Knowledge Gain and the Impact of Stress in a Virtual Reality-based Medical Emergencies Training with Automated Feedback – A Randomized Controlled Trial.

Marco Lindner, Tobias Leutritz, Joy Backhaus, Sarah König, Tobias Mühling

Submitted to: Journal of Medical Internet Research
on: October 11, 2024

Table of Contents

Knowledge Gain and the Impact of Stress in a Virtual Reality-based Medical Emergencies Training with Automated Feedback – A Randomized Controlled Trial.

Marco Lindner[1]; Tobias Leutritz[1] PhD; Joy Backhaus[1]; Sarah König[1] MD; Tobias Mühling[1] MD

[1]Institute of Medical Teaching and Medical Education Research University Hospital Würzburg Würzburg DE

Corresponding Author:
Tobias Mühling MD
Institute of Medical Teaching and Medical Education Research
University Hospital Würzburg
Joseph-Schneider-Straße 2
Würzburg
DE

Abstract

Background: A significant gap exists in the knowledge and procedural skills of medical graduates when it comes to managing emergencies. In response, highly immersive virtual-reality (VR)-based learning environments have been developed to train clinical competencies.

Objective: This study aimed to assess the effectiveness of virtual-reality-based (VR-based) simulation training, augmented with automated feedback, compared to video seminars at improving emergency medical competency among medical students. Furthermore, the study investigated the relationship between learning outcomes and physiological stress markers.

Methods: 72 senior medical students underwent VR-based emergency training (intervention) or viewed video seminars (control) on two topics (acute myocardial infarction and exacerbated chronic obstructive pulmonary disease) in an intra-individual crossover design. Levels of applied knowledge were assessed objectively by open-response tests pre/post-intervention and after 30 days. Additionally, two electrodermal activity markers representing physiological stress response were measured during VR sessions using a wearable sensor. Participants also rated their estimated learning success and perceived stress.

Results: Immediately after the intervention, short-term knowledge gain was comparable between both groups. However, VR training led to significantly better long-term knowledge gain compared to video seminars (VR: 17.8 ± 15.1%, control: 11.9 ± 18.0%, difference: -5.9, 95% CI [-11.5, -0.4]). Participants rated the VR training as significantly more effective for learning. While physiological stress markers generally increased during VR sessions, they correlated only weakly and negatively with knowledge gains. No correlation was found between perceived stress and knowledge outcomes.

Conclusions: VR-based simulation training with automated feedback can provide substantial long-term learning advantages over a traditional method in the context of emergency-medicine education. Although VR training induced stress, it had only a weak impact on performance and self-assessment. Given the time constraints faced by clinical educators, self-moderated VR-based learning proves to be a valuable addition to medical training.

(JMIR Preprints 11/10/2024:67412)
DOI: https://doi.org/10.2196/preprints.67412

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?
✓ **Please make my preprint PDF available to anyone at any time (recommended).**
Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.
Only make the preprint title and abstract visible.
No, I do not wish to publish my submitted manuscript as a preprint.
2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?
✓

Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).
Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain v
Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <a href="http

Original Manuscript

Knowledge Gain and the Impact of Stress in a Virtual Reality-based Medical Emergencies Training with Automated Feedback – A Randomized Controlled Trial.

Marco Lindner[1], Tobias Leutritz[1], Joy Backhaus[1], Sarah König[1], Tobias Mühling[1]*
[1]Institute of Medical Teaching and Medical Education Research, University Hospital Würzburg, Würzburg, Germany
*Corresponding author

Abstract:

Background: A significant gap exists in the knowledge and procedural skills of medical graduates when it comes to managing emergencies. In response, highly immersive virtual-reality (VR)-based learning environments have been developed to train clinical competencies.

Objective: This study aimed to assess the effectiveness of VR-based simulation training, augmented with automated feedback, compared to video seminars at improving emergency medical competency among medical students. Furthermore, the study investigated the relationship between learning outcomes and physiological stress markers.

Methods: 72 senior medical students underwent VR-based emergency training (intervention) or viewed video seminars (control) on two topics (acute myocardial infarction and exacerbated chronic obstructive pulmonary disease) in an intra-individual crossover design. Levels of applied knowledge were assessed objectively by open-response tests pre/post-intervention and after 30 days. Additionally, two electrodermal activity markers representing physiological stress response were measured during VR sessions using a wearable sensor. Participants also rated their estimated learning success and perceived stress.

Results: Immediately after the intervention, short-term knowledge gain was comparable between both groups. However, VR training led to significantly better long-term knowledge gain compared to video seminars (VR: 17.8 ± 15.1%, control: 11.9 ± 18.0%, difference: -5.9, 95% CI [-11.5, -0.4]). Participants rated the VR training as significantly more effective for learning. While physiological stress markers generally increased during VR sessions, they correlated only weakly and negatively with knowledge gains. No correlation was found between perceived stress and knowledge outcomes.

Conclusion: VR-based simulation training with automated feedback can provide substantial long-term learning advantages over a traditional method in the context of emergency-medicine education. Although VR training induced stress, it had only a weak impact on performance and self-assessment. Given the time constraints faced by clinical educators, self-moderated VR-based learning proves to be a valuable addition to medical training.

Keywords: Virtual Reality, Automated Feedback, Emergency Medicine, Medical Education, Knowledge Gain, Electrodermal Activity, Skin Conductance

Introduction

Although emergency medicine constitutes a fundamental component of most medical education curricula, existing evidence highlights deficiencies in differential diagnosis, initial treatment, and verbal handover during medical emergencies among graduates [1–3]. In response to such deficits, highly immersive virtual-reality (VR) learning environments have been developed to train resuscitation [4], clinical reasoning [5], team communication among paramedics [6], and the use of specific medical equipment [7]. A significant advantage of these software-based simulations is the potential incorporation of automated feedback (e.g., insights into beneficial or detrimental actions, end-of-session evaluations), which can enhance learning even in the absence of a tutor. The acceptance and the educational achievement of VR-based training programs are generally rated highly by participants [8,9], fostering the belief that they will play a crucial role in the future of emergency training. However, especially in the context of VR-based simulations in emergency medicine, the objective learning outcomes, as assessed through skills or knowledge tests, have not been systematically investigated to meet the quality criteria of meta-analyses [8,10].

When evaluating the learning outcomes of VR-based training, it is important to consider that immersive environments may influence the learning process in various ways. Evidence strongly supports the notion that VR increases learner motivation and improves the contextualization of the learning experience [11]. In a recent research project of ours, we explored different types of perceived stress encountered during VR-based training in medical emergencies [12]: We identified the sense of control within the scenario and the challenge of making correct decisions as stressors, which exhibit an inverse U-shaped and a negative relationship with the estimated learning success, respectively. These observations align with those of similar studies, in which excessive stress was found to undermine cognitive abilities [13–15].

Stress, undeniably, encompasses both psychological and physiological components. A more comprehensive understanding of stress responses could thus be achieved through the investigation of perceived stress along with physiological stress markers [16]. Electrodermal activity (EDA) as a measure of sympathetic nervous system activity is frequently employed to monitor psychophysiological activation in stressful conditions continuously [17]. EDA includes various parameters, among which the skin conductance level (SCL) and the frequency of non-specific skin conductance responses (NSSCR) are particularly suitable for assessing background psychophysiological activation in the absence of specific stimuli [18]. Both physiological stress markers can be recorded by wearable sensors (e.g., wristbands), causing minimal or no disruption to the simulation [19,20].

To evaluate objectively the learning outcomes of self-moderated VR-based emergency training with automated feedback and simultaneously to gain insight into the impact of different stress dimensions on the learning process, we aimed to answer the following questions:

1. Whether VR-based emergency simulation training is superior to video seminars for short- and long-term learning outcomes;
2. How perceived stress and physiological stress markers correlate during VR training sessions;
3. To what extent perceived stress and physiological stress markers align with learning outcomes and student s' demographics and characteristics.

Methods

VR-based simulation training and automated feedback

We utilized the training software STEP-VR (version 0.11b), which was co-developed together with ThreeDee GmbH (Munich, Germany), along with the previously described and recently evaluated hardware setup [12]. For the intervention, two VR-based case scenarios from the subject of internal emergency medicine were utilized: (A) chest pain resulting from acute myocardial infarction (MI), and (B) dyspnea caused by exacerbation of chronic obstructive pulmonary disease (COPD). Both scenarios were developed by board-certified specialists in internal medicine and emergency medicine with additional qualifications in medical education, based on guidelines from the relevant professional societies [21,22]. The medical content of the scenarios and their evaluation has been described in detail elsewhere [12,23].

Moreover, for this study, the software was complemented with an automated feedback system, providing recommendations based on medical guidelines. This system included various feedback modalities:

1. Positive notifications for correctly executed actions;
2. Warnings for actions that could potentially be harmful;
3. Prompts for missed time-critical actions using a “scaffolding” approach, in which additional cues are provided to guide the user without directly specifying the required action;
4. Access to the results of diagnostics (e.g. CT-scans) through a virtual representation of a patient documentation system;
5. A final evaluation presented in a checklist format, indicating the completion of relevant actions.

For points 1 – 4, brief messages, accompanied by an acoustic signal, were displayed immediately following the action. Users also had the option of accessing more detailed feedback through the virtual computer menu. Feedback components in the adapted version of STEP-VR are shown in Figure 1.

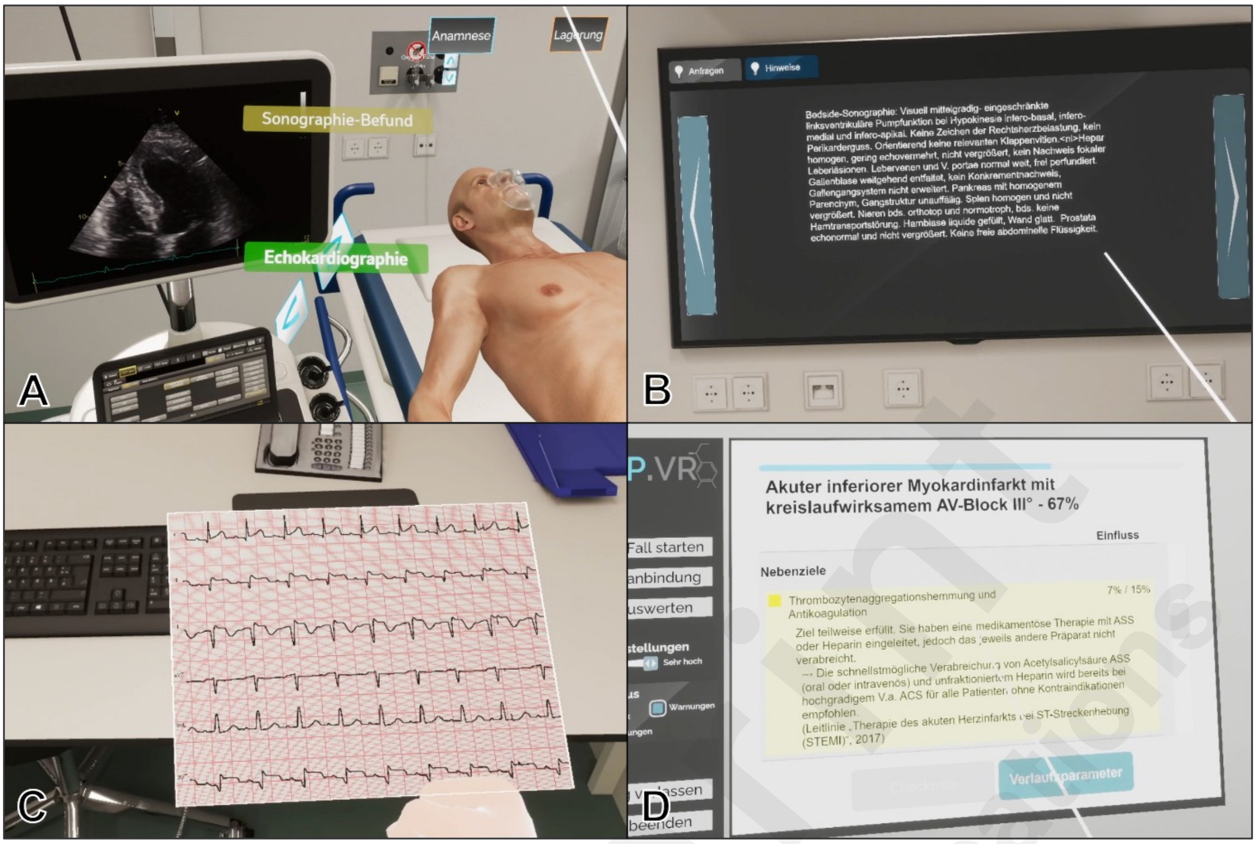


Figure 1: Feedback components in the adapted software. (A) positive notifications (green) for correctly executed actions, (B) results of diagnostics in the virtual computer menu or (C) as direct output from medical devices (parameters calculated by dynamic physiology), and (D) final evaluation in checklist format.

Study design and data collection

The study was conducted at the skills lab of a German medical school with a six-year medicine degree curriculum between June 2022 and May 2023. To ensure a uniform baseline knowledge level, only medical students who had passed the written examination in internal medicine (regularly at the end of the eighth semester) were eligible to enroll. Additionally, students with known epilepsy or severe simulation sickness were not considered eligible for this study. Recruitment was through semester mailing lists, with participants being offered a 15 Euro voucher as incentive. The simulation training itself took place outside regular teaching hours in the form of individual sessions. To avoid motivation bias among participants in a potential control group, we employed an intra-individual cross-over design. This design required all participants to complete two scenarios, each lasting 35 minutes: one utilizing VR (intervention) and the other in the form of a video seminar (control).

At the start of the intervention phase, all participants completed a five-minute VR tutorial, during which they could familiarize themselves with the functionality of the empty virtual emergency room while listening to explanations via an audio commentary. Following the tutorial, participants engaged in 20 minutes of a self-directed learning activity within the scenario, during which automated feedback was provided (learning phase, LP). During this self-moderated period, users decided themselves which measures they wanted to perform. Afterwards, participants reviewed the final evaluation checklist in the virtual doctors' office. To enhance content retention, they were given an additional ten minutes to revisit the scenario and practice the medical procedures correctly (repetition phase, RP). Throughout the whole session, a student tutor was available for technical support. As control, video seminars for scenario A and B lasting 35 minutes each were created comprising the same learning content in the form of a slide-based presentation. The lecturer interspersed content delivery with questions designed to foster reflection among the learners. Similar to the intervention, during the last 10 minutes participants were free to memorize the final summary or review specific slides from the presentation. All participants were randomly assigned to one of two groups. The first started with case A as VR scenario and then continued with case B as video seminar, while the second started with case A as video seminar and continued with case B as VR scenario. The random allocation sequence of whether Case A or B was completed as an intervention or control was conducted by ML using computer-generated random numbers, who assigned participants to the intervention. Blocks of size four were utilized, ensuring that the groups remained comparable throughout the recruitment process.

Students were asked to take applied knowledge tests with open-ended questions on the two case scenarios, both before and after the intervention or control session, following a pre-/post-test design. Additionally, students completed an online survey, in which they rated their estimated learning success (after both the intervention and control) and perceived stress (only following the intervention), along with providing person-specific data. To assess knowledge retention, the knowledge test was repeated after 30 days. Figure 2 provides an overview of the study design and data collection process.

Assuming a moderate effect size, a significance level of 0.05, and a power of 80%, the estimated sample size was approximately 63 participants. To account for a projected dropout rate of 10% before retention, we adjusted the total number of participants to 72.

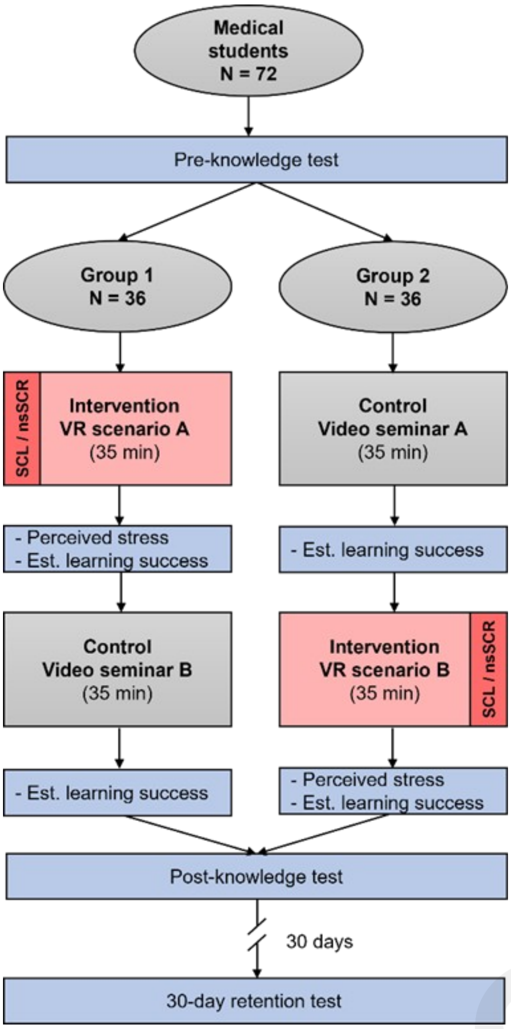


Figure 2: Study design and data collection process for knowledge tests and estimated learning success, as well as physiological stress markers (SCL and NSSCR) and perceived stress during intervention and control. Participants completed one scenario in VR and another as video seminar.

Questionnaires and measures

Knowledge tests

To assess the objective learning outcome, eight open-ended questions were created based on the scenario content, adhering to the guidelines provided by professional societies [21,22]. These questions focused on applied knowledge related to diagnosing and initiating treatment in the two emergency scenarios. Questions covered topics that were taught in both the intervention and control settings. A corresponding rubric was developed (Supplement Table 3) and tested during a pilot phase with a small number (n=8) of participants. It was subsequently refined to achieve optimal discriminative power. Participants could score a maximum of 1-3 points for correct answers, depending on the difficulty and complexity of the content. Student performance was rated according to the percentage of correctly answered questions. The pre-, post- and 30-day retention tests of knowledge all featured identical open-ended questions. Correct answers were never disclosed to participants. The responses were evaluated by two independent reviewers (a board-certified specialist in internal medicine with additional specialization in medical education (TM), and a medical student and trained skills lab tutor with 3 years of experience in rating of OSCEs (ML)) in a blinded manner. In cases of disagreement between the raters, the respective answers were considered correct.

Short and long-term knowledge gains were calculated as the differences between the post-test or 30-day-retention test and the pre-test, respectively.

Self-assessment questionnaires

Perceived stress: In a previous study [12], we identified three main stress factors in VR-based emergency training: “sense of control” (ability to impact patient outcomes), “challenge to perform” (difficulty in making correct decisions using specific knowledge), and “social interaction to learn” (feeling of being watched and judged by others). Owing to time constraints and practical considerations, the current study assessed these factors with just one relevant item each, based on their strong performance in the previous exploratory factor analysis. Students rated items on a 5-point Likert scale with 1 representing the minimum and 5 the maximum value. A separate overall rating for “total perceived stress” was evaluated based on a single item using the same scale. It should be noted that the questionnaire was used exclusively for the intervention (VR).

For the correlation analysis, the average of all four items (three factors + total perceived stress) was calculated and labelled as “average perceived stress”.

Estimated learning success: A similar approach was taken for the student self-assessment of estimated learning success represented by two factors previously identified as “didactic value” (participants' perceived general effectiveness of VR simulation) and “individual learning benefit” (the positive impact on the theoretical/practical skills of each participant). In this study, each factor was also evaluated with only one single item on the described 5-point Likert scale. Survey questionnaires with their items are listed in Supplement Table 2. This questionnaire was used for both the intervention and the control groups.

To calculate the correlations, the average of the two items was taken and labelled as "average estimated learning".

Measurement of physiological stress markers

EDA was measured using the Empatica E4 wristband, following the manufacturers guide and good practice recommendations [18]. As there were no defined stimuli present during the VR scenarios, we opted to measure the continuous parameters SCL and NSSCR, suited to capture spontaneous fluctuations in EDA in the presence of an ongoing sustained stimulus over a period of time [24]. During completion of the tutorial, 3 minutes of baseline data were recorded for each participant. Recording during the intervention phase was divided according to the study design into the LP and RP. SCL and temperatures of all recordings were within normal ranges of 0.05-60µS and 30-40°C, respectively [25]. Data were processed with R (Version 4.3.1, RRID:SCR_001905) using the wearables package [26]. We calculated the median values of SCL and NSSCR per minute during the respective observation phases (baseline, LP and RP), resulting in six parameters.

For correlation analysis, we wanted to account for inter-individual variability of SCL and NSSCR, as well as for the level of stress generated using the VR environment itself (without emergency scenario). Thus, the values of the individual baseline recording were subtracted from LP and RP data for every participant, yielding four parameters (two baseline-corrected physiological stress markers combined with two observation phases): SCL (LP), SCL (RP), NSSCR (LP), and NSSCR (RP).

Statistical analysis

All data were analyzed by employing descriptive statistics such as counts (both absolute and percentages), means, and standard deviations (SD). Regarding primary outcome data from knowledge tests, differences and 95% confidence intervals were also reported. As most of the parameters were not distributed normally (e.g. measures of EDA), we chose non-parametric tests: Wilcoxon rank-sum test was used for comparisons between groups and Spearman’s rank test was used to calculate correlation coefficients (r). Curvilinear relationships were calculated using quadratic regression analyses, and the fit of the models was expressed by R^2. The inter-rater reliability was calculated using Cohen's Kappa to quantify the agreement between the assessments of the independent reviewers. For all calculations excluding EDA data, GraphPad Prism (version 10.1.2, RRID:SCR_002798) was used. Reporting was primarily conducted in accordance with the CONSORT-EHEALTH recommendations [27].

Ethical considerations

Human subject ethics review approval: The local institutional review and ethics board judged the project as not representing medical or epidemiological research on human subjects and as such adopted a simplified assessment protocol. The project was approved without any reservation under the proposal number 20210917-01.

Informed consent: Students were informed about the study, and their participation was voluntary. Written informed consent was obtained in printed form from all participants, who were also provided with information on data processing for the analysis and the publication of results. Contact details were supplied for participants wishing to withdraw their consent to data processing. The decision to participate or not had no consequences on the students' academic progress.

Privacy and confidentiality: Survey data from the questionnaires were collected anonymously using the EvaSys® platform (Lüneburg, Germany). Data were processed and stored in accordance with local data protection laws.

Compensation details: A 15-euro book voucher was handed to the participants upon completion of the retention test.

Results

Participants, knowledge tests, and estimated learning success

A total of 72 participants were recruited for the study. All participants also completed the 30-day retention test, resulting in no dropouts. Person-specific factors are listed in Table 1. The gender distribution (64% female participants) was representative for students in the degree program in medicine. Overall, there was only little prior experience with VR and 3D applications.

Characteristics and experiences	Subgroup	n	%
Gender	Female	46	64
	Male	25	35
	Diverse	1	1
Academic semester	8	20	28
	9	23	32
	10	26	36
	>10	3	4
Experience in emergency medicine (e.g., voluntary service, clinical rotation)	Yes	32	44
	No	40	56
Frequency of of 3D-application use	Never	49	68
	Rarely	12	17
	Sometimes	3	4
	Frequent	7	10
	Daily	1	1
Cumulative experience with VR-applications	None	30	42
	<1h	26	36
	1-5h	13	18
	6-10h	1	1
	>10h	2	3

Table 1: Person-specific factors of participants.

Interrater reliability was consistently high in the knowledge test with a Cohen's Kappa of

0.927 averaged across all questions. Student performance is depicted in Figure 3. The prior knowledge of participants was similar in both the intervention and control groups for the pre-test (intervention: 57.8 ± SD 16.1%, control: 57.4 ± 16.4%, difference: -0.4, 95% CI [-5.8, 4.9]). Both groups exhibited a similar short-term increase immediately following the respective learning sessions (post-test: intervention: 84.4 ± 9.6%, control: 84.6 ± 11.0%, difference: 0.2, 95% CI [-3.2, 3.6]). However, the 30-day retention test revealed differences: knowledge acquired in the VR scenario was retained significantly better than that acquired from the video seminar (intervention: 75.4 ± 12.5%, control: 69.0 ± 14.5%, difference: -6.4, 95% CI [-10.9, -1.9]). At the level of scenarios, the difference in the 30-day retention test was more noticeable for the case of MI (intervention: 75.2 ± 15.3%, control: 67.4 ± 13.9%, difference: -7.8, 95% CI [-14.8, -0.8]). Conversely, for COPD, the difference was slightly less pronounced (intervention: 75.6 ± 9.0%, control: 70.5 ± 15.3%, difference: -5.1, 95% CI [-11.0, 0.9]). To account for prior knowledge, short-term gains (intervention: 26.6 ± 15.3%, control: 27.2 ± 16.0%, difference: 0.6, 95% CI [-4.5, 5.8]) and long-term gains (intervention: 17.8 ± 15.1%, control: 11.9 ± 18.0%, difference: -5.9, 95% CI [-11.5, -0.4]) were calculated.

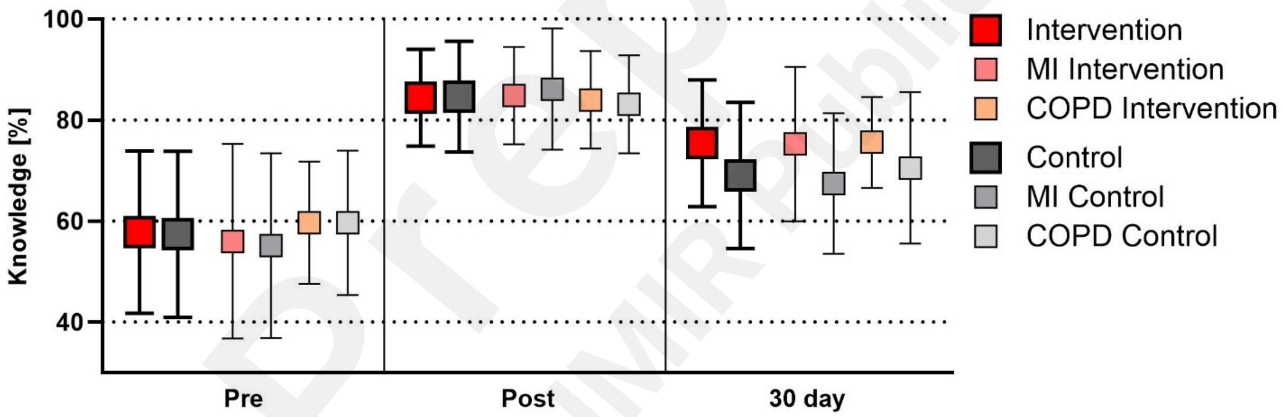


Figure 3: Knowledge test results as percentages of correct responses before (pre), immediately after (post) and 30 days after the intervention. Data are given in mean (box) with corresponding error bars (SD). The larger symbols and lines indicate overall results, while the smaller symbols detail the results for the individual scenarios.

On self-assessment by participants, the two items of estimated learning success in the VR scenarios approached the maximum possible value of 5 (didactic value: 4.83 ± 0.41 and individual learning benefit 4.85 ± 0.40). In contrast, these items in the control group had lower scores, only slightly above average (didactic value: 3.44 ± 1.00, individual learning benefit: 3.15 ± 1.15). Furthermore, the differences between the groups were statistically significant (each $p < .001$).

Physiological stress markers and perceived stress during VR scenarios

For all participants, parameters of EDA were recorded during the completion of the VR scenarios, as illustrated in Figure 4. SCL and nsSCR exhibited increases from baseline values (1.39 ± 1.98 µS and 3.05 ± 2.16 /min, respectively) to higher levels during the LP (3.63 ± 4.71 µS and 4.81 ± 3.16 /min, respectively), further rising to their peak values during the RP (4.16 ± 4.43 µS and 5.38 ± 3.26 /min, respectively). The differences between the average values of LP and RP for both SCL and nsSCR and their respective baseline values were highly significant (both $P < .001$). Throughout the study, there was considerable variability in the values, as indicated by relatively high standard deviations.

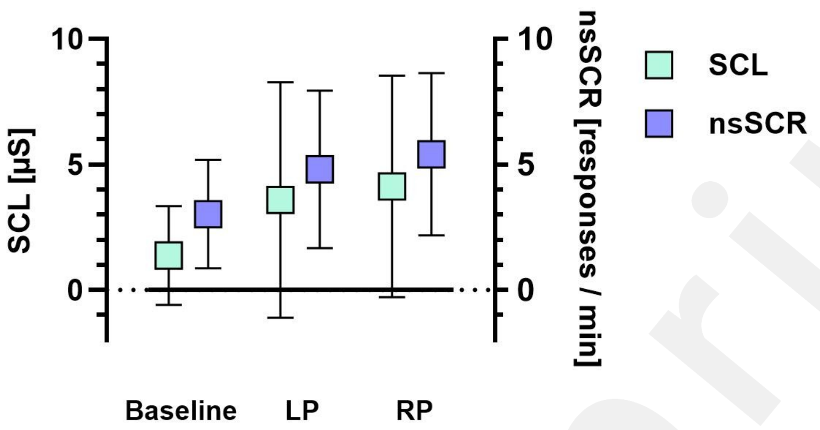


Figure 4: The progression of physiological stress markers SCL [µS] and nsSCR [responses per minute] during the completion of the VR scenarios. Data are given in mean (box) with corresponding error bars (SD).

Within the perceived stress questionnaire used in the VR scenarios, the factor "challenge to perform" received average ratings (3.00 ± 1.08) among participants, followed by the "sense of control" (2.51 ± 0.95) and low ratings for "social interaction to learn" (1.37 ± 0.80). Accordingly, moderate levels were reported for the “total perceived stress” item (2.86 ± 0.94).

Analysis of correlations

To investigate any potential influence of the baseline-corrected physiological stress markers (SCL and nsSCR) on short or long-term knowledge gains and students’ self-assessments (average perceived stress and average learning gain), correlations were calculated (Figure 5). The parameters within the same modality (knowledge gains and stress markers) correlated well with each other. However, weak and non-significant negative correlations were observed between short and long-term knowledge gains and baseline-corrected physiological stress markers recorded during the LP. The respective correlations were even less pronounced during the RP. Subsequently, quadratic regression analyses were performed to explore further the relationships between knowledge gains and average perceived stress as well as baseline-corrected physiological stress markers, respectively. While some plots visually suggested the presence of inverted U-shaped trendlines, all R^2 values fell below 0.2, indicating only weak correlation (data not illustrated).

Average perceived stress and average estimated learning exhibited only minimal correlation with knowledge gains and baseline-corrected physiological stress markers (all r < 0.2). Notably, there was no correlation between average perceived stress and average estimated learning (r = -0.09). All r-values are depicted in Figure 5 and corresponding P-values are listed in Supplement Table 1.

To evaluate the potential influence of (nominally and ordinally distributed) person-specific factors, group comparisons were conducted. No significant differences were found between gender, academic semester, prior experience level in emergency medicine, or experience with 3D or VR applications in relation to short and long-term knowledge gain or baseline-corrected physiological stress markers.

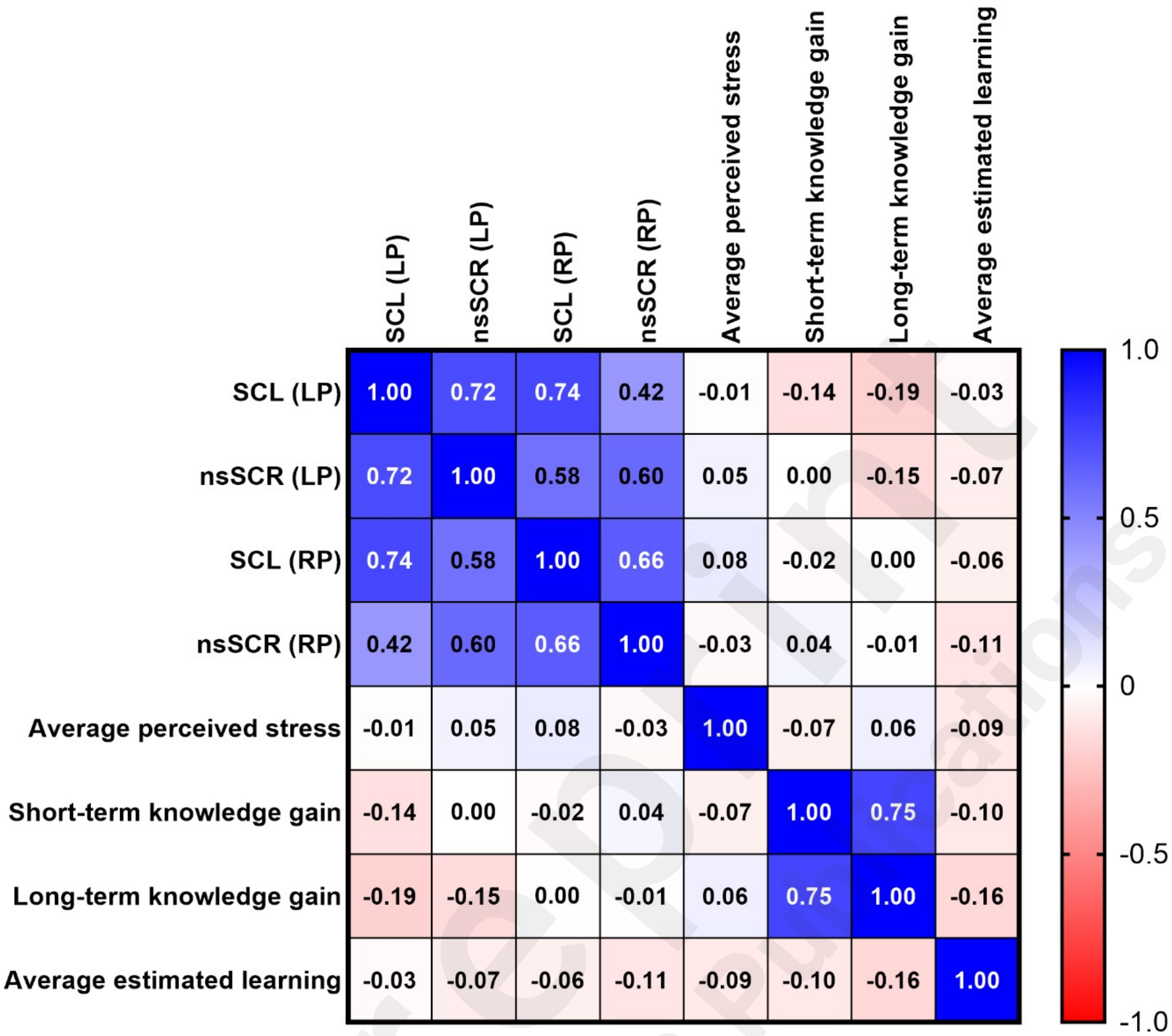


Figure 5: Correlation matrix between stress markers, knowledge gains, average perceived stress, and average estimated learning. The heatmap scale (right) indicates the color coding for positive (blue) and negative (red) correlations. The more intense the color, the higher the correlation coefficient. Corresponding significance values are listed in Supplement Table 1.

Discussion

This study, conducted within the curricular framework of a degree course in medicine, employed a randomized, controlled design to assess short and long-term knowledge gains, as well as estimated learning success resulting from self-moderated learning in a complex VR-based medical emergencies training. Moreover, we investigated the influence of physiological stress markers and perceived stress among students.

Focusing on our primary objective, we found significantly greater knowledge retention after 30 days following a 35-minute self-moderated VR learning session with automated feedback, when compared to a video seminar of equivalent duration and content. The moderate effect size must also be considered in the context of the high prior knowledge (almost 60%) and the very explicit presentation of medical facts in the control group. While VR-based programs are increasingly integrated into medical curricula, small sample sizes and heterogeneous study designs hinder clear conclusions on the effectiveness of learning with such programs [28]. To date, the number of randomized, controlled trials is limited, and results are partly contradictory: For VR-based resuscitation training, two studies reported different results. The VR-training group had either better or worse no-flow times, dependent on which control group was chosen (web-based vs. "classic" basic life support course) [4,29]. Another study examining the effectiveness of VR-based pediatric intubation training reported similar scores compared to learning with a checklist [30]. Although it is known that context-specific knowledge tends to be anchored in memory better, we found no studies that included long-term outcomes by measuring knowledge or skill retention in VR-based emergency training.

A lack of significant correlations between objective knowledge gains and estimated learning success has been observed among learners in various medical disciplines. Both medical students and junior doctors often face challenges in accurately self-assessing their own competence in various clinical skills [31,32]. Research from cognitive science suggests that self-assessment of learning is influenced more by affective factors, such as enthusiasm for a subject or a teaching method, rather than by objective knowledge gain [33]. This seems particularly plausible in the context of engaging and innovative teaching methods that predominantly elicit positive emotional responses among learners. We conclude that self-reported learning outcomes - especially those from such methods - should ideally be supplemented or even substituted by objective assessments of competence.

The second objective of our study was to quantify physiological stress responses during VR-based learning and correlate them with levels of perceived stress. We successfully measured two physiological stress markers, SCL and nsSCR, in all participants. Both parameters exhibited similar increases during the scenarios, consistent with literature benchmarks [18]. Current evidence suggests that autonomic nervous system responses may be less pronounced in VR settings compared to real-life situations [34], yet more prominent than those observed in similar 2D screen settings [35]. Notably, SCLs were higher than average in stress-inducing VR-based disaster scenarios [36] and from VR environments designed specifically to trigger stress [37], indicating an authentic representation of the VR case scenarios. It is surprising that both physiological stress markers were lower during the LP (presumably the more stressful activity) than during the RP (where learning content was supposed to be reproduced). This is plausible for SCL, which can drift towards higher values over the recording period under certain circumstances (e.g. change in skin temperature), as reported [18]. However, we also observed a similar increase in nsSCR during the RP, indicating that the rise from LP to RP was not merely an artifact of SCL. One possible explanation is that participants might have found it challenging to concentrate on reproducing the medical procedures during the RP, especially after just receiving (potentially poor) results from the final evaluation. This type of stress during the RP might differ qualitatively from that experienced during the LP, potentially explaining the differing correlation levels observed between the two phases (see next paragraph). Interestingly, we found no correlation between the physiological stress markers and the identified items for perceived stress and the average perceived stress, respectively. This finding contrasts with a recent review, which reported that the majority (~87%) of studies across various contexts found a positive association between EDA and perceived stress [38]. However, the reported degree of correlation varies strongly even within the specific context of simulation training, from no correlation in a simulated airway emergency [39] to a very high correlation during surgical skills training [40]. These findings underscore the need for further research into different measures of stress responses in VR settings (e.g. heart rate variability or salivary cortisol) to improve our understanding and contextualize our results.

Lastly, we explored the relationship between physiological stress markers and perceived stress in relation to knowledge outcomes. Our findings revealed a weak negative correlation between physiological stress markers recorded during the LP and long-term knowledge gain, while correlations with stress markers recorded during the RP were virtually absent. It is generally recognized that affects and emotions play a pivotal role in influencing learning and memory performance [28]. Nonetheless, the exact relationship between physiological stress parameters and learning or performance outcomes remains somewhat ambiguous. For instance, in similar training sessions designed to establish intraosseous access [41], conduct pediatric intubation [42], or executed simulated assessments [43], key physiological stress markers like salivary cortisol and heart rate variability displayed no correlation with performance outcomes. Contrastingly, broader data from simulation contexts generally suggest that stress adversely affects learning outcomes [15,44], and that stress management training can improve performance during critical simulated situations [30]. The lack of correlation between average perceived stress and average estimated learning success aligns with the divergent objective outcomes (knowledge and physiological stress response), which was anticipated, given the absence of consistent correlations between subjective and objective measures within the same domain. To elucidate these discrepancies further, a future research direction could involve a more detailed qualitative approach, such as the repeated assessment of subjective stress dimensions throughout the VR scenario. This method could provide deeper insight into how stress dynamics uniquely affect individual learning trajectories in immersive virtual environments.

Strengths

The study benefited primarily from a relatively large sample size, which was representative of the medical student population at the study location. Knowledge was assessed at multiple time points and results largely corresponded across two scenarios with distinct content, thereby increasing the generalizability of the study. Established physiological stress markers such as SCL and nsSCR were recorded using validated tools, alongside a baseline measurement, augmenting the robustness of the findings. The greatest strength, however, is the randomized, controlled, and single-blinded study design in which these measures were conducted.

Limitations

Several factors limit the generalizability of this study, however. Firstly, the sample consisted of medical students from a single academic institution, which may not reflect broader demographics. Secondly, the learning sessions for participants were relatively short, with a duration of only 35 minutes. It would be interesting, albeit logistically challenging, to explore (dose-dependent) effects of repeated learning sessions. Thirdly, the participants in this study were not blinded regarding whether they were currently receiving the intervention or control condition, which may have influenced their performance. We aimed to minimize at least the motivational bias by using an intra-individual cross-over design, where each participant received both the intervention and control in different case scenarios. Lastly, our study focused solely on applied knowledge. Given the simplified representation of invasive procedures in the program and the lack of haptic feedback from VR controllers, we chose not to assess practical skills. Future research should investigate whether VR can effectively support long-term skill-based learning outcomes.

Conclusion

This study provides new insights into the durability and effectiveness of learning outcomes achievable through VR technology in emergency simulation training implemented in the curriculum of a degree course program in medicine. The results strongly suggest that the immersive, practical learning experiences offered by VR-based scenarios yield long-term educational benefits. Given the time constraints faced by clinical educators and the heterogeneity of curricula regarding complex emergencies, self-moderated VR-based learning stands out as a valuable addition to medical education. Moreover, employing physiological stress markers within VR scenarios could provide a novel approach to identify moments when stress impacts knowledge acquisition. Leveraging such data could enhance the learning experience, for instance, by adapting the presentation of information or minimizing distractions in the virtual environment in future implementations.

Acknowledgments

We would like to thank André Schulz (University of Luxembourg), Ivo Käthner and Daniel Gromer (both from the University of Würzburg), as well as Morris Gellisch (University of Bochum) for their assistance in analyzing the EDA data. Moreover, we thank Sarah Rickenbacher-Frey (Kantonsspital Aarau, Switzerland), as well as Thomas Sauter and Tanja Birrenbach (Inselspital Bern, Switzerland) for their didactic conception of the COPD scenario. Lastly, we would like to thank Andrew Entwistle for his critical review of the manuscript.

Funding

Funding for this work was provided by the *Research Support Award for 2022* (Forschungsförderpreis 2022) from the non-profit Vogel-Foundation Dr. Eckernkamp.

Data Availability

The datasets generated during and analyzed during this study are available from the corresponding author on reasonable request.

Conflict of Interest

TM was involved in the software development process of STEP-VR. All other authors report no conflict of interest.

Author contributions

TM was responsible for the conceptualization of the study, acquiring funding, supervising the execution, analyzing the knowledge and EDA data, and writing the original draft. ML contributed to the conceptualization, conducted the investigation, and performed analysis of knowledge and EDA data. JB assisted in conceptualization of the study and performed statistical analysis. TL was involved in EDA data curation, formal analysis, and the review, and editing of the manuscript. SK also contributed to the conceptualization and participated in writing, review, and editing of the manuscript.

Generative AI disclosure

No generative AI was used in the writing of this manuscript.

Supplement Table 1: P-values of the correlation matrix (Figure 3) between baseline-corrected parameters of the physiological stress response and knowledge tests at different time points, as well as the subjective measures of perceived stress and estimated learning success. P-values below the significance level of 0.05 are printed in bold.

	SCL (LP)	nsSCR (LP)	SCL (RP)	nsSCR (RP)	Average perceived stress	Short-term knowledge gain	Long-term knowledge gain	Estimated learning success
SCL (LP)		**<.001**	**<.001**	**<.001**	.91	.26	.12	.83
nsSCR (LP)	**<.001**		**<.001**	**<.001**	.67	.99	.23	.56
SCL (RP)	**<.001**	**<.001**		**<.001**	.52	.89	.00	.60
nsSCR (RP)	**<.001**	**<.001**	**<.001**		.78	.73	.94	.36
Average perceived stress	.91	.67	.52	.78		.58	.59	.43
Short-term knowledge gain	.26	.99	.89	.73	.58		**<.001**	.40
Long-term knowledge gain	.12	.23	.00	.94	.59	**<.001**		.18
Estimated learning success	.83	.56	.60	.36	.43	.40	.18	

Supplement Table 2: Subjective measures questionnaire containing questions on perceived stress and estimated learning success (1 to 5 on a 5-point Likert scale).

Perceived Stress	
1.	I felt stressed because many things in the scenario were beyond my control.
2.	I felt stressed because I lacked the expertise to handle the case.
3.	The presence of the tutor put pressure on me.
4.	Overall, I felt stressed during the completion of the case scenario.
Estimated Learning Success	
5.	The teaching method is suitable as a learning tool for acquiring emergency medical knowledge.
6.	I benefited from the recent learning module in terms of my own practical skills.

References

1. Monrouxe LV, Grundy L, Mann M, John Z, Panagoulas E, Bullock A, Mattick K. How prepared are UK medical graduates for practice? A rapid review of the literature 2009-2014. BMJ Open 2017;7(1):e013656. PMID:28087554
2. Till Johannes Bugaj, Christoph Nikendei, Jan Benedikt Groener, Jan Stiepak, Julia Huber, Andreas Möltner, Wolfgang Herzog, Ansgar Koechel. Ready to run the wards? – A descriptive follow-up study assessing future doctors' clinical skills.
3. Burridge S, Shanmugalingam T, Nawrozzadeh F, Leedham-Green K, Sharif A. A qualitative analysis of junior doctors' journeys to preparedness in acute care. BMC Med Educ 2020;20(1):12. PMID:31931784
4. Issleib M, Kromer A, Pinnschmidt HO, Süss-Havemann C, Kubitz JC. 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 2021;29(1):27. PMID:33526042
5. Mahling M, Wunderlich R, Steiner D, Gorgati E, Festl-Wietek T, Herrmann-Werner A. Virtual Reality for Emergency Medicine Training in Medical School: Prospective, Large-Cohort Implementation Study. J Med Internet Res 2023;25:e43649. PMID:36867440
6. Lerner D, Mohr S, Schild J, Göring M, Luiz T. An Immersive Multi-User Virtual Reality for Emergency Simulation Training: Usability Study. JMIR Serious Games 2020;8(3):e18822. PMID:32735548
7. Wolff G, Bruno RR, Reiter M, Kantzow B, Kelm M, Jung C. Virtual reality device training for extracorporeal membrane oxygenation. Crit Care 2020;24(1):390. PMID:32616025
8. Kim H-Y, Kim E-Y. Effects of Medical Education Program Using Virtual Reality: A Systematic Review and Meta-Analysis. Int J Environ Res Public Health 2023;20(5). PMID:36900904
9. Mergen M, Graf N, Meyerheim M. Reviewing the current state of virtual reality integration in medical education - a scoping review. BMC Med Educ 2024;24(1):788. PMID:39044186
10. Liu JYW, Yin Y-H, Kor PPK, Cheung DSK, Zhao IY, Wang S, Su JJ, Christensen M, Tyrovolas S, Leung AYM. The Effects of Immersive Virtual Reality Applications on Enhancing the Learning Outcomes of Undergraduate Health Care Students: Systematic Review With Meta-synthesis. J Med Internet Res 2023;25:e39989. PMID:36877550
11. Gutiérrez F, Pierce J, Vergara VM, Coulter R, Saland L, Caudell TP, Goldsmith TE, Alverson DC. The effect of degree of immersion upon learning performance in virtual reality simulations for medical education. Stud Health Technol Inform 2007;125:155-160. PMID:17377256
12. Mühling T, Späth I, Backhaus J, Milke N, Oberdörfer S, Meining A, Latoschik ME, König S. Virtual reality in medical emergencies training: benefits, perceived stress, and learning success. Multimedia Systems 2023;29(4):2239-2252. doi:10.1007/s00530-023-01102-0
13. Wetzel CM, Black SA, Hanna GB, Athanasiou T, Kneebone RL, Nestel D, Wolfe JHN, Woloshynowych M. The effects of stress and coping on surgical performance during simulations. Ann Surg 2010;251(1):171-176. PMID:20032721
14. McGraw LK, Out D, Hammermeister JJ, Ohlson CJ, Pickering MA, Granger DA. Nature, correlates, and consequences of stress-related biological reactivity and regulation in Army nurses during combat casualty simulation. Psychoneuroendocrinology 2013;38(1):135-144. PMID:22710003
15. Mauriz E, Caloca-Amber S, Córdoba-Murga L, Vázquez-Casares AM. Effect of Psychophysiological Stress and Socio-Emotional Competencies on the Clinical Performance of Nursing Students during a Simulation Practice. Int J Environ Res Public

Health 2021;18(10). PMID:34069709
16. Gellisch M, Bablok M, Brand-Saberi B, Schäfer T. Neurobiological stress markers in educational research: A systematic review of physiological insights in health science education. Trends in Neuroscience and Education 2024;37:100242. doi:10.1016/j.tine.2024.100242
17. Bhoja R, Guttman OT, Fox AA, Melikman E, Kosemund M, Gingrich KJ. Psychophysiological Stress Indicators of Heart Rate Variability and Electrodermal Activity With Application in Healthcare Simulation Research. Simul Healthc 2020;15(1):39-45. PMID:32028446
18. Boucsein W. Electrodermal Activity. Berlin [u.a.]: Springer; 2012. ISBN:978-1-4614-1125-3.
19. Fadeev KA, Smirnov AS, Zhigalova OP, Bazhina PS, Tumialis AV, Golokhvast KS. Too Real to Be Virtual: Autonomic and EEG Responses to Extreme Stress Scenarios in Virtual Reality. Behav Neurol 2020;2020:5758038. PMID:32256856
20. Halbig A, Latoschik ME. A Systematic Review of Physiological Measurements, Factors, Methods, and Applications in Virtual Reality. Front. Virtual Real. 2021;2. doi:10.3389/frvir.2021.694567
21. Ibanez B, James S, Agewall S, Antunes MJ, Bucciarelli-Ducci C, Bueno H, Caforio ALP, Crea F, Goudevenos JA, Halvorsen S, Hindricks G, Kastrati A, Lenzen MJ, Prescott E, Roffi M, Valgimigli M, Varenhorst C, Vranckx P, Widimský P. 2017 ESC Guidelines for the management of acute myocardial infarction in patients presenting with ST-segment elevation: The Task Force for the management of acute myocardial infarction in patients presenting with ST-segment elevation of the European Society of Cardiology (ESC). Eur Heart J 2018;39(2):119-177. PMID:28886621
22. Vogelmeier C, Buhl R, Burghuber O, Criée C-P, Ewig S, Godnic-Cvar J, Hartl S, Herth F, Kardos P, Kenn K, Nowak D, Rabe KF, Studnicka M, Watz H, Welte T, Windisch W, Worth H. Leitlinie zur Diagnostik und Therapie von Patienten mit chronisch obstruktiver Bronchitis und Lungenemphysem (COPD). [Guideline for the Diagnosis and Treatment of COPD Patients - Issued by the German Respiratory Society and the German Atemwegsliga in Cooperation with the Austrian Society of Pneumology]. Pneumologie 2018;72(4):253-308. PMID:29523017
23. Rickenbacher-Frey S, Adam S, Exadaktylos AK, Müller M, Sauter TC, Birrenbach T. Development and evaluation of a virtual reality training for emergency treatment of shortness of breath based on frameworks for serious games. GMS J Med Educ 2023;40(2):Doc16. PMID:37361243
24. Hugo F. Posada-Quintero and Ki H. Chon. Innovations in Electrodermal Activity Data Collection and Signal Processing: A Systematic Review.
25. Kleckner IR, Jones RM, Wilder-Smith O, Wormwood JB, Akcakaya M, Quigley KS, Lord C, Goodwin MS. Simple, Transparent, and Flexible Automated Quality Assessment Procedures for Ambulatory Electrodermal Activity Data. IEEE Trans Biomed Eng 2018;65(7):1460-1467. PMID:28976309
26. Looff P de, Duursma R, Noordzij M, Taylor S, Jaques N, Scheepers F, Schepper K de, Koldijk S. Wearables: An R Package With Accompanying Shiny Application for Signal Analysis of a Wearable Device Targeted at Clinicians and Researchers. Front Behav Neurosci 2022;16:856544. PMID:35813597
27. Eysenbach G. CONSORT-EHEALTH: improving and standardizing evaluation reports of Web-based and mobile health interventions. J Med Internet Res 2011;13(4):e126. PMID:22209829
28. Wu T-C, Ho C-TB. A scoping review of metaverse in emergency medicine. Australas Emerg Care 2023;26(1):75-83. PMID:35953392
29. Moll-Khosrawi P, Falb A, Pinnschmidt H, Zöllner C, Issleib M. Virtual reality as a

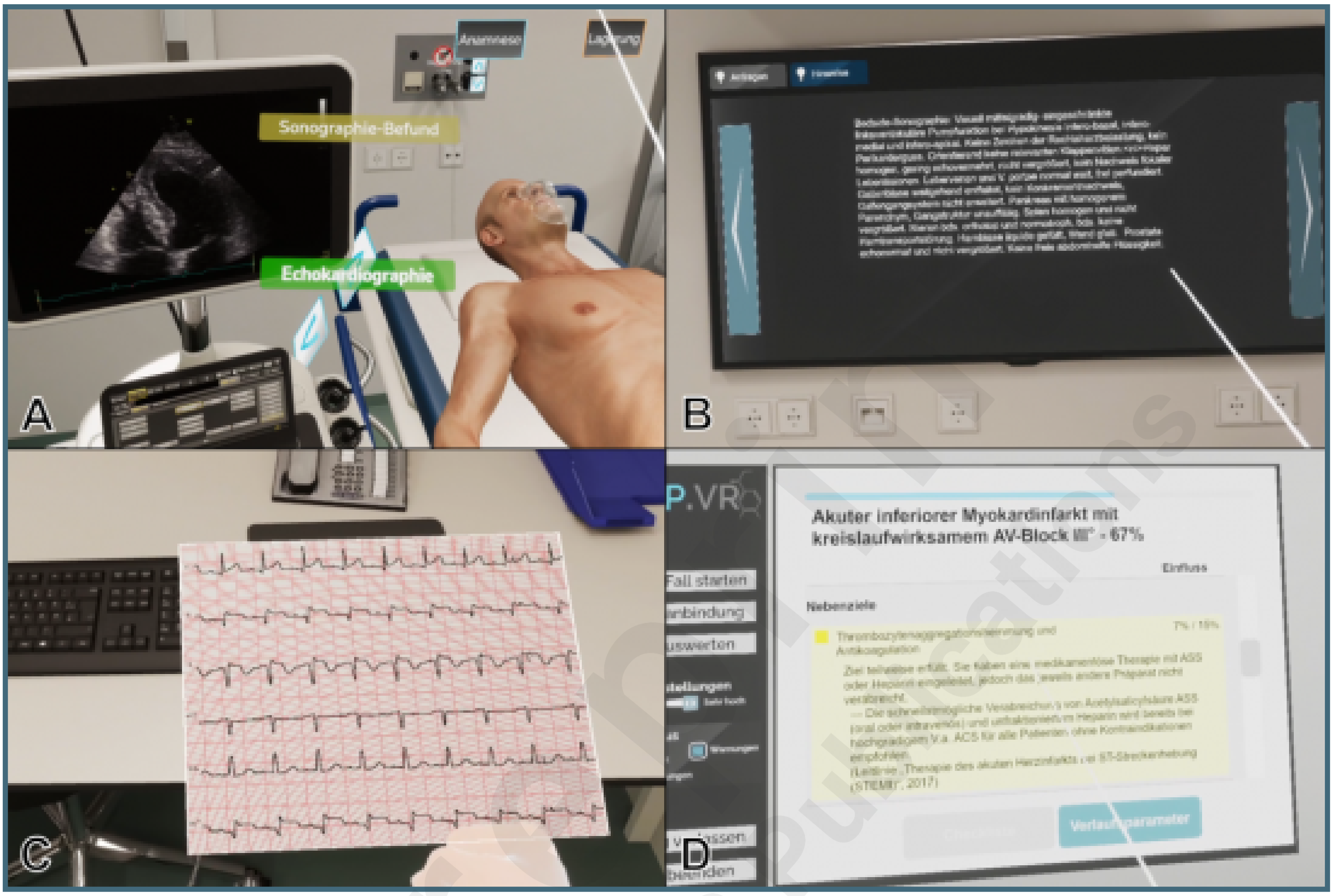
teaching method for resuscitation training in undergraduate first year medical students during COVID-19 pandemic: a randomised controlled trial. BMC Med Educ 2022;22(1):483. PMID:35733135
30. Agasthya N, Penfil S, Slamon N. Virtual Reality Simulation for Pediatric Airway Intubation Readiness Education. Cureus 2020;12(12):e12059. PMID:33447488
31. Gabbard T, Romanelli F. The Accuracy of Health Professions Students' Self-Assessments Compared to Objective Measures of Competence. Am J Pharm Educ 2021;85(4):8405. PMID:34283796
32. Keicher F, Backhaus J, König S, Mühling T. Virtual reality for assessing emergency medical competencies in junior doctors - a pilot study. Int J Emerg Med 2024;17(1):125. PMID:39333858
33. SITZMANN T, ELY K, BROWN KG, BAUER KN. Self-Assessment of Knowledge: A Cognitive Learning or Affective Measure? Academy of Management Learning & Education 2010;9(2):169-191 URL: http://www.jstor.org/stable/25682447 [accessed 2024-01-24].
34. Marín-Morales J, Higuera-Trujillo JL, Guixeres J, Llinares C, Alcañiz M, Valenza G. Heart rate variability analysis for the assessment of immersive emotional arousal using virtual reality: Comparing real and virtual scenarios. PLoS One 2021;16(7):e0254098. PMID:34197553
35. Kuhne C, Kecelioglu ED, Maltby S, Hood RJ, Knott B, Ditton E, Walker FR, Kluge MG. Direct comparison of virtual reality and 2D delivery on sense of presence, emotional and physiological outcome measures. Front. Virtual Real. 2023;4. doi:10.3389/frvir.2023.1211001
36. Prachyabrued M, Wattanadhirach D, Dudrow RB, Krairojananan N, Fuengfoo P. Toward Virtual Stress Inoculation Training of Prehospital Healthcare Personnel: A Stress-Inducing Environment Design and Investigation of an Emotional Connection Factor. In: 2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR): IEEE; uuuu-uuuu. ISBN:978-1-7281-1377-7. p. 671–679.
37. Tao K, Huang Y, Shen Y, Sun L. Automated Stress Recognition Using Supervised Learning Classifiers by Interactive Virtual Reality Scenes. IEEE Trans Neural Syst Rehabil Eng 2022;30:2060-2066. PMID:35857724
38. Klimek A, Mannheim I, Schouten G, Wouters EJM, Peeters MWH. Wearables measuring electrodermal activity to assess perceived stress in care: a scoping review. Acta Neuropsychiatr 2023:1-11. PMID:36960675
39. Phitayakorn R, Minehart RD, Hemingway MW, Pian-Smith MCM, Petrusa E. Relationship between physiologic and psychological measures of autonomic activation in operating room teams during a simulated airway emergency. Am J Surg 2015;209(1):86-92. PMID:25454964
40. Georgiou K, Larentzakis AV, Khamis NN, Alsuhaibani GI, Alaska YA, Giallafos EJ. Can Wearable Devices Accurately Measure Heart Rate Variability? A Systematic Review. Folia Med (Plovdiv) 2018;60(1):7-20. PMID:29668452
41. Ghazali DA, Darmian-Rafei I, Ragot S, Oriot D. Performance Under Stress Conditions During Multidisciplinary Team Immersive Pediatric Simulations. Pediatr Crit Care Med 2018;19(6):e270-e278. PMID:29432402
42. Brahim Bensouda, Romain Mandel, Abdelwaheb Mejri, Jean Lachapelle, Marie St-Hilaire, Nabeel Ali. Effect of an audience on trainee stress and performance during simulated neonatal intubation: a randomized crossover trial.
43. Stein C. The effect of clinical simulation assessment on stress and anxiety measures in emergency care students. Afr J Emerg Med 2020;10(1):35-39. PMID:32161710
44. Mills B, Carter O, Rudd C, Claxton L, O'Brien R. An experimental investigation into the extent social evaluation anxiety impairs performance in simulation-based learning

environments amongst final-year undergraduate nursing students. Nurse Educ Today 2016;45:9-15. PMID:27429397

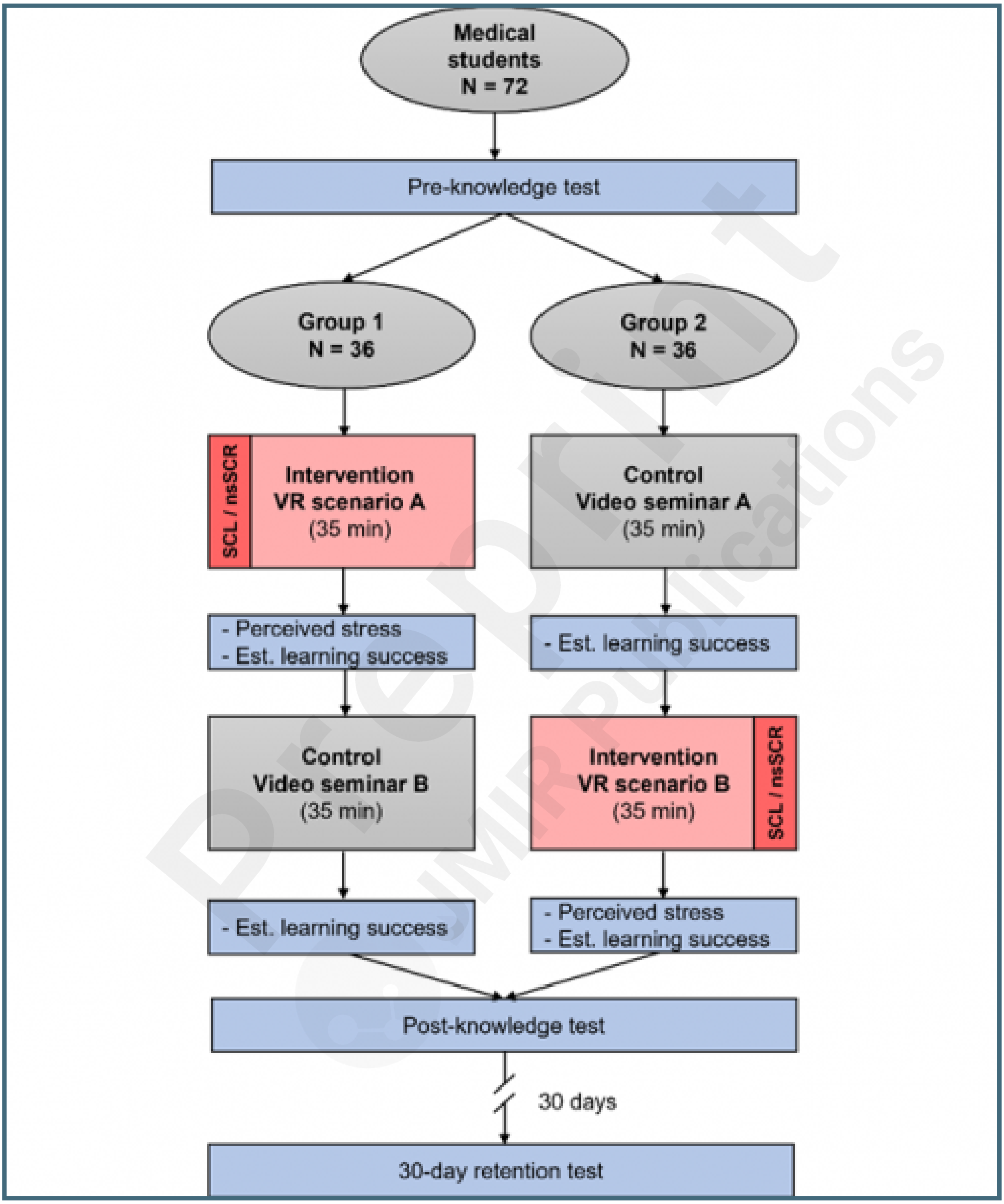
Supplementary Files

Figures

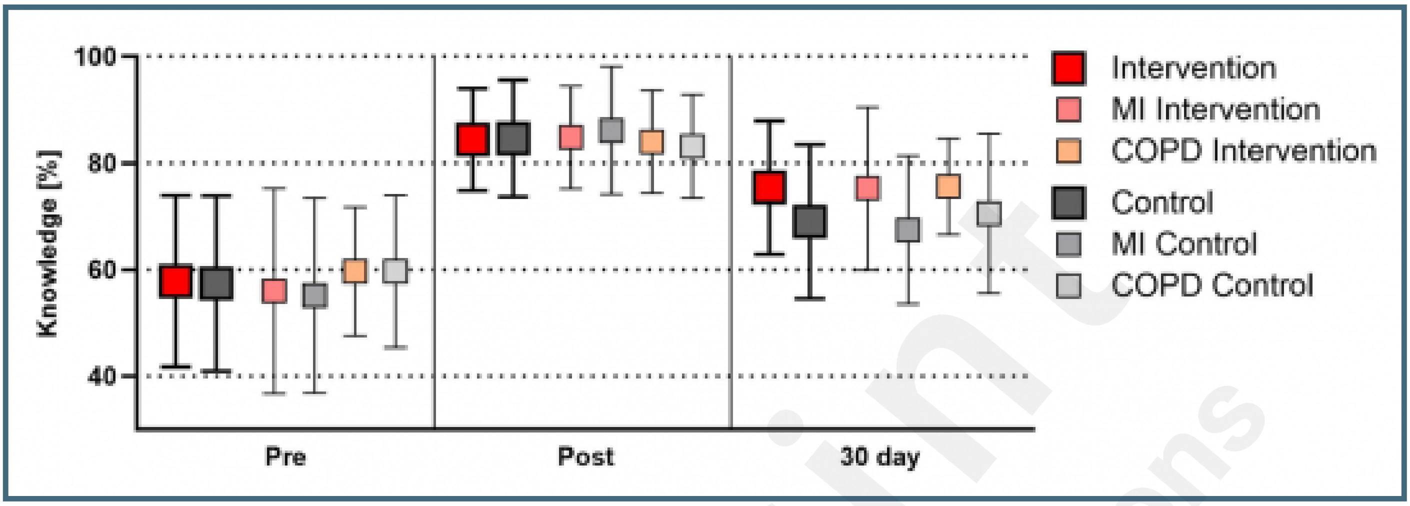
: Feedback components in the adapted software. (A) positive notifications (green) for correctly executed actions, (B) results of diagnostics in the virtual computer menu or (C) as direct output from medical devices (parameters calculated by dynamic physiology), and (D) final evaluation in checklist format.



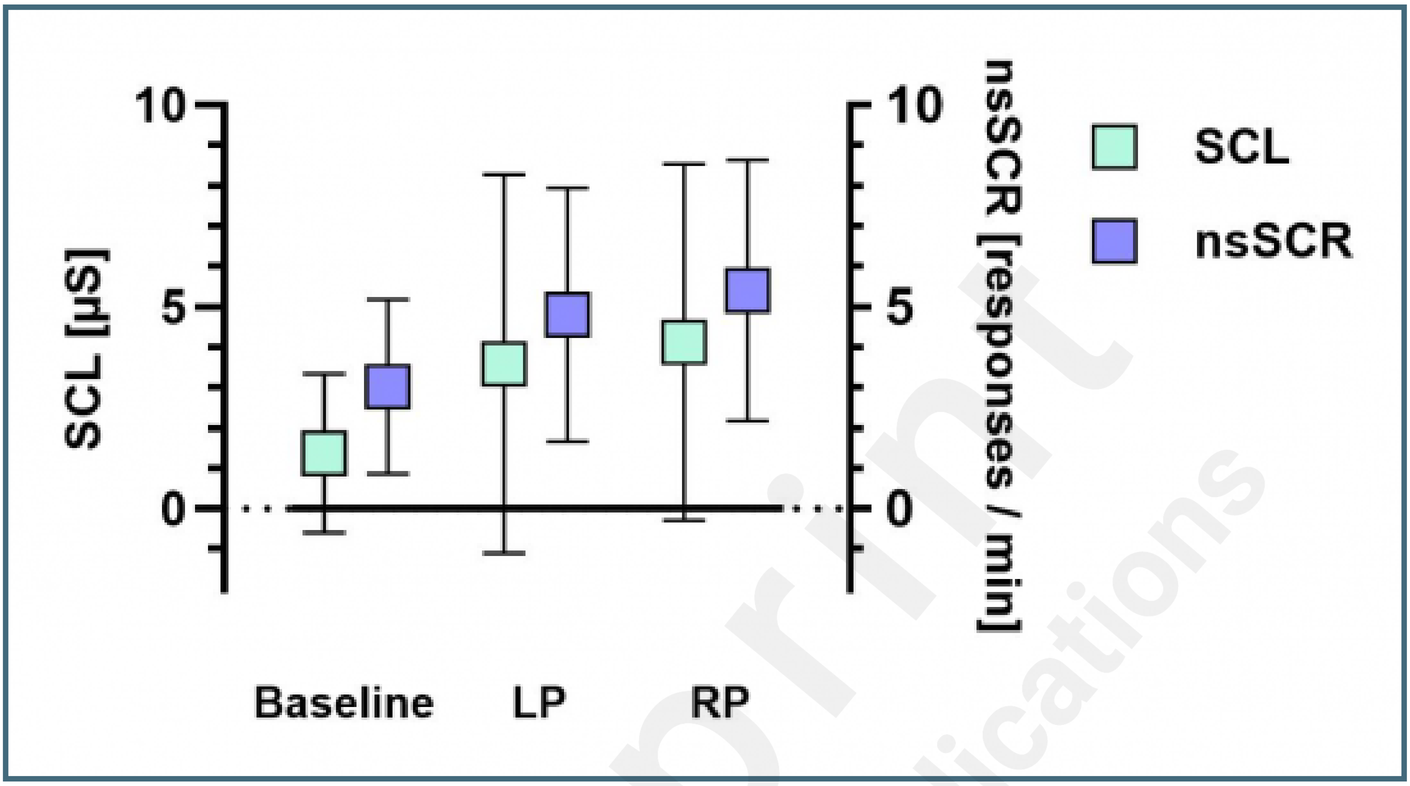
Study design and data collection process for knowledge tests and estimated learning success, as well as physiological stress markers (SCL and NSSCR) and perceived stress during intervention and control. Participants completed one scenario in VR and another as video seminar.



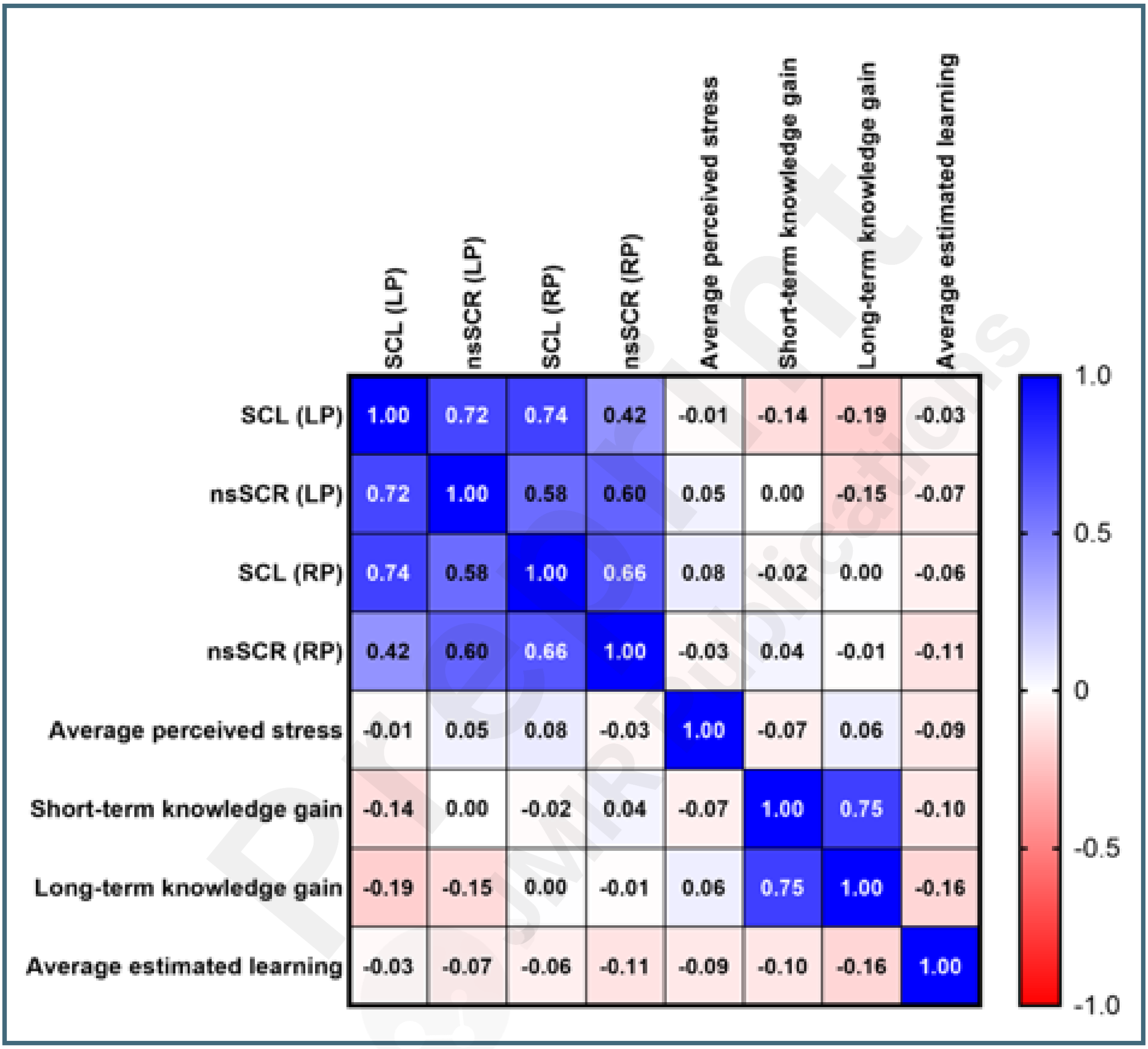
Knowledge test results as percentages of correct responses before (pre), immediately after (post) and 30 days after the intervention. Data are given in mean (box) with corresponding error bars (SD). The larger symbols and lines indicate overall results, while the smaller symbols detail the results for the individual scenarios.



The progression of physiological stress markers SCL [µS] and nsSCR [responses per minute] during the completion of the VR scenarios. Data are given in mean (box) with corresponding error bars (SD).



Correlation matrix between stress markers, knowledge gains, average perceived stress, and average estimated learning. The heatmap scale (right) indicates the color coding for positive (blue) and negative (red) correlations. The more intense the color, the higher the correlation coefficient. Corresponding significance values are listed in Supplement Table 1.



Multimedia Appendixes

Supplement Table 3: Rubric for open-ended questions (German) applied for knowledge tests.
URL: http://asset.jmir.pub/assets/c9b07a6665ab343bd6ada925f8efc812.docx

Feedback components in the adapted software (Video).
URL: http://asset.jmir.pub/assets/e5653d5e9719970ba60ddb6c40c00814.mp4