2$  to be significant ( $\chi^2(98, n=907)=856.50, p<0.05$ ), and NNFI (0.89) to be somewhat lower than the 0.90 cut-off value. However, CFI (0.90) was shown to be satisfactory. All hypothesized factor loadings, covariances, error residuals, and factor residuals were found to be significant [20].

Presence, i.e., the feeling of actually being in the virtual environment, as well as the illusion of location and plausibility are, in our view, key quality criteria for the user experience of an XR application. The measurement of presence, i.e., the user's feeling of being directly in the virtualized action, will be discussed in more detail in the Methods section. However, we want to establish the connection to the technical perspective here and thus state that HMDs have an outstanding reputation with regard to the creation of a sense of presence, i.e., they have high immersive potential compared to other virtualization methods [27]. In this respect, Salomoni et al. [27] point out the high importance of the graphical user interface and the way in which the user interface can be operated by the users, which also has an effect on the expression of the sense of presence.

The IPQ addresses questions representing Involvement, Spatial Presence, Realism, and a General Factor, with responses using seven-point rating scales (0–6) and high scores being positive for each construct. At this point, we quote Schubert et al. because there seems to be some misunderstandings about scaling: "All items have a range from 0 to 6. The left endpoint of the scale is always 0, the right endpoint is always 6" [28]. The IPQ was tested with good reliability (Cronbach's alpha > 0.7 for all subscales) and factor validity. For details, see Schubert et al. [23].

In highly immersive HMD-based XR, user experience can be affected by side effects, which are called simulator sickness. Symptoms include headaches, impaired vision, and dizziness. There are objective and subjective methods for assessing the severity of simulator sickness, with the most popular being Kennedy's SSQ [29]. The SSQ includes three constructs – Disorientation, Nausea, and Oculomotor – which are answered using a four-point rating scale. Each construct as well as the total score has its own weighting (Table 3). Weighted scores < 5 are considered "negligible", scores of 5–10 are considered "minimal", scores of 10–15 are considered "significant", and scores of 15–20 are considered "of concern". A total score above 20 is considered "bad" [30]. The use of the SSQ is critically debated in the application domain of VR due to its development, which was aimed at flight simulators rather than the latest generation of HMD-VR [30, 31]. Nevertheless, because SSQ has been used in numerous studies for reporting VR-related adverse events [29, 32], it was used here to allow comparability. Symptoms of simulator sickness can last from minutes to hours and increase with the duration of VR exposure, although no general pattern is currently discernible. Studies in VR settings have shown that the probability of symptoms is heterogeneous [32, 33].

The usability of a virtual learning system should be as high as possible to allow participants to focus on the learning objectives. A system with poor usability runs the risk of participants losing engagement and using cognitive resources to operate the system that are actually needed to master the learning objectives. The SUS is a simple and technology-independent questionnaire to evaluate the usability of a system. It comprises ten questions, which are answered on a five-point rating scale and evaluated as a percentage on the grade scale. Rule of thumb: "[...] products that scored in the 90 s were exceptional, products that scored in the 80 s were good, and products that scored in the 70 s were acceptable. Anything below a 70 had usability issues that were cause for concern." [22]. The SUS was

tested with excellent reliability (Cronbach's  $\alpha=0.91$ ) and acceptable factorial validity [22, 34].

## 2.9 Measures and covariates, power and precision

All statistical operations were performed using RStudio (Version 2022.02.1 + 461). First, the data was prepared. This was followed by a pattern analysis of missing values and a descriptive-statistical evaluation.

Subgroups were compared using Welch's *t*-Test ( $\alpha=0.05$ , two-sided) after ruling out extreme deviation from the normal distribution (R package "stats"). While the pooled *t*-test according to Student is based on a paired calculation of standard errors, the *t*-test with Welch estimation applied here is based on an unpooled calculation of standard errors, which does not require heteroskedacity, but is at least as reliable as pooled calculation when presented in terms of Type I error [35–37]. The power of the Welch's *t*-test is similar to that of the Student's *t*-test even when the population variances are equal. With regard to unequal sample sizes, we note the robustness of the Welch approximation compared to the Student's *t*-test, especially in the absence of gross violations of normal distribution [38].

Correlations were tested with Pearson correlation with Holm correction (R-package "rstatix"). The sample size of the two subgroups resulted in a beta error probability of 0.43 for the *t*-Test for two independent means (effect size  $d=0.5$ ) and 0.49 for the calculation of the correlation coefficient by Pearson (effect size  $q=0.6$ ) for independent samples ("g\*power", Version 3.1). Considering that the size of the two subgroups is not extremely unbalanced, it can be assumed that the calculation of the Welch's *t*-Tests and the Pearson correlation is robust for the present sample [39].

The three genders were compared with the Tukey–Kramer-Test ( $\alpha=0.05$ ) under adjustments for multiple comparisons ("stats" package). This test uses the harmonic mean of sample numbers and can therefore be used when sample numbers in the groups being compared are different [40].

To test interrelatedness, Cronbach's  $\alpha$  and McDonald's  $\omega$  were computed (R-package "psych") (Table 2). The raw data is publicly available at the FORDATIS research data repository [41].

## 3 Results

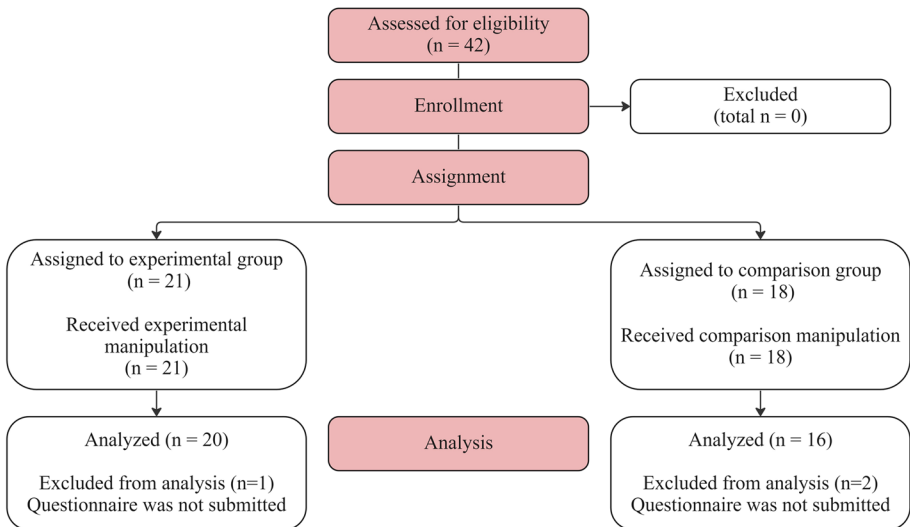
### 3.1 Participant flow

A total of 42 participants were distributed class-wise between the experimental group ( $n=21$ ) and the comparison group ( $n=18$ ). The sample sizes of the EG and the CG are different because the participants were recruited from the paramedic courses of two schools. The courses have different numbers of participants because there are no uniform guidelines for the size of the courses. Other factors, such as voluntary participation and absence due to illness, also come into play and are difficult to plan in advance. All participants in both groups completed the experiment. In the experimental group, one person opened the online questionnaire but did not submit it. In the comparison group, there were two persons with this same behavior. Figure 6 shows the flow of participants.

**Table 2** Reliability measurements

	Internal Consistency	
	Cronbach's Alpha	McDon-ald's Omega (h)
<b>SIMS</b>		
Intrinsic Motivation	0.76	0.76
Identified Regulation	0.81	0.80
External Regulation	0.90	0.89
Amotivation	0.83	0.70
<b>SSQ</b>		
Nausea	0.53	0.57
Disorientation	0.67	0.49
Oculomotor	0.77	0.54
<b>IPQ</b>		
General Factor	*	*
Involvement	0.48	0.37
Realism	0.77	0.76
Spatial Presence	0.63	0.51
<b>SUS</b>	0.79	0.74

\*Single item factor



**Fig. 6** Flow of participants

**3.2 Baseline data**

Age was distributed without practice-relevant differences in the two groups, but gender was not. There were significantly more males than females and non-binary persons (Table 3).

**Table 3** Age and sex – Complete sample and subgroups

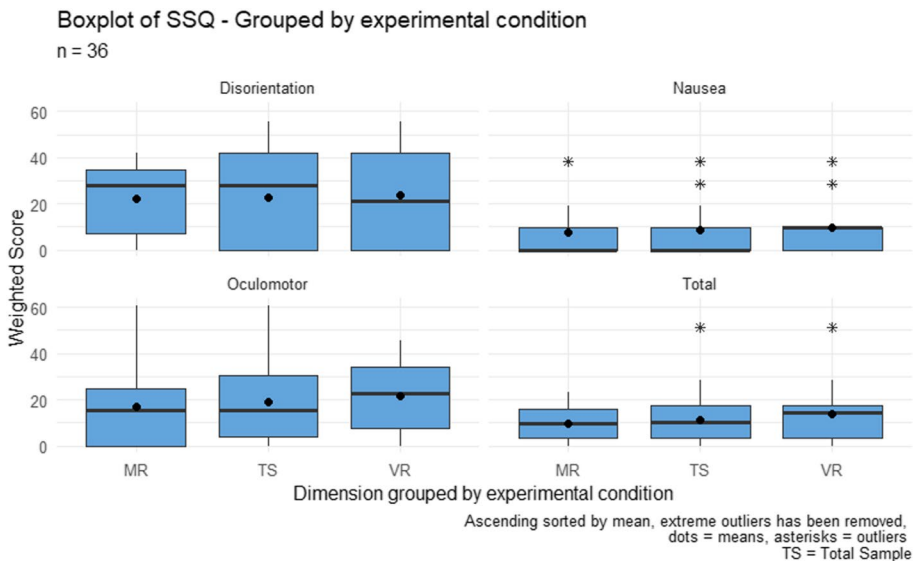
	Age Mean (SD)	Gender Absolute Frequency		
		Male	Female	Non-binary
Total	22.2 (2.1)	22	12	1
MR Group (EG)	22.4 (1.7)	13	5	1
VR Goup (CG)	22.0 (2.5)	9	0	0

### 3.3 Descriptive statistics

Missing data: One person did not answer any of the sociodemographic questions, and three persons did not answer the age question. The remaining variables were partially not answered by one person, without any specific pattern. Three persons did not submit the questionnaire at all. Table 4 shows the location and dispersion parameters (RQ1 & RQ2).

Figure 7 shows the results of the SSQ dimensions for the total sample (RQ1) grouped by experimental conditions (RQ2). Extreme outliers have been removed; i.e., we removed values that were 1.5 times the interquartile range above the third quartile or 1.5 times the interquartile range below the first quartile. An analysis at the item level shows that all those who reported moderate to severe problems with focusing (Items: “Difficult focusing”) and/or seeing clearly (Item: “Blurred vision”) wore a visual aid (glasses or contact lenses).

The SUS is additionally interpreted as a percentage on the grade scale [22, 34]. Table 5 shows the grouped results. Here, the SUS score was divided into sections (F to A) and underlaid with acceptability ranges to simplify interpretation according to Bangor. In each case, the numbers represent the frequency of participants allocated to the respective section after weighting.



**Fig. 7** Boxplot of SSQ scores

**Table 4** Means and standard deviations of constructs and factors

	Total Sample	MR Group	VR Group
	Mean (SD) Min.   Max	Mean (SD) Min.   Max	Mean (SD) Min.   Max
SIMS*			
Intrinsic Motivation	5.32 (1.07) 2.25   7.0	5.22 (1.24) 2.25   7.0	5.44 (0.82) 3.5   7.0
Identified Regulation	4.97 (1.27) 1.5   7.0	5.09 (1.4) 1.5   7.0	4.83 (1.12) 2.5   7.0
External Regulation	2.44 (1.59) 1   6.5	2.1 (1.24) 1   5.0	2.86 (1.89) 1   6.5
Amotivation	1.92 (0.99) 1   4.75	1.56 (1.25) 1   3.0	2.36 (1.14) 1   4.75
SSQ Total Score	11.43 (10.5) 0   51.4	9.6 (7.36) 0   23.22	13.73 (13.36) 0   51.4
Nausea	8.48 (12.2) 0   38.16	7.63 (12.23) 0   38.16	9.54 (12.07) 0   38.16
Disorientation	29.0 (28.89) 0   139.2	24.36 (19.09) 0   69.6	37.7 (34.8) 0   139.2
Oculomotor	20.84 (19.8) 0   83.38	17.06 (18.05) 0   60.64	25.58 (21.74) 0   83.38
IPQ*			
General Factor	4.94 (0.76) 4.0   6.0	5.0 (0.82) 4.0   6.0	4.88 (0.72) 4.0   6.0
Involvement	3.16 (0.96) 1.75   5.5	2.91 (0.68) 2.25   4.25	3.47 (1.17) 1.75   5.5
Realism	3.17 (0.95) 0.5   5.0	3.21 (1.05) 0.5   5.0	3.12 (0.84) 1.0   4.25
Spatial Presence	4.24 (1.05) 1.4   6.0	4.44 (0.98) 3.0   5.8	4.0 (1.11) 1.4   6.0
SUS	68.54 (15.58) 22.5   95.0	71.12 (13.34) 50.0   87.5	65.31 (17.91) 22.5   95.0

\*Total score not applicable

**Table 5** System Usability Scale – Grade scale (according to Bangor et al., 2009 [34])

Scale	F (0–60)	D (60–70)	C (70–80)	B (80–90)	A (90–100)
SUS total	8	11	6	10	1
MR Group	5	3	5	4	3
VR Group	3	7	2	3	1
Acceptability ranges	not acceptable or marginal low	marginal high	acceptable		

With respect to sample size, absolute frequencies are shown

### 3.4 Inferential statistics

**Mean score differences** Construct-level and factor-level *t*-Tests revealed a significant mean score difference for Amotivation ( $p=0.021$  [CI95=-1.47,-0.13]). No other significant differences were found regarding the groups. A significant mean difference ( $p=0.03$  [CI95: -26.47,-1.33]) between the three genders was found only for SUS between females (mean = 60.42, SD = 18.49) and males (mean = 74.32, SD = 11.42).

**Correlations** With reference to RQ2, significant correlations of  $r > 0.3$  are reported below. In both the VR group and the MR group, Identified Regulation (VR group:  $r=0.67$ , MR group:  $r=0.64$ ) and Intrinsic Motivation (VR group:  $r=0.70$ , MR group:  $r=0.83$ ) are highly correlated with Usability. Similarly, Spatial Presence is highly correlated with Usability (VR group:  $r=0.61$ , MR group:  $r=0.71$ ), Identified Regulation (VR group:  $r=0.55$ , MR group:  $r=0.70$ ), and Intrinsic Motivation (VR group:  $r=0.51$ , MR group:  $r=0.61$ ) in both groups. Involvement is highly negatively correlated with External Regulation ( $r=-0.67$ ) and Amotivation ( $r=-0.53$ ) only in the VR group. The MR group is further distinguished by the high negative correlation of VR Sickness with Intrinsic Motivation ( $r=-0.53$ ) and Usability ( $r=-0.63$ ). Further differences of the MR group are the high negative correlation of Usability with External Regulation ( $r=-0.69$ ) and Amotivation ( $r=-0.51$ ). Table 6 shows a summary of the correlations.

**Table 6** Significant correlations

Correlated Constructs	r	p	CI95	
			Lower Bound	Upper Bound
<b>VR Group</b>				
Identified Regulation ~ Usability	0.76	<0.001	0.42	0.91
Intrinsic Motivation ~ Usability	0.70	0.003	0.31	0.89
Spatial Presence ~ Usability	0.61	0.013	0.16	0.84
Spatial Presence ~ Identified Regulation	0.55	0.028	0.07	0.82
Spatial Presence ~ Intrinsic Motivation	0.51	0.043	0.02	0.80
Involvement ~ External Regulation	-0.67	0.004	-0.88	-0.27
Involvement ~ Amotivation	-0.53	0.036	-0.81	-0.04
<b>MR Group</b>				
Identified Regulation ~ Usability	0.74	<0.001	0.42	0.89
Intrinsic Motivation ~ Usability	0.83	<0.001	0.61	0.93
Spatial Presence ~ Usability	0.71	<0.001	0.38	0.88
Spatial Presence ~ Identified Regulation	0.70	<0.001	0.36	0.88
Spatial Presence ~ Intrinsic Motivation	0.61	0.006	0.21	0.83
VR-Sickness ~ Intrinsic Motivation	-0.53	0.019	-0.79	-0.10
VR-Sickness ~ Usability	-0.63	0.004	-0.84	-0.24
Usability ~ External Regulation	-0.69	0.001	-0.87	-0.34
Usability ~ Amotivation	-0.51	0.027	-0.78	-0.07

*p* values are Holm-adjusted for multiple testing

### 3.5 Secondary findings

In addition to the quantifiable results presented so far, we present three secondary findings. These are to be understood as observations of the study team. **First:** The technical environment can be challenging. This is mainly due to the various system components that need to be connected. Each component is critical to success. As a trivial example, empty batteries in one of the controllers will cause significant delays. In the course of the project, the development and subsequent use of a checklist became established. An example is provided in the appendix. **Second:** Wired HMDs can affect the sense of security and presence. At the time of the study, there were no high-performance wireless solutions available that could be integrated into the image setup without latencies. **Third:** Small-scale operations, such as the dispensing of medication with the controllers, were time-consuming to learn and perform. Suitable data gloves that allow intuitive movement were not available at the time the study was conducted.

## 4 Discussion

### 4.1 Interpretation

Regarding **RQ 1**, which includes the complete sample, it can be stated that the participants had a high level of Intrinsic Motivation and Identified Regulation. Low Amotivation and low External Regulation complete the picture. Overall, this means that the participants indicated a high level of intrinsic interest. The low External Regulation may have been caused by the integration into the curriculum itself, but, like Amotivation, it must not be overestimated due to the low expression. Nevertheless, it is worth questioning how the participants were prepared for the simulation and the evaluation by the training centers and whether competing events, such as imminent examinations, should be better considered in terms of time. Identified Regulation and Intrinsic Regulation were highly correlated with Spatial Presence, similar to Lerner et al. [42].

**VR Sickness:** The complete sample shows a mean value of the Total Score, which must be considered “significant”. Likewise, the mean values for Disorientation and Oculomotor need to be discussed in order to initiate improvement measures. The reasons for the high expression of Disorientation and Oculomotor are due to the structure of the SSQ: Both constructs contain items that address eye problems; in part, the same item is used in several constructs, like “Difficulty focusing”. In both Oculomotor and Disorientation, this item was the leading problem. Therefore, it must be emphasized that individual adaptation of the HMD is essential. Not only must the HMD fit securely without slipping, but the interpupillary distance and the distance from eye to lens must also be adjusted. It should be noted here that a secure fit is difficult to achieve due to personal protective precautions (i.e., hair net). This is also crucial for a good visual user experience and can possibly explain the high symptom expression in our case. Also, the limitations of the system for people with visual impairments need to be explored. In case of continuous use, individual solutions, such as corrective lenses for the HMD, may have to be provided.

**Presence:** The participants indicated a medium to high level for all facets. In detail, this means that Spatial Presence has already reached a satisfactory level, while Realism and Involvement are not a cause for concern, although there is still room for improvement. A comparison with the study by Girau et al. [16], which deals with a similar scenario in

MR and reports better results for realism, spatial presence, and involvement, supports the assumption that improvements are possible. In this respect, reference can be made to Muckler [43], who makes effective simulation dependent on the ability of teachers to help trainees overcome their disbelief. The following suggestion for improvement should be considered: Effective simulation depends on the ability of teachers to enable learners to overcome their disbelief. Here, the pedagogical professional must also be held responsible. The goal is to enable learners to accept the otherwise unrealistic aspects of a clinical simulation, according to Muckler. Muckler further notes: "The ability to suspend disbelief enhances the participant's level of immersion in the simulation." To overcome the suspension of disbelief, a fiction contract can be used. This is a "psychological contract" between learners and teachers, in the sense of establishing a "safe container", which includes the following components: clarifying expectations, attending to logistic details, and declaring and enacting a commitment to respecting learners and concern for their psychological safety [44]. Furthermore, controlling the application with respect to the graphical user interface can be discussed, for example in terms of a diegetic user interface as described by Salomoni et al. [27]. Concretely, it can be discussed whether a diegetic application would improve the virtualized fine motor tasks (e.g., drawing up medications) in terms of usability and realism.

The high correlation of Usability and Spatial Presence in both groups highlights the great importance of the spatial dimension of presence. Due to the connection between embodiment and presence, improvements could also be achieved through stronger embodiment [45]. This includes, for example, individualizable avatars (which are very important in embodiment [46]), wearing, e.g., professional clothing with which the respective group can identify themselves. In terms of embodiment and self-identification through avatars, we would like to highlight the paper by Dalgarno and Lee [47] for further explanations.

In terms of usability, the scores are close to the cut-off of 70, which suggests that improvements are necessary, but also that an intermediate level has been reached. Participants' comments indicate that the wired HMD was particularly annoying. Similarly, there was criticism regarding the fact that very small-scale movements, such as drawing up medications, would require a high level of attention and distract from the actual learning goal due to the unnatural imitation of movement by the controllers and occasional tracking problems with very fine movements. Unfortunately, due to a lack of devices with adequate performance, it was not possible to introduce "data gloves" or wireless HMDs as originally planned. There is a lot of ongoing development in this field, and we expect XR to get to a new level of immersion with the release of improved, affordable "data gloves". However, this idea should be pursued in future projects. It is worth investigating whether increasing the number of sensors could enhance the precision of haptic movement tracking. In the present configuration, the position of the HMDs and controllers is determined through triangulation using two external sensors, both of which require an unobstructed line of sight to the devices. Any obstruction to this sight line can cause momentary position loss and sudden "jumps" within the simulation. Introducing additional "lighthouses" might increase the stability and accuracy of device positioning. The high positive correlation between Usability and Identified Regulation in both groups as well as the high negative correlation of Involvement with External Regulation and Amotivation in the VR Group suggest the immense importance of targeted, planned pedagogical guidance.

The results regarding Usability and Presence are similar to Lerner et al. [42] in the predecessor VR project EPICSAVE and to Schild et al. [8] from previous evaluations.

With respect to **RQ 2**, a significant mean difference between the MR Group and the VR Group was found only for Amotivation. It is surprising that Amotivation was lower in the



experimental group, but no significant differences could be found for Intrinsic Motivation, Identified Regulation, and External Regulation. This can be seen as being in contrast to [4, 18, 19]. We point out the very different objectives and methodologies, which means that a comparison of the studies would require a deeper discussion of the methods and methodological effects. One possible explanation for the different findings is that the extraneous cognitive load [48] required by the operation outweighed the perception of the manikin during training. In future studies, it would be important to perform cognitive load measurements in addition to the current study design.

Understanding the underlying learning effectiveness is inherently intricate and methodologically challenging, as Hattie suggests [49]. Therefore, a longitudinal study should be conducted over several training sessions on different days to account for learning effects.

**In summary**, the described findings regarding Realism, Involvement, and Usability imply that there are main prerequisites that enable the acquisition of competencies as a basal criterion: The technical setting must be mastered in a suitable classroom without ifs and buts in order to avoid presence dropouts and distractors. VR/MR can only develop its full effect if used in a planned manner by pedagogical professionals. Pre-briefing, fiction contract, and structured debriefing can be assumed to be vital, especially with regard to the perception of Realism and Involvement. This hypothesis needs to be tested in future learning-outcome-oriented studies. Since competence-based assessment of learning outcomes is effort-intensive, a first step may be to consider measuring objectively assessable performance metrics such as accuracy, timing, and sequencing of actions [50]. A practical example that can be further extended is the measurement of no-flow time during BLS-CPR in VR by Issleib et al. [51]. Independent variables to be considered include the impact of familiarization and habituation on learning outcome. Abelson et al. [1] also highlight the importance of familiarization in prehospital case simulations.

## 4.2 Generalizability and limitations

The results show that trainees at paramedic training schools are open to technology and highly motivated. These are two important foundations for the use of “new” educational technologies.

In this study, internal validity may be biased by the novelty effect [52]. It should be noted that (without competent pedagogical supervision) a trainee’s engagement gradually wanes as the now familiar game elements and mechanisms no longer entertain, challenge, or satisfy the user [53]. Furthermore, it can be assumed that the operation of the virtual learning environment represents moderate to high extrinsic cognitive demand. This can lead to failure to achieve the learning outcomes. Both of these limitations can be addressed by repeating the measurement of the MR and VR application over multiple instructional sessions on different days. Regarding the low interrelatedness from the presence dimension Involvement, this has to be tested in larger samples in the field of HMD simulation.

In view of the sample size, the generalizability of this study must be discussed critically. Reproducibility should be examined in further studies. We would once again like to draw attention to the particularities of the restrictions imposed by Covid-19 at the time of the study, which did not allow larger samples to be included. Furthermore, in the case of self-report measures, it must be considered in the research design that considerable bias may occur [54].

Due to the restrictions on teaching associated with the Covid-19 pandemic, it was not possible to establish larger longitudinal learning outcome measurements. Regarding the

determination of knowledge gains in VR settings for paramedics and nursing students, we refer to the study by Schild et al. [8] in comparable settings.

VR and MR are inconceivable without an interface between man and machine. Following Joisten et al. [55], there is a relationship to the human lifeworld, a certain depth of intervention in the human lifeworld, and an influence on professional and learning practice. This repercussion on humans requires a techno-ethical perspective, which has not been adopted in this paper for the sake of the two research questions. However, a more detailed techno-ethical analysis is needed in future research.

With regard to external validity, an important strength of this study should be highlighted, namely, its practical relevance, as the evaluation was done in real learning situations at the educational partners. Thus, it can be assumed that the results can be transferred very well to social reality. However, the findings cannot be extrapolated uncritically to an international sample. Different training prerequisites as well as differences in professional socialization and understanding of roles must be considered. This is, however, mainly a question of pedagogical design and less one of basic technology-related impact factors.

### 4.3 Implications

The results of the survey and the evaluation itself have shown that the technical setting is demanding. Aspects such as customization of the HMD must be mastered in order to enable appropriate usability and a positive learning experience. To realize both, a "simulation faculty" is needed that can synergistically combine technical aspects and pedagogical aspects. Working with a checklist is recommended.

The results show that there is high motivation to learn in XR scenarios on the part of the paramedic trainees, which should be exploited as a potential. Where this potential should be used in comparison to conventional settings in terms of the relationship between cost, time required for the setup, and learning success must be weighed. Conceivable are primarily learning situations that are difficult to produce using conventional means, such as certain injuries or mass casualty incidents.

This leads to the next considerations that are familiar from conventional settings: Educational variables must be clarified and planned in advance. Even before the simulation, competence-based learning objectives must be defined. These considerations lead to the need to consider *ex ante* the competence-oriented learning goal also in terms of cognitive, affective, and psychomotor learning goals, including achievability within the available time. In order to consider an adequate level of stress and strains, the extrinsic cognitive load caused by the learning environment must also be included in these considerations. To strengthen involvement, a fiction contract is recommended. To enable reflection as an integral part of a case simulation, a structured debriefing should be established.

That an MR scenario is superior to a VR scenario could not be shown with this study. However, as a consequence for research, further studies designed as repeated measures in a within-subjects design are recommended to investigate the role of confounding variables such as familiarization effects, learning effects, and cognitive load.

For future studies, we recommend first an extension of the methodology, for example by focusing on competence-oriented outcome parameters. These could be structured according to curricular learning objectives. Second, we recommend that the sample be expanded and varied in terms of sample size and professional affiliation. In addition to larger samples, emergency nurses, medical staff, and other health professionals could be included. Third, variation in setting and intervention is recommended, such as future wired HMDs

versus wireless HMDs. In general, the challenging technical environment, the demanding didactic considerations, the pedagogical interventions, and the organization on site make an interprofessional study team indispensable.

#### 4.4 Conclusion

This study has a high degree of innovation as it is the first to address location-independent training from two different sites and to combine this with the pedagogical implications of interprofessional team training in EMS.

Looking at the results in terms of participant motivation, MR and VR appear to be learning environments that provide a sense of presence and realism and allow for psychological engagement of the participants. Participants in paramedic training were found to have high intrinsic and self-regulated motivation. For a planful and goal-oriented approach, we recommend creating a simulation checklist that addresses specific points of the XR simulation. Apparent trivialities, such as adequate fit of the HMD, can be critical factors for success. A competence-oriented approach is also recommended for XR simulations, as are a fiction contract and a structured debriefing, just like in a conventional case simulation. Technically, it is desirable to implement both wireless HMD and "data gloves". In the overall appraisal, we conclude that XR is not just a gimmick, but a promising learning method when used by a competent simulation faculty. XR case simulations are difficult to run by teachers as "lone wolves". Further studies addressing competence-based learning outcomes, techno-ethical perspectives and the role of habituation effects, learning effect, and cognitive load are needed.

## Appendix

### Additional resources

#### A. Checklist

##### Pedagogical Considerations:

- Competence-based learning objectives defined and discussed at the faculty?
- Control of the scenario (e.g., symptom progression or improvement) discussed at the faculty?
- Information about participants prepared with regard to goals and framework conditions?
- Fiction contract prepared?

##### Classroom:

- Playing field size sufficient?
- Playing field size in all scenarios in XR self-assessed?
- Tracking of people, hand and finger movements, and MR elements robust?
- Lighthouse placement appropriate?

##### System:

- For wired systems: Cables disentangled?
- Tripping hazards in the physical space eliminated?
- Controllers charged?
- For laptops: Connected to the power supply?
- Sound setting: Volume and sound direction appropriate?

### HMD:

- Participant wearing a hair net as PPE?
- HMD secure but not too tight? If necessary, consider several adjustment options.
- Eye-display distance adjusted?
- Pupil distance adjusted? (Not possible with every HMD).

### Hygiene and Personal Protective Equipment (PPE):

- Hair net, disposable gloves, mouth-nose mask provided if necessary?
- Hand sanitizer provided for participants and faculty at reasonable locations?
- Surface disinfection for grip surfaces in the room, manikin, and XR equipment available?
- Rapid tests or vaccination status required and participants and faculty informed?

### Participants:

- Educated about possible adverse side effects?
- Educated about importance of grids as spatial boundaries in XR?
- Physical condition of the day? Ready for the simulation?
- Fiction contract done? (Can also be done at class level instead of pairwise).

## B. Examination and treatment procedures

Examination/treatment	General description of examination and treatment	Description of examination and treatment in XR scenario (implications underlined)
Assessment of the accident		
Scene	Assessment of weather, location, need for further rescue resources, etc	The site of the operation is in the patient's garden, weather OK, visual flight conditions for possible helicopter transport, no further rescue equipment required. <u>Assessment as non-critical</u>
Safety	Self-protection comes first and is the top priority. Before approaching an accident scene, any hazards to the patient and the team must be eliminated	A kettle grill in the garden smokes only minimally, no further sources of danger. <u>Assessment as safe</u>
Situation	If safety is ensured at the accident site, the accident is recorded, including the so-called trauma kinematics	A man (approx. 45 years old) sitting in a garden chair has a visible burn injury mainly in the thoracic area and on the right arm. <u>Assessment as a potentially critical patient, communication of the outcome to the team</u>

Examination/treatment	General description of examination and treatment	Description of examination and treatment in XR scenario (implications underlined)
<b>Primary Survey</b>		
Airway management and cervical spine stabilization	For oxygen uptake, the patient needs a clear airway (A) and intact breathing (B). These are checked first. If the airway is not clear, the cause must be remedied quickly, e.g., by suctioning	Airways clear, no soot, no singed nasal hairs, no edema/swelling, no pathological flow sounds. <u>Assessment as non-critical</u>
Breathing (ventilation)	Breathing can be assessed by clinical examination with auscultation of the chest and calculation of the respiratory rate. Pulse oximetry provides further valuable information	Bilateral vesicular breath sounds, tidal volume sufficient but painful breath movements, respiratory rate 25/min. <u>Assessment as non-critical, oxygen application via oxygen mask</u>
Circulation (hemorrhage and perfusion)	Severe hemorrhage is one of the major causes of death after trauma. Therefore, relevant hemorrhage must be diagnosed or ruled out at an early stage	Peripheral pulse well palpable, regular, frequency between 85 – 100/min, recapillarization of the nail bed in under 2 s., non-invasive blood pressure 140/85 mmHg. <u>Assessment as non-critical</u>
Disability	In an initial neurological examination, the patient's state of consciousness and pupils should be checked	No neurological deficit, BG always above 100 mg/dl, pupils react laterally and promptly. <u>Assessment as non-critical</u>
Environment	Initially, the patient should be completely undressed and subjected to a brief head-to-toe examination in order to detect relevant severe injuries with a possible short-term impact on vital functions at an early stage	Injuries as described, temperature initially 35.5 °C ear, NRS initially 8 (severe pain). <u>Initiation of adequate pain relief and protection against further cooling via the injured body surface; this also includes wound care</u>

**Acknowledgements** We thank Prof. Dr. Vera Hagemann (University of Bremen, Germany) and Dr. David Häske (University Hospital Tübingen, Germany) for the unpaid scientific review as part of the institute's internal review process. We also thank PD Dr. Thomas Luiz (Fraunhofer IESE) for medical advice and Sonnhild Namingha (Fraunhofer IESE) for linguistic review.

The ViTAWiN project, which was the starting point for this article, was funded by the German Federal Ministry of Education and Research (BMBF), grant number 01PV1800.

**Authors' contributions** CE was responsible for the implementation of the evaluation. The evaluation concept was established and the data was analyzed in joint coordination with the consortium leader, Hannover University of Applied Sciences and Arts (Hannover, Germany). PD, DS, and CE were responsible for the interpretation of the data and the coherence of the resulting conclusion. Image processing was the task of DS, while PD was mainly responsible for literature research.

**Funding** Open Access funding enabled and organized by Projekt DEAL. The ViTAWiN project, which was the starting point for this article, was funded by the German Federal Ministry of Education and Research (BMBF), grant number 01PV1800. CE and DS were involved in the ViTAWiN project as research fellows at Fraunhofer IESE. They were responsible for the didactic design, parts of the evaluation, and emergency medical consultation.

**Data availability** The anonymized raw data is publicly available through the research data repository "FOR-DATIS" [41].

## Declarations

**Ethics approval and consent to participate** All participants were informed in advance, both verbally and in writing, about the objectives, background, risks, and privacy aspects. Voluntary participation and the right to delete the collected data were an explicit part of the survey. Written informed consent was obtained from all participants. Participation could be terminated or interrupted at any time and for any reason. A positive vote of the Commission for Transparency and Ethics in Research (Chair: Prof. Dr. Dieter Weber) of Hannover University of Applied Sciences and Arts is available (02 Sept. 2020). No reference number is available.

**Consent for publication** Written consent was also obtained for the production and publication of photographic material.

**Competing interests** CE and DS were involved in the ViTAWiN project as research fellows at Fraunhofer IESE. They were responsible for the didactic design, parts of the evaluation, and emergency medical consultation. No other conflicts exist. PD was not involved in the ViTAWiN project. No conflicts exist.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

1. Abellsson A, Rystedt I, Suserud B-O, Lindwall L (2014) Mapping the use of simulation in prehospital care - a literature review. *Scand J Trauma Resusc Emerg Med* 22:22. <https://doi.org/10.1186/1757-7241-22-22>
2. Boyle M, Williams B, Burgess S (2007) Contemporary simulation education for undergraduate paramedic students. *Emerg Med J* 24(12):854–857. <https://doi.org/10.1136/emj.2007.046318>
3. Braçq M-S, Michinov E, Jannin P (2019) Virtual reality simulation in nontechnical skills training for healthcare professionals: a systematic review. *Simul Healthc* 14(3):188–194. <https://doi.org/10.1097/SIH.0000000000000347>
4. Birtill M, King J, Jones D, Thyer L, Pap R, Simpson P (2023) The use of immersive simulation in paramedicine education: a scoping review. *Interact Learn Environ* 31(4):2428–2443. <https://doi.org/10.1080/10494820.2021.1889607>
5. Ryan GV, Callaghan S, Rafferty A, Higgins MF, Mangina E, McAuliffe F (2022) Learning outcomes of immersive technologies in health care student education: systematic review of the literature. *J Med Internet Res* 24(2):e30082. <https://doi.org/10.2196/30082>
6. Milgram P, Kishino F (1994) A taxonomy of mixed reality visual displays. *IEICE Trans Inf Syst* E77-D(12):1321–1329
7. Project description ViTAWiN. Available at: [https://www.iese.fraunhofer.de/en/customers\\_industries/referenzprojekt-vitawin.html](https://www.iese.fraunhofer.de/en/customers_industries/referenzprojekt-vitawin.html). Last accessed on: 05 Oct 2023
8. Schild J, Elsenbast C, Carbonell G ViTAWiN - developing multiprofessional medical emergency training with mixed reality:1–9. <https://doi.org/10.1109/SEGAH52098.2021.9551890>
9. 2018 IEEE 6<sup>th</sup> International Conference on Serious Games and Applications for Health (SeGAH): IEEE; 2018
10. Schild J, Lerner D, Misztal S, Luiz T EPICSAVE — Enhancing vocational training for paramedics with multi-user virtual reality:1–8. <https://doi.org/10.1109/SeGAH.2018.8401353>
11. Colonna AL, Robbins R, Stefanucci J, Creem-Regeh S, Patterson B, Engel BT et al (2022) Trauma bay virtual reality—a game changer for ATLS instruction and assessment. *J Trauma Acute Care Surg* 93(3):353–359. <https://doi.org/10.1097/TA.0000000000003569>

12. Chaabane S, Etienne A-M, Schyns M, Wagener A (2022) The impact of virtual reality exposure on stress level and sense of competence in ambulance workers. *J Trauma Stress* 35(1):120–127. <https://doi.org/10.1002/jts.22690>
13. Mills B, Dykstra P, Hansen S, Miles A, Rankin T, Hopper L et al (2020) Virtual reality triage training can provide comparable simulation efficacy for paramedicine students compared to live simulation-based scenarios. *Prehosp Emerg Care* 24(4):525–536. <https://doi.org/10.1080/10903127.2019.1676345>
14. Harrington CM, Kavanagh DO, Quinlan JF, Ryan D, Dicker P, O’Keeffe D et al (2018) Development and evaluation of a trauma decision-making simulator in Oculus virtual reality. *Am J Surg* 215(1):42–47. <https://doi.org/10.1016/j.amjsurg.2017.02.011>
15. Paletta L, Schneeberger M, Reim L, Kallus W, Peer A, Schonauer C et al (2022) Work-in-Progress—Digital human factors measurements in first responder virtual reality-based skill training. In: 2022 8<sup>th</sup> International Conference of the Immersive Learning Research Network (iLRN). IEEE, pp 1–3. Last accessed on: 23 Sep 2023
16. Girau E, Mura F, Bazzuro S, Casadio M, Chirico M, Solari F et al (2019) A Mixed Reality system for the simulation of emergency and first-aid scenarios. *Annu Int Conf IEEE Eng Med Biol Soc* 2019:5690–5695. <https://doi.org/10.1109/EMBC.2019.8856777>
17. Melo M, Gonçalves G, Monteiro P, Coelho H, Vasconcelos-Raposo J, Bessa M (2022) Do multi-sensory stimuli benefit the virtual reality experience? A systematic review. *IEEE Trans Vis Comput Graph* 28(2):1428–1442. <https://doi.org/10.1109/TVCG.2020.3010088>
18. Dinh HQ, Walker N, Hodges LF, Song C, Kobayashi A (1999) Evaluating the importance of multi-sensory input on memory and the sense of presence in virtual environments. In: *Proceedings IEEE Virtual Reality (Cat. No. 99CB36316)*: IEEE Comput. Soc, pp 222–228. Last accessed on: 23 Sep 2023
19. Biocca F, Kim J, Choi Y (2001) Visual touch in virtual environments: an exploratory study of presence, multimodal interfaces, and cross-modal sensory illusions. *Presence: Teleop Virt* 10(3):247–65. <https://doi.org/10.1162/105474601300343595>
20. Guay F, Vallerand RJ, Blanchard C (2000) On the assessment of situational intrinsic and extrinsic motivation: the Situational Motivation Scale (SIMS). *Motiv Emot* 24(3):175–213. <https://doi.org/10.1023/A:1005614228250>
21. Brooke J (1995) SUS: a quick and dirty usability scale. *Usability Eval. Ind.* 189
22. Bangor A, Kortum PT, Miller JT (2008) An empirical evaluation of the system usability scale. *Int J Hum-Comput Interact* 24(6):574–594. <https://doi.org/10.1080/10447310802205776>
23. Schubert T, Friedmann F, Regenbrecht H (2001) The experience of presence: factor analytic insights. *Presence: Teleop Virt* 10(3):266–81. <https://doi.org/10.1162/105474601300343603>
24. Kennedy RS, Lane NE, Berbaum KS, Lilienthal MG (1993) Simulator sickness questionnaire: an enhanced method for quantifying simulator sickness. *Int J Aviat Psychol* 3(3):203–220. [https://doi.org/10.1207/s15327108ijap0303\\_3](https://doi.org/10.1207/s15327108ijap0303_3)
25. Shkëmbi F, Treska V (2023) A review of the link between self-efficacy, motivation and academic performance in students. *Eur J Soc Sci* 10(1):23–31
26. Deci EL, Ryan RM (1980) The empirical exploration of intrinsic motivational processes. In: *Advances in experimental social psychology*, vol 13. Elsevier, pp 39–80. Last accessed on: 23 Sep 2023
27. Salomoni P, Prandi C, Rocchetti M, Casanova L, Marchetti L, Marfia G (2017) Diegetic user interfaces for virtual environments with HMDs: a user experience study with oculus rift. *J Multimodal User Interfaces* 11(2):173–184. <https://doi.org/10.1007/s12193-016-0236-5>
28. igroup.org – project consortium. igroup presence questionnaire (IPQ) Database; 2016. Available at: <http://www.igroup.org/pq/ipq/data.php>. Last accessed on: 23 Sep 23
29. Dużmańska N, Strojny P, Strojny A (2018) Can simulator sickness be avoided? A review on temporal aspects of simulator sickness. *Front Psychol* 9:2132. <https://doi.org/10.3389/fpsyg.2018.02132>
30. Stanney KM, Kennedy RS, Drexler JM (1997) Cybersickness is not simulator sickness. *Proc Hum Factors Ergon Soc Ann Meet* 41(2):1138–1142. <https://doi.org/10.1177/107118139704100292>
31. Stone Iii WB (2017) Psychometric evaluation of the Simulator Sickness Questionnaire as a measure of cybersickness. Last accessed on: 23 Sep 2023
32. Kim HK, Park J, Choi Y, Choe M (2018) Virtual reality sickness questionnaire (VRSQ): motion sickness measurement index in a virtual reality environment. *Appl Ergon* 69:66–73. <https://doi.org/10.1016/j.apergo.2017.12.016>
33. Chen W, Chen JZ, So R (2011) Visually induced motion sickness: effects of translational visual motion along different axes. *Contemporary Ergonomics and Human Factors* 2011:281–287. <https://doi.org/10.1201/b11337-47>
34. Bangor A, Kortum P, Miller J (2009) Determining what individual SUS scores mean: adding an adjective rating scale. *J Usability Stud* 4:114–123

35. Coombs WT, Algina J, Oltman DO (1996) Univariate and multivariate omnibus hypothesis tests selected to control type I error rates when population variances are not necessarily equal. *Rev Educ Res* 66(2):137–179. <https://doi.org/10.3102/00346543066002137>
36. Moser BK, Stevens GR (1992) Homogeneity of variance in the two-sample means test. *Am Stat* 46(1):19–21. <https://doi.org/10.1080/00031305.1992.10475839>
37. Moser BK, Stevens GR, Watts CL (1989) The two-sample t test versus satterthwaite's approximate f test. *Commun Stat – Theory Methods* 18(11):3963–3975. <https://doi.org/10.1080/03610928908830135>
38. Stonehouse JM, Forrester GJ (1998) Robustness of the t and U tests under combined assumption violations. *J Appl Stat* 25(1):63–74. <https://doi.org/10.1080/02664769823304>
39. Norman G (2010) Likert scales, levels of measurement and the “laws” of statistics. *Adv Health Sci Educ Theory Pract* 15(5):625–632. <https://doi.org/10.1007/s10459-010-9222-y>
40. Lee S, Lee DK (2018) What is the proper way to apply the multiple comparison test? *Korean J Anesthesiol* 71(5):353–360. <https://doi.org/10.4097/kja.d.18.00242>
41. Elsenbast C (2022) Psychometric and sociodemographic raw data of training evaluation of paramedic trainees with extended reality in classroom setting within project ViTAWiN. Available at: <https://fordatis.fraunhofer.de/handle/fordatis/296>. Last accessed on: 23 Sep 23
42. Lerner D, Mohr S, Schild J, Göring M, Luiz T (2020) An immersive multi-user virtual reality for emergency simulation training: usability study. *JMIR Serious Games* 8(3):e18822. <https://doi.org/10.2196/18822>
43. Muckler VC (2017) Exploring suspension of disbelief during simulation-based learning. *Clin Simul Nurs* 13(1):3–9. <https://doi.org/10.1016/j.ecns.2016.09.004>
44. Rudolph JW, Raemer DB, Simon R (2014) Establishing a safe container for learning in simulation: the role of the presimulation briefing. *Simul Healthc* 9(6):339–349. <https://doi.org/10.1097/SIH.000000000000047>
45. Riva G (1999) From technology to communication: psycho-social issues in developing virtual environments. *J Vis Lang Comput* 10(1):87–97. <https://doi.org/10.1006/jvlc.1998.0110>
46. Fowler C (2015) Virtual reality and learning: where is the pedagogy? *Brit J Educational Tech* 46(2):412–422. <https://doi.org/10.1111/bjet.12135>
47. Dalgarno B, Lee MJW (2010) What are the learning affordances of 3-D virtual environments? *Brit J Educational Tech* 41(1):10–32. <https://doi.org/10.1111/j.1467-8535.2009.01038.x>
48. Sweller J (2010) Element interactivity and intrinsic, extraneous, and germane cognitive load. *Educ Psychol Rev* 22(2):123–138. <https://doi.org/10.1007/s10648-010-9128-5>
49. Hattie J (2012) *Visible learning for teachers: maximizing impact on learning* / John Hattie. Routledge, London. Last accessed on: 23 Sep 2023
50. Baetzner AS, Wespi R, Hill Y, Gyllencreutz L, Sauter TC, Saveman B-I et al (2022) Preparing medical first responders for crises: a systematic literature review of disaster training programs and their effectiveness. *Scand J Trauma Resusc Emerg Med* 30(1):76. <https://doi.org/10.1186/s13049-022-01056-8>
51. Issleib M, Kromer A, Pinnschmidt HO, Süß-Havemann C, Kubitz JC (2021) Virtual reality as a teaching method for resuscitation training in undergraduate first year medical students: a randomized controlled trial. *Scand J Trauma Resusc Emerg Med* 29(1):27. <https://doi.org/10.1186/s13049-021-00836-y>
52. Tsay CH-H, Kofinas AK, Trivedi SK, Yang Y (2020) Overcoming the novelty effect in online gamified learning systems: an empirical evaluation of student engagement and performance. *Comput Assist Learn* 36(2):128–46. <https://doi.org/10.1111/jcal.12385>
53. Brickman P, Campbell D (1971) Hedonic relativism and planning the good society. Available at: <https://api.semanticscholar.org/CorpusID:140481039>. Last accessed on: 23 Sep 2023
54. Carpenter SK, Witherby AE, Tauber SK (2020) On students’ (mis)judgments of learning and teaching effectiveness. *J Appl Res Mem Cogn* 9(2):137–151. <https://doi.org/10.1016/j.jarmac.2019.12.009>
55. Joisten K, Thiemer N, Renner T, Janssen A, Scheffler A (2022) Focusing on the ethical challenges of data breaches and applications. In: 2022 IEEE International Conference on Assured Autonomy (ICAA). IEEE, pp 74–82. Last accessed on: 23 Sep 2023

**Publisher’s Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



## Authors and Affiliations

Christian Elsenbast<sup>1</sup>  · Philipp Dahlmann<sup>2</sup>  · Dominik Schnier<sup>1</sup> 

✉ Christian Elsenbast  
christian.elsenbast@iese.fraunhofer.de

Philipp Dahlmann  
philipp.dahlmann@th-deg.de

Dominik Schnier  
dominik.schnier@iese.fraunhofer.de

<sup>1</sup> Fraunhofer Institute for Experimental Software Engineering IESE, Research Program Digital Healthcare, Fraunhofer-Platz 1, 67663 Kaiserslautern, Germany

<sup>2</sup> Deggendorf Institute of Technology, Dieter-Görlitz-Platz 1, 94469 Deggendorf, Germany