



Feasibility study on virtual reality-based training for skin cancer screening: Bridging the gap in dermatological education

Marvin Mergen¹ · Lisa Will² · Norbert Graf¹ · Marcel Meyerheim¹

Received: 25 April 2024 / Accepted: 22 August 2024 / Published online: 10 September 2024
© The Author(s) 2024

Abstract

Despite growing interest in the inclusion of virtual reality (VR) in medical education, its full potential for immersive and interactive training remains underutilized, particularly in dermatology. As part of a multidisciplinary project, we tested the feasibility of integrating a VR training scenario for a complete skin cancer screening into the curriculum with 58 medical students. The evaluation focused on simulation usability, cognitive task load, immersion, change in perceived competence, suitability, and cybersickness using established and adapted questionnaires. Participants reported rather high levels of system usability and immersion and medium cognitive task load. Self-assessed competence in performing a skin cancer screening increased significantly after the training simulation.

Prior skin cancer screening experience correlated positively with self-assessed competence and theoretical knowledge, which themselves were neither related to gender nor age. Age correlated negatively with perceived usability and immersion, enjoyment of learning during the simulation and assessed appropriateness of the simulation to learn a dermatological full-body examination while being positively correlated with perceived cybersickness. Gender was significantly associated with the need for technical help during the simulation and openness to new technologies.

As a blueprint of a feasibility evaluation, this study can contribute to further refinement of the presented and relatable VR applications in medical curricula.

Keywords Virtual reality · Artificial Intelligence · Medical education · Dermatology · Skin cancer screening

✉ Marvin Mergen
marvin.mergen@uks.eu

¹ Department of Pediatric Oncology and Hematology, Faculty of Medicine, Saarland University, Building 9, Kirrberger Strasse 100, Homburg 66421, Germany

² Clinic for Dermatology, Venereology and Allergology, Faculty of Medicine, Saarland University, Building 6, Kirrberger Strasse 100, Homburg 66421, Germany

1 Introduction

1.1 VR in medical education and dermatology

Virtual Reality (VR) has been gaining increasing prominence in medical education, revolutionizing the way medical professionals are trained. It offers immersive and interactive simulations that provide a safe environment for learners to practice and repeat various medical procedures (Mergen et al., 2024). VR's successful integration in surgical fields (Co et al., 2023; Ntakakis et al., 2023), emergency medicine (Abbas et al., 2023; Mühling et al., 2023; Walls et al., 2024) and preclinical anatomy (Abundez Toledo et al., 2024; García-Robles et al., 2024), has demonstrated its potential to enhance training and improve clinical skills.

Although VR technologies have garnered substantial interest among dermatologists, their integration in this field has lagged behind surgery (Gençoğlu, 2024). Nonetheless, VR technology seems to be one of the most promising futuristic trends in dermatological education (Das, 2024), particularly because of its visual nature. Recently published articles that outline best practices for research in VR and Augmented Reality (AR) in dermatology demonstrate the growing interest of the scientific community in this topic (Muralidharan et al., 2024; Obagi et al., 2020).

A review by Bonmarin et al. explores the status and potential future developments of AR and VR technologies in dermatology (Bonmarin et al., 2022) by investigating their use in education, diagnostics, and dermatologic surgery.

In education, AR and VR can enhance the learning experience by providing realistic 3D models of skin diseases, which promote and improve diagnostic skills and knowledge retention. In dermatologic surgery, AR and VR have been used to reduce patient anxiety and assist in complex procedures and in diagnostics. In a clinical setting Artificial Intelligence (AI) combined with VR has shown promise in detecting skin cancer.

During our literature research for similar training concepts in dermatology, we found one AR-based solution called mARble-dermatology, which offers users to learn about dermatological diseases with digital flashcards and virtual projections of lesions onto real skin examples. The authors did not find a significant benefit in terms of immediate learning success when learning with the mobile application, while a slight increase of long-term knowledge retention was examined (Noll et al., 2017).

Furthermore, we identified a study by Ranpariya et al. that examined the benefits of a VR memory palace for memorizing treatment protocols for atopic dermatitis (Ranpariya et al., 2022). The study involved a small cohort of medical students, dermatology residents and research fellows, who interacted with the VR environment for about 10 min per session. The results indicated high satisfaction among participants, with most finding the VR memory palace easy to set up and navigate. They also considered the mnemonics helpful for learning. The study suggests that the application could be a valuable addition to dermatology education, enhancing recall and engagement compared to traditional methods.

In February 2024, a study was published in which the same VR simulation application that we employed was tested at a different university following a distinct course design. They found that the use of VR in medical education was technically feasible and well-received by students, who viewed it as a valuable supplement to the curriculum. From the students' subjective perspective, the VR scenario contributed to a good learning outcome, particularly in enhancing their competence in performing skin cancer screenings. While preparation and follow-up are important for overall learning, the VR scenario was identified as having the most significant impact on students' perceived ability to conduct these screenings (Junga, Schmidle et al., 2024).

In addition to differences in course design, we employed a more detailed exploratory analysis and considered differences and correlations between subgroups. A comparison to their results is included in our discussion.

1.2 Didactic concepts

VR in education is underpinned by several key didactic concepts that enhance the learning experience. The following four didactic concepts collectively demonstrate how VR can improve medical education by providing variable and useful immersive scenarios.

Constructivist Learning posits that learners actively construct knowledge based on their existing understanding and past experiences (Mishra, 2023). VR supports this by immersing learners in interactive, context-rich environments where they can engage with virtual patients, enabling them to learn from their actions and outcomes and build upon their experiences made during the simulation.

The Situated Learning Theory emphasizes the effectiveness of learning when it takes place in the context in which the knowledge will be applied (Lave & Wenger, 1991). VR creates authentic learning environments that simulate real clinical situations, allowing students to gain practical experience. This immersion helps reinforce theoretical knowledge through practical application, thereby enhancing comprehension and retention.

The Experiential Learning Theory by Kolb describes learning as a cyclical process involving four stages: concrete experience, reflective observation, abstract conceptualization, and active experimentation (Kolb, 1984).

VR facilitates this process by enabling students to perform medical procedures, analyze their performance, apply theoretical concepts, and improve through repeated practice in a safe, controlled environment.

The Cognitive Load Theory addresses the load on working memory during learning, highlighting the need to balance intrinsic, extraneous, and germane cognitive loads (Plass et al., 2010).

VR can reduce extraneous cognitive load by presenting information visually and interactively, thus aiding comprehension. It enables learners to grasp complex medical concepts through immersive visualizations and interactive simulations, fostering a deeper understanding in a realistic and context-rich setting.

1.3 Motivation, use case and goal

Among the essential skills that medical practitioners should possess is the ability to conduct proficient skin cancer screenings. Early detection of skin cancer is crucial for patient outcomes and the corresponding skills should not be limited to dermatologists alone: general practitioners and other medical professionals must also be trained to identify suspicious skin lesions and refer patients to specialists for timely follow-up diagnosis and treatment. Unfortunately, skin cancer—especially undetected malignant melanoma (MM)—continues to claim numerous lives and can rapidly lead to metastases (Rashid et al., 2023).

Despite the significance of skin cancer screening, medical students still rarely have the opportunity to actually practice this skill themselves during their medical training. Patients with suspicious skin lesions are often directed to immediate surgical intervention, making it challenging for students to gain hands-on experience. Furthermore, the nature of a comprehensive full-body examination in a traditional medical curriculum is impeded by time constraints on wards and both patient and student discomfort, especially in terms of the essential genital examination. The same applies for training simulations with actors imitating patients.

In light of these challenges, our study aims to address this educational gap by testing the feasibility of integrating a self-developed VR simulation course for skin cancer screening into the medical curriculum at our medical faculty. The VR simulation allows medical students to independently examine a virtual, interactive patient in an immersive 3D environment of a dermatological practice. During the simulation, students are required to take the patient's health history, examine, document and evaluate skin lesions, making decisions on which lesions require further diagnostics, treatment or observation.

Through this research project we seek to explore and demonstrate the potential of VR as an innovative and effective tool in training medical students for skin cancer screening. By providing a safe and realistic environment for learning, we aim to empower future medical professionals with the necessary skills to identify and respond to skin cancer in its early stages, ultimately contributing to improved patient outcomes and reduced mortality rates associated with undetected skin cancer.

Within this scope, the following research questions (RQ) are addressed in this study. While the overall aim of the study is the assessment of feasibility of this new learning method, the RQs were formulated to investigate whether subgroups of participants (e.g., regarding gender, age, prior VR or skin cancer screening experience, self-efficacy) differed from one another and to identify possible correlations. The objective was to determine which factors might influence the assessment of the VR training.

RQ 1: How do medical students evaluate the usability of the VR simulation?

Since usability in terms of user-friendliness could influence the overall assessment of the suitability of the presented VR simulation for teaching, this aspect was captured as independent variable.

RQ 2: What is the cognitive task load experienced by medical students during skin cancer screening training in the VR simulation?

Cognitive task load during learning should be neither too low nor too stressful in training situations to ensure effective learning (Nelson & Elias, 2023).

RQ 3: To what extent do the medical students feel present and immersed during the VR simulation?

Immersion and presence in training simulations enhance realism, allowing users to engage deeply with the environment, enabling skill development by closely mimicking real-world scenarios.

RQ 4: To what extent does the VR course change medical students' self-perceived competence in skin cancer screening?

Training using the VR simulation is only worthwhile if an increase in competence is realized and perceived afterwards. While the validity of self-assessments is currently under discussion, there are indeed studies showing that students' self-perceptions correlate with external evaluations (Spoto-Cannons et al., 2019).

RQ 5: How suitable is the VR course for training a skin cancer screening?

Suitability depends on whether simulation tasks are feasible and the training is perceived as meaningful or beneficial according to the students' opinion as end users.

RQ 6: Do participants experience symptoms of Cybersickness during the VR simulation?

Evaluating cybersickness is crucial to ensure user comfort and safety of the virtual experience. To account for equal opportunities to use the simulation, these negative effects must be limited.

2 Methods

The methodology of this feasibility study involved the planning and integration of a VR simulation as part of the BMBF-educational project “medical tr.AI.ning” (Mergen et al., 2023a) involving the Saarland University, the University of Münster, the University of Fine Art Saar (HBKsaar), the German Research Center for Artificial Intelligence (DFKI), and the University of Applied Sciences Münster (FHMS), initiated in December 2021.

This VR simulation served as the inaugural application developed within the scope of the medical tr.AI.ning - project.

2.1 Organization

2.1.1 Context

The study took place at the Medical Faculty of Saarland University, Campus Homburg and targeted participants from the fifth or sixth clinical semester, depending on their educational schedule. The VR simulation was seamlessly incorporated into the mandatory curriculum for the summer semester of 2023 by integrating it into the dermatology practical course.

This practical course is divided into two subsequent blocks, Block A (April to May 2023) and Block B (June to July 2023). The VR simulation was only integrated into Block B. The weekly practical courses are split up into two sessions of 90 min.

The simulation session took place in the skills lab at Campus Homburg, where groups of six students participating in each session were briefed on the procedure and introduced into dermatological skin cancer screening by a qualified dermatologist before starting the VR simulation. Participants were then divided into three sub-groups of two individuals and assigned to three separate rooms.

Ethical approval was given by the 'Ärztchamber des Saarlandes', Germany (No. 244/21). Moreover, we obtained written informed consent from all participants with information on General Data Protection Regulation (GDPR)-compliant data storage and processing. Participation in the study was voluntary. Students who did not want to participate could still take part in the course and the VR simulation without completing the study questionnaires.

2.1.2 Course scheme

The whole course scheme is presented in Fig. 1. Each VR session involved one participant performing the skin cancer screening in VR while being observed by another student.

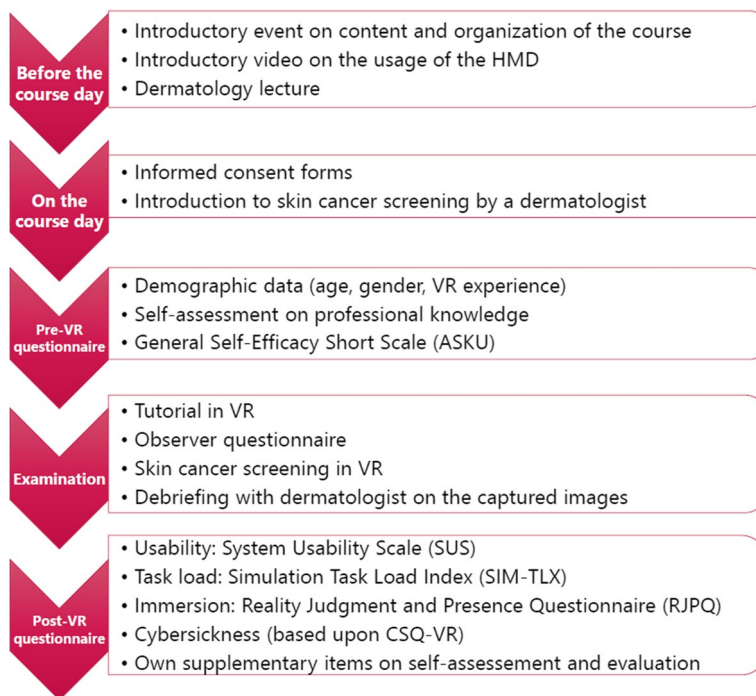


Fig. 1 Time schedule and content of the course

Before starting the simulation, the students were asked to complete a pre-VR questionnaire on demographic information as well as prior experience with both VR and skin cancer screening.

A tutorial before the actual skin cancer screening served to familiarize participants with the VR controls and interactions. After completion, participants started the VR simulation standing in the waiting area of a virtual dermatological practice where the patient was waiting for them in one of the examination rooms. The students were free to choose how they carried out the complete examination routine. A virtual dermatoscope could be used to take a limited number of photos of suspicious skin lesions.

During the simulation, the observing student followed the examiner's steps in VR on a computer screen and completed a designated questionnaire that was comparable to an Objective Structured Clinical Examination (OSCE) (Harden & Gleeson, 1979). This questionnaire served as a checklist to comprehensively assess the completeness and correctness of the skin cancer screening.

After completing the examination, the roles were swapped, and the observing student completed the tutorial and the subsequent skin cancer screening.

After concluding the VR simulations, participants discussed their photos of skin lesions with the dermatologist. Subsequently, they completed the post-VR questionnaire as described in detail in Sect. 2.5.

Throughout the course and study, participants were continuously supported by at least one medical didactic expert, a dermatologist, and a minimum of two technical support assistants.

2.1.3 Hardware and software

For the VR simulation, the researchers utilized the Valve Index VR-Kits, in combination with high-end laptops (32 GB RAM, 11th Gen Intel(R) Core(TM) i7-11800H, Nvidia GeForce(R) RTX(TM) 3080 16 GB). The software platform employed was the latest regularly updated SteamVR version, and the simulation was programmed using Unity (Version 2021.3.15f1).

Questionnaires were completed by all students using an iPad Pro 11" and the online survey tool 'LimeSurvey'.

2.2 Learning goals

With this VR course, students should learn to perform a complete and structured skin cancer screening with a whole-body skin examination and to evaluate skin lesions in terms of their possible malignancy using the ABCDE rule (Friedman et al., 1985).

This rule, introduced by Robert Friedman in 1985, outlines that asymmetry (A), border irregularity (B), color variegation (C), and a diameter generally greater than 6 mm (D) are key features that aid in the early diagnosis of malignant melanoma. In 2004, the rule was expanded to include "evolution" (E), which accounts for changes in size, shape, color, surface features, or symptoms, recognizing the dynamic nature of melanoma malignancy (Duarte et al., 2021).

In the currently expanded German national learning catalog for medical school (NKLM 2.0) (Medizinischer Fakultätentag, 2021) these goals are associated with clinical-practical skills, i.e.

- VIII.7.-02: The graduate will apply examination skills in a manner that is situationally adequate, hygienically sound, and respectful to patients.
- VIII.7–02.1.15: They can perform a clinical examination of the skin and skin appendages as part of the basic whole-body examination.
- Entrustable Professional Activity (EPA) 1, i.e., taking medical history and performing a clinical examination (ten Cate, 2005; ten Cate, 2013).

2.3 Simulation task

During the simulation, the participants were instructed to act in the role of a medical student in a dermatological practice, taking over a skin cancer screening from their supervising dermatologist who was unable to attend due to an urgent call. The participants were asked to perform a complete skin cancer screening and to document any detected abnormalities on the virtual patient's skin with photos using a digital dermatoscope for a follow-up discussion. To focus on the detection of the actual abnormalities, the number of possible savable pictures was limited to five.

The location and presence of inconspicuous nevi and up to two malignant melanomas (see Fig. 2) varied between simulation runs and was randomly chosen

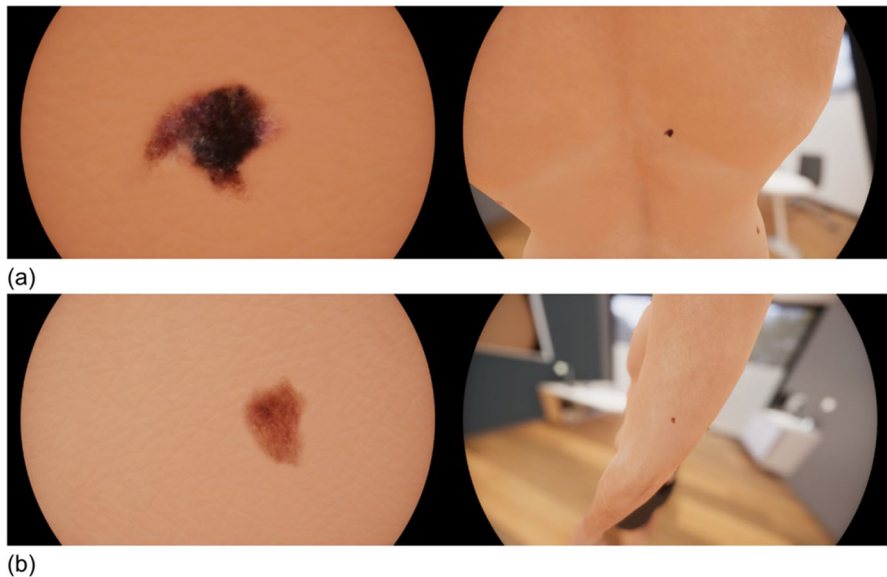


Fig. 2 Examples for a malignant melanoma (a) and nevus (b) in the VR application

from limited sets of pre-defined distributions grouped into 5 cases at the beginning of the simulation to prevent any memory effect among students:

- Case 1: 28 inconspicuous nevi, 1 malignant melanoma (head scalp)
- Case 2: 28 inconspicuous nevi, 2 malignant melanomas (back center and left hip)
- Case 3: 28 inconspicuous nevi, 1 malignant melanoma (right foot sole)
- Case 4: 28 inconspicuous nevi, 2 malignant melanomas (left gluteal cheek and left armpit)
- Case 5: 28 inconspicuous nevi, 1 malignant melanoma (head neckline).

The whole course design and application development phase with skin lesions included qualitative feedback iterations with local dermatologists.

2.4 Questionnaires

The list of all questionnaires and their items can be found in Supplement 1.

Before the simulation started, we asked the students in the pre-VR questionnaire about demographic data, prior experience with VR and skin cancer screening, theoretical knowledge about skin diseases, estimated suitability of VR as learning method, their openness to embrace novel technologies, as well as their general self-efficacy using the “Allgemeine Selbstwirksamkeit Kurzskala” (ASKU) questionnaire (Beierlein et al., 2012). The ASKU scale consists of three 5-Point Likert items focusing on the reliance on one’s own abilities in difficult situations, independence in problem-solving and the management of challenging tasks. Self-assessed agreement with each item is rated on a scale from 1 (*does not apply at all*) to 5 (*applies completely*). To obtain the ASKU score, the answers to the individual items of a test person are averaged such that the resulting score varies between 1 (low self-efficacy) and 5 (high self-efficacy) according to the authors’ guidelines.

Throughout the simulation, the student that was not currently acting in the VR environment was able to monitor the examiner’s performance in VR on a screen and filled out the observer questionnaire resembling an OSCE. This questionnaire acted as a checklist to assess if all the elements of the dermatological full-body examination were performed, including considerations like patient privacy and medical history. Moreover, the sequence of the participant’s steps in the simulation was recorded.

Upon finishing the simulation, participants proceeded to complete the post-VR questionnaire (see Fig. 1). Its components are described in the following:

The System Usability Scale (SUS) (Brooke, 1996) was used to assess the application’s usability, which can broadly be defined as “general quality of the appropriateness to a purpose of any particular artefact”. Based on its 10 Likert items, we calculated the resulting SUS score according to the author’s guidelines which spans from 0 (low usability) to 100 (high usability).

To assess cognitive load during the VR simulation, the Simulation Task Load Index (SIM-TLX) (Harris et al., 2020) was applied which is based on 9 items rated on a scale from 0 (*very low*) to 10 (*very high*). The final scores result from the sum

of all ratings, while the last item's rating is reversed, and has a possible range from 0 (low task load) to 90 (high task load).

To measure immersion in the simulation, we used the Reality Judgment and Presence Questionnaire (RJPQ) (Baños et al., 2000). Its 18 items are grouped into three factors: attention/absorption, reality judgment and external/internal correspondence.

Each item is rated on a scale from 0 (*not at all*) to 10 (*absolutely*). The final scores for the three factors resulted from the averaged ratings on the respective items. Having developed the virtual skin lesions by ourselves (as work of Leon Pielage, University of Münster) using generative AI-based algorithms, we added one item on their perceived realism using the same scale.

The term “Cybersickness” comprises symptoms such as nausea, headache or dizziness due to a incongruence of vestibular (inner ear) and visual information which can be experienced by VR users (Rebenitsch & Owen, 2016). We used a shortened version of the Cybersickness in Virtual Reality (CSQ-VR) questionnaire (Kourtesis et al., 2023) to assess to which degree participants experienced any related symptoms by rating their occurrence with one 7-Point Likert item (1 = *not at all* to 7 = *extremely strong*) and naming the symptoms experienced.

The last ten 5-Point Likert items of the post-VR questionnaire focused on course evaluation and subjective learning outcomes (1 = *do not agree at all* to 5 = *fully agree*). The selection of questionnaires and additional items were carefully aggregated by the co-authors in collaboration with the medical training-consortium and partially derived from the Training Evaluation Inventory (TEI) (Ritzmann et al., 2014) and the Learning Experience Questionnaire (LEQ) (Gruber & Strömsten, 2021). At the end of the post-VR questionnaire, participants were able to provide additional feedback through free text answers which were evaluated in terms of content, i.e., positive, neutral or negative statements as well as recommendations and translated from German into English.

2.5 Statistics

IBM SPSS Statistics Version 28.0.1.1 was used for descriptive statistics to illustrate the composition of the study cohort and present the results regarding the defined research questions. The application was also used for all other statistical analyses investigating differences between subgroups within the study cohort and correlations between scores. A significance level of 0.05 was used for all statistical tests (two-tailed).

Chi-Square-Test of Independence and Homogeneity were used to analyze associations between nominal variables. In case of expected frequencies below 5, Fisher's exact test or Fisher-Freeman-Halton exact test (FFH) were used, depending on the dimensions of the contingency table. In case of significant associations between nominal variables, the Phi Coefficient ϕ (in case of two dichotomous variables) or Cramér's V (if one of the two nominal variables had more than two categories) were computed to determine effect sizes and interpreted based on their absolute values according to literature (small ≥ 0.1 , medium ≥ 0.3 , large ≥ 0.5) (Cohen, 1988). Spearman-rho correlation r_s was computed to check for monotonic associations between ordinal variables and its effect size interpreted based on its absolute value according to literature (weak ≥ 0.1 , moderate ≥ 0.3 , strong ≥ 0.5) (Cohen, 1988).

To check for significant differences between groups considering ordinal variables (such as responses to single Likert items) non-parametric tests were used depending on the number of groups, i.e., the Mann–Whitney-U Test in case of two groups or the Kruskal–Wallis-Test in case of three or more groups. In case of significant differences between groups, effect size r was computed (Fritz et al., 2012) and interpreted based on its absolute value according to literature (small ≥ 0.1 , medium ≥ 0.3 , large ≥ 0.5) (Cohen, 1988). The Wilcoxon-Signed-Rank-Test was used to check for significant differences between repeated measures (paired samples) of ordinal variables.

The distribution of metrical variables, such as aggregated Likert scale scores derived from several Likert items, were tested for both normality using the one-sample Shapiro–Wilk test and variance homogeneity using the Levene’s test. If both conditions were met, the parametric unpaired two-sample t-test was applied to check for significant differences between groups. Otherwise, non-parametric tests were used (Mann–Whitney U test or Kruskal–Wallis test). To check for significant differences between age groups considering ordinal and metric variables, few age range groups were pooled, i.e., 26 to 30 year-olds and 31 to 33 year-olds, to meet the Kruskal–Wallis test’s requirement to have group sizes of at least 5 samples each.

Metric variables with normal distribution are presented as mean value with standard deviation, non-normal variables are reported as median value with interquartile range (IQR). The distributions of normal and non-normal variables are visualized using boxplots for the sake of uniformity.

3 Results

3.1 Pre-VR-questionnaires

3.1.1 Study population

In total, 58 medical students completed the simulation and the respective questionnaires with 35 female (60%) and 23 male participants (40%), a gender distribution which is in line with both the cohort of a previous survey at our faculty (Mergen et al., 2023b), $\chi^2(1, N=58)=0.54$, $p=0.461$, and the current distribution among German medical students (*Studierende der Medizin nach Geschlecht bis 2022/2023*, 2023), $\chi^2(1, N=58)=0.41$, $p=0.523$. The participants’ age ranged from 22 to 34 years ($Mdn=24$, $IQR=23–26$) with more than half belonging to the group of 23-year-olds ($n=19$, 33%) and 24-year-olds ($n=15$, 26%). No significant difference was detected between genders in terms of age distribution, $U=350$, $z=-0.86$, $p=0.390$. Most students were right-handed ($n=54$, 93%), while 3 were left-handed and 1 person both-handed. This approximately resembles the ratio of right- and left-handed people in Germany (Papadatou-Pastou et al., 2020).

3.1.2 Prior VR and skin cancer screening experience

Assessing prior VR experience, two thirds ($n=38$, 66%) claimed having used a head mounted display (HMD) before our study at least once or rarely while 20 students

(34%) reported they have never used HMDs. None of the participants claimed having used HMDs on a regular basis.

60% ($n=35$) have already attended an educational format with VR. One participant has visited such a course without wearing an HMD, while the opposite case holds true for 4 students: they have used an HMD, but outside of an educational setting. A significant association of strong effect was detected between prior VR-experience and prior participation in an educational format with VR, $\chi^2(1, N=58)=39.07, p<0.001, \phi=0.82$.

Considering prior skin cancer screening experience, most students stated to only possess theoretical background knowledge ($n=35, 60\%$), followed by having no knowledge at all ($n=11, 19\%$) or having only observed an examination ($n=9, 16\%$). Three participants have performed the examination under supervision ($n=3, 5\%$), while none has performed a skin cancer screening without supervision.

3.1.3 Self-assessed competence and knowledge

Prior to engaging in the VR skin cancer screening simulation, an analysis of the initial self-assessment responses from the cohort of 58 participants revealed the following insights:

A relevant proportion of participants lacked confidence in their competency to perform the skin cancer screening, with 28 participants (48%) expressing their uncertainty. Meanwhile, 15 participants (26%) adopted a neutral stance, and 5 participants (9%) identified themselves as having limited competence in this area. Regarding self-perceived theoretical knowledge of skin diseases, the participants' responses varied: 14 participants (24%) agreed that they possessed sufficient knowledge, while 19 participants (33%) remained neutral, and a substantial proportion of 25 participants (43%) disagreed. A significant moderate positive correlation was detected between prior skin cancer screening experience and both the self-assessed competence, $r_s(56)=0.40, p=0.002$, and theoretical knowledge, $r_s(56)=0.36, p=0.005$.

Prior to the VR simulation, most participants ($n=40, 69\%$) perceived the VR format as a valuable tool for enhancing their competences. However, 11 participants (19%) maintained neutral opinions, while 7 participants (12%) did not find the VR format conducive to their learning needs.

However, a substantial number of participants ($n=43, 74\%$) agreed to enjoy trying out and using new technological advancements. A smaller fraction ($n=7, 12\%$) did not share this attitude. 8 participants (14%) were indifferent to this question. A significant difference of medium effect was detected between male (mean rank of 37.13) and female participants (mean rank of 24.49), $U=227, z=-2.98, p=0.003, r_s(56)=0.40$. Thus, male participants were more likely to assess their openness to new technologies higher.

3.1.4 General self-efficacy (ASKU)

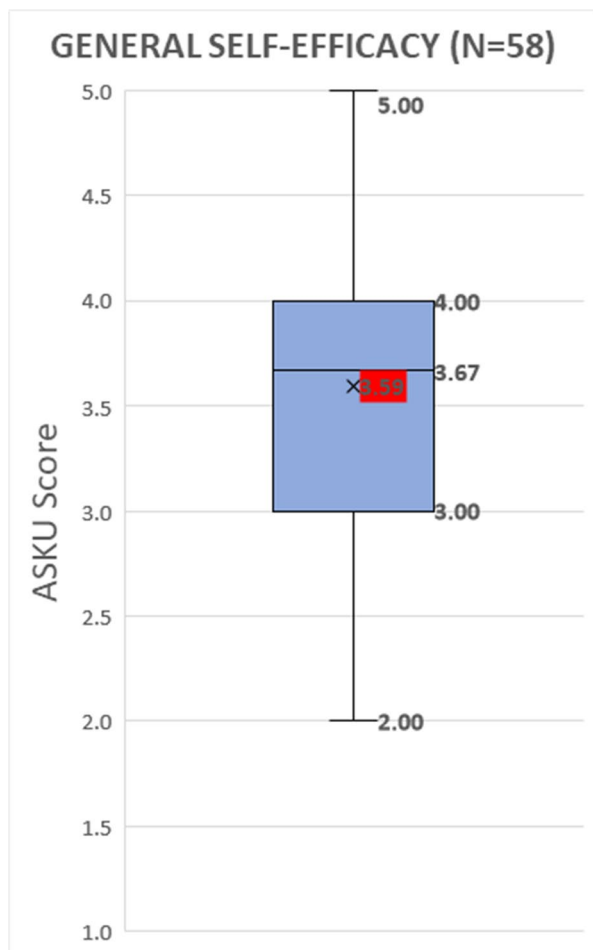
Regarding the participants' assessed self-efficacy through ASKU, more than half ($n=31, 53\%$) believed in their ability to proficiently manage demanding and intricate tasks. 25 participants (43%) partially agreed with this statement, while a minor group of 2 participants (3%) expressed low agreement.

In terms of their confidence in their ability to overcome challenges, 37 participants (64%) agreed that they could effectively navigate various problems on their own. Concurrently, 19 participants (33%) held a partially affirmative perspective on this matter, while a minor proportion of 2 participants (3%) indicated weak confidence.

Considering their reliance on their abilities during demanding situations, 33 participants (57%) were convinced by their skills, while 23 participants (40%) reported a certain level of trust in their abilities, whereas 2 participants (3%) expressed low certainty under such circumstances. None of the participants responded with “do not agree” to any of the items.

Based upon the average score across the three items, we calculated the ASKU score with a median score of 3.67 (IQR=3.00–4.00) (Fig. 3) with no significant differences neither regarding gender, $U=305.5$, $z=-1.59$, $p=0.111$, nor age groups, $\chi^2(5, N=58)=5.39$, $p=0.371$. Nonetheless, we found strong positive significant

Fig. 3 Box plot for general self-efficacy represented by ASKU score as assessed by the study participants



correlations between the ASKU score and its three items accounting for high internal consistency, all $r_s(56) \geq 0.86$, $p < 0.001$.

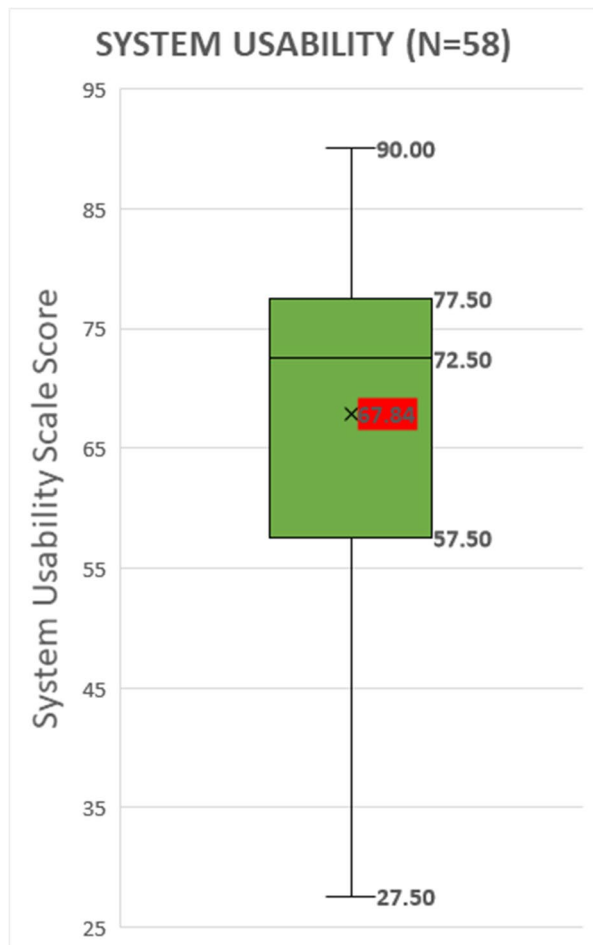
3.2 Post-VR questionnaires

3.2.1 System Usability Scale (SUS) (RQ 1)

The average SUS Score among our cohort was 67.8 ($SD = 14.0$) with a median score of 72.5 (IQR = 57.5–77.5) (see Fig. 4) and without significant differences between neither male and female participants, $U = 295.5$, $z = -1.18$, $p = 0.726$, nor prior VR experience, $U = 427$, $z = 0.80$, $p = 0.224$.

According to literature (Bangor et al., 2009), this score can be considered as D grade, which means marginally high acceptable or respectively at the higher range of the adjective ratings “OK” close to “GOOD” (see Fig. 5). A significant moderate

Fig. 4 Box plot for system usability represented by SUS as assessed by the study participants



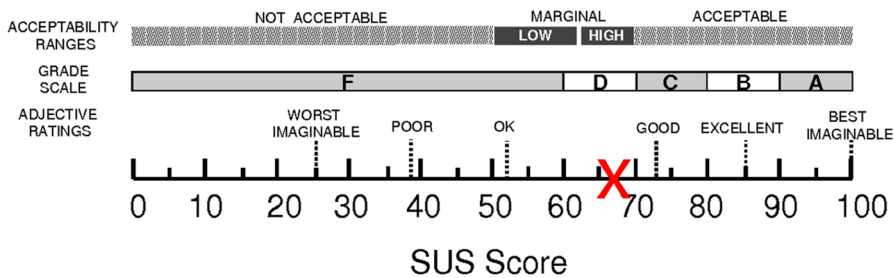


Fig. 5 The average SUS score of 67.8 ($SD=14.0$) as assessed by the participants translates adjective ratings between “OK” and close to “GOOD” according to literature (Bangor et al., 2009)

negative correlation was found regarding age, $r_s(56)=-0.42$, $p<0.001$. With higher age, the SUS decreased. Another significant weak negative correlation revealed between the SUS’s item on how safe participants felt using the system and previous skin cancer experience, $r_s(56)=-0.28$, $p=0.033$. Female participants (mean rank of 33.09) were more likely than men (mean rank of 24.04) to agree to the item that they would need the help of a tech savvy person to use the system with a significant difference of small effect, $U=528$, $z=2.09$, $p=0.04$, $r=0.27$. Detailed results considering more SUS items can be found in Supplement 1.

3.2.2 Cognitive load (SIM-TLX) (RQ 2)

The SIM-TLX score could be assumed to be normally distributed among our cohort as assessed with the Shapiro–Wilk Test ($W=0.985$, $p=0.697$). The mean SIM-TLX score of 26.19 ($SD=10.5$) indicated a medium task load (see Fig. 6). No significant differences were observed between genders, $t(56)=-0.24$, $p=0.813$, between age groups, $\chi^2(5, N=58)=6.03$, $p=0.304$, or prior VR-experience, $t(56)=1.83$, $p=0.072$.

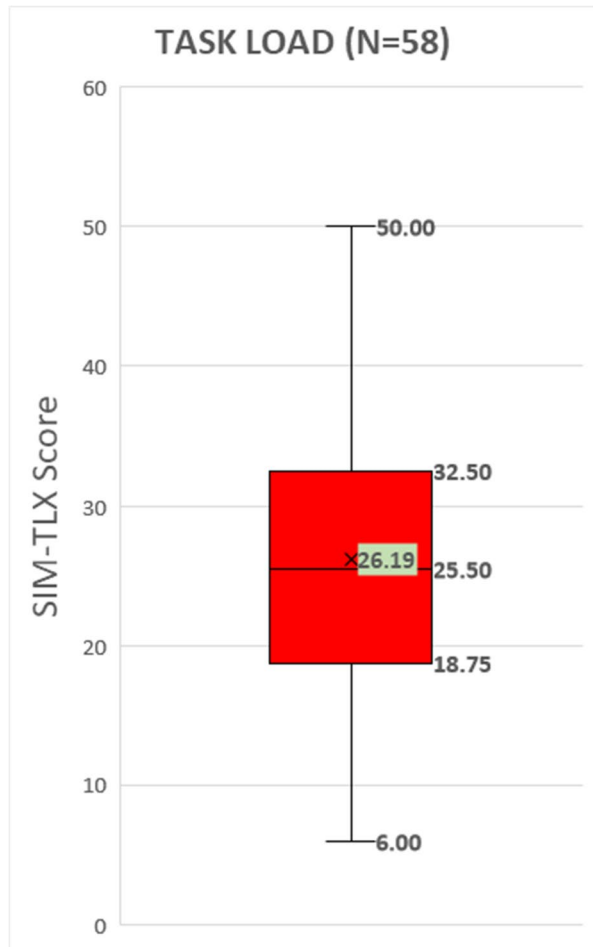
The SIM-TLX score correlated strongly negatively with the SUS score, $r_s(56)=-0.69$, $p<0.001$, as well as moderately negatively with the SUS’s item on feeling confident using the system, $r_s(56)=-0.31$, $p=0.019$, and weakly negatively with the SUS’s item on the openness to new technologies, $r_s(56)=-0.27$, $p=0.040$.

3.2.3 Immersion (RJPQ) (RQ 3)

Focusing on the attention/absorption score, the surveyed students ($n=58$) answered with a median score of 5.75 (IQR=4.69–7.06) (see Fig. 7A).

The reality judgment score revealed similar results with a median score of 4.94 (IQR=4.13–6.13) (see Fig. 7B). A slightly higher index was found examining the internal/external correspondence score with a median score of 6.42 (IQR=4.96–7.67) (see Fig. 7C). Regarding the latter score, a moderate negative correlation was detected considering age, $r_s(56)=-0.41$, $p=0.001$, and there was a significant difference of medium effect between students with (mean rank of

Fig. 6 Box plot for perceived task load represented by SIM-TLX score as assessed by the study participants



21.63) and without (mean rank of 33.64) prior VR-experience, $U=537.5$, $z=2.58$, $p=0.010$, $r=0.34$.

The realism of the self-developed skin lesions revealed a median score of 8 (IQR=6–9) with a significant difference of small effect between males (mean rank of 24.17) and females (mean rank of 33), $U=525$, $z=1.980$, $p=0.048$, $r=0.260$.

3.2.4 Subjective learning outcomes after the VR simulation (RQ 4) and perceived suitability (RQ 5)

53 participants (91%) agreed to having fun learning with our simulation (see Fig. 8) and 54 (93%) considered the acquired competences useful for performing a skin cancer screening. Nevertheless, the results' variance of the self-assessed competence was higher, with 25 (43%) stating that they feel confident, 16 (28%) neither agreeing nor disagreeing and 17 (29%) perceiving themselves with a low competence. Still,

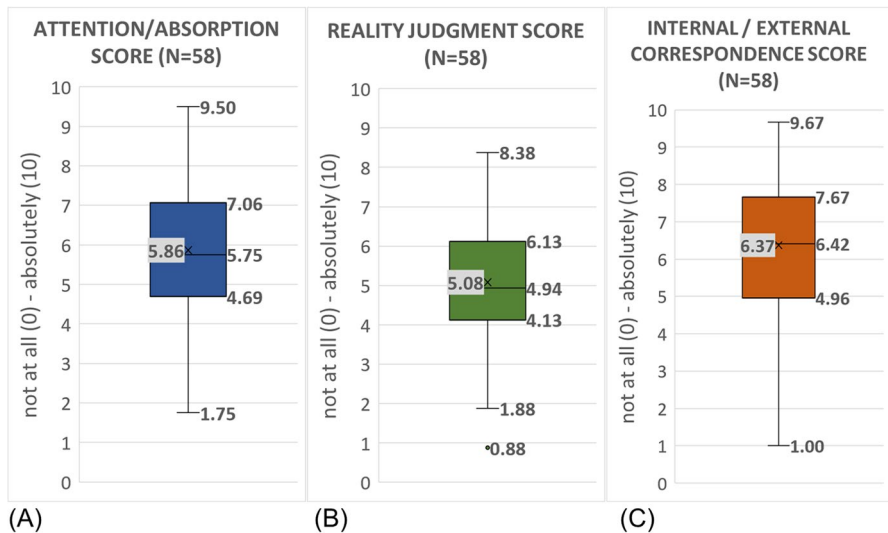


Fig. 7 Box plots for the scores representing attention/absorption (A), reality judgment (B) and internal/external correspondence (C) as assessed by the participants through RJPQ

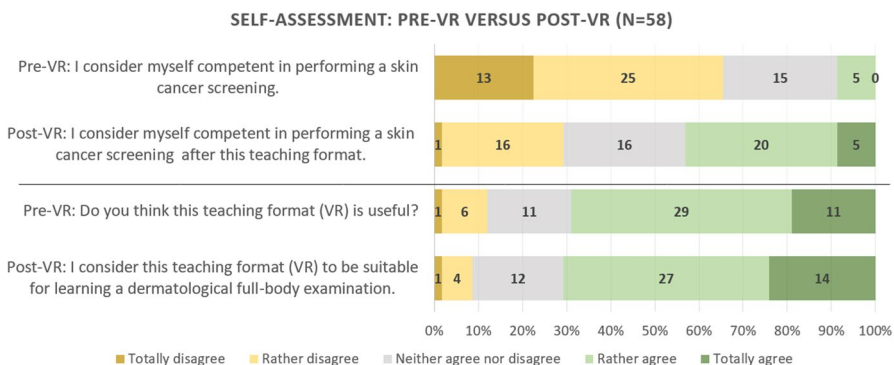


Fig. 8 Bar charts representing self-assessments and perceived benefit by the participants pre- and post-VR

51 (88%) students gave positive feedback about the suitability of the VR simulation for future skin cancer screenings and 41 (71%) consider VR technology as appropriate for learning a dermatological whole-body examination. Considering age, weak negative correlations were detected with both—enjoyment of learning with the simulation, $r_s(56) = -0.27$, $p = 0.043$, and assessment of the appropriateness of the simulation to learn a full-body dermatological examination, $r_s(56) = -0.30$, $p = 0.023$.

Almost all participants were satisfied with the amount of help received if needed during the simulation ($n = 55$, 95%) and most evaluated the course as engagingly challenging ($n = 46$, 79%). Furthermore, the possibility to test procedures that are

required in future clinical practice was positively rated by 49 students (85%) and 39 (67%) expect long-term knowledge retention. Overall, 46 students (79%) considered the time invested in the VR simulation as useful.

Comparing pre-post-test scores, we found a significant difference of large effect regarding self-assessed competence in performing a skin cancer screening before (median score of 2 (IQR 2–3)) versus after the simulation (median score of 3 (IQR 2–4)), $z=0.75$, $p<0.001$, $r=0.68$.

3.2.5 Observer questionnaire (RQ 5)

During our simulation, participants engaged in pairs of two, with one student starting with the examination in VR and the other as an observer, resulting in 30 initial examiners and 28 initial observers. In case of uneven group sizes, one of the technical assistants took over the role of the observer and completed the corresponding questionnaire.

As specified above, the pre-defined five different cases were assigned randomly to the participants, i.e., case 1 ($n=9$, 16%), case 2 ($n=14$, 24%), case 3 ($n=13$, 22%), case 4 ($n=11$, 19%) and case 5 ($n=11$, 19%). To prevent any bias from unbalanced case distribution, we checked for any significant dependence between case distribution and subgroups within our cohort which could not be statistically confirmed. Thus, the case distribution could be assumed to be independent of gender, FFH=5.39, $p=0.251$, age, $\chi^2(4, N=58)=3.31$, $p=0.508$, prior VR experience, FFH=3.10, $p=0.581$, prior skin cancer screening experience $\chi^2(4, N=58)=0.15$, $p=0.997$, and order of the observer, FFH=2.00, $p=0.785$.

The most frequently missed examination steps included neglecting to close the sliding door of the examination room to protect the patient's privacy ($n=23$, 47%) and omitting hand disinfection/wearing gloves ($n=20$, 35%). A significant difference of medium effect was detected between participants taking over the examiner role first ($n=9$, 30% closed the door) and those observing first ($n=22$, 79% closed the door), $\chi^2(1, N=58)=13.73$, $p<0.001$, $\phi=0.49$. Similarly, less of the initially examining participants ($n=11$, 37%) remembered hand disinfection/gloves in contrast to the initially observing participants ($n=27$, 96%) marked by a significant difference of large effect, $\chi^2(1, N=58)=22.90$, $p<0.001$, $\phi=0.63$. Considering these two tasks, no significant differences were observed regarding gender, age, or prior experience in skin cancer examination.

Conversely, consistent adherence was observed in taking the patient's medical history in terms of covering e.g., pre-existing conditions, allergies and medication. Additionally, throughout the examination, tasks like utilizing the dermatoscope and photo documentation were consistently executed by all participants. All other examination steps, including to obtain agreement in the examination as well as noticing and describing skin lesions, were performed by between 91 and 98% of the students.

3.2.6 Simulation task completion (RQ 5)

Independently of the observer questionnaire, we evaluated how many of the participants detected and took a photo of the actual malignant melanomas present in

each of the five cases. 45 students (78%) documented all present malignant melanomas, whereas 13 (22%) missed one. This ratio was statistically independent of gender, $\chi^2(1, N=58)=0.55$, $p=0.534$, age, $\chi^2(1, N=58)=0.37$, $p=0.546$, prior VR experience, $p=0.534$, prior skin cancer screening experience, $U=312$, $z=0.41$, $p=0.679$, and order of the observer, $\chi^2(1, N=58)=0.55$, $p=0.534$. However, a significant association of strong effect was detected between missed melanomas and the case number, $FFH=18.52$, $p<0.001$, Cramer's $V=0.62$. The highest percentage of missed melanoma occurred in case 4 ($n=8$, 73%), in which two melanomas were present. The missed melanoma located under the left armpit was the same among all participants. In contrast, none of the malignant melanomas on the head were missed (case 1 and 5). In case 2, 2 out of 14 participants (14%) missed the melanoma on the left hip while 3 out of 13 participants (23%) missed the melanoma on the right foot sole in case 3.

3.2.7 Cybersickness (RQ 6)

In our cohort, less than half of the participants reported no symptoms of cybersickness after the simulation at all ($n=23$, 40%). The remaining participants assessed very few ($n=12$, 21%), few ($n=11$, 19%), moderate ($n=7$, 12%), strong ($n=4$, 7%), and very strong side effects ($n=1$, 2%). Among those participants, mentioned symptoms covered mainly dizziness ($n=17$, 49%), nausea ($n=6$, 17%), headache ($n=3$, 9%), watery/tired eyes ($n=2$, 6%), blurred vision ($n=1$, 3%), panic ($n=1$, 3%), loss of orientation ($n=1$, 3%), and a self-described phenomenon of a “strange feeling in head” ($n=1$, 3%). The median intensity of side effects reported was 2 (IQR 1–3) which equals ‘few side effects’. A weak positive correlation among ages, $r_s(56)=0.30$, $p=0.023$, and a significant difference of medium effect between participants with (mean rank of 23.87) and without VR experience (mean rank of 40.2) was detected, $U=166$, $z=-3.650$, $p<0.001$, $r=0.48$. The intensity of experienced cybersickness correlated strongly positively with the SIM-TLX score, $r_s(56)=0.56$, $p<0.001$, and moderately negatively with the SUS score, $r_s(56)=-0.46$, $p<0.001$.

3.2.8 Feedback comments

The end of the post-simulation questionnaire enabled participants to provide us with free text feedback which were received from 16 participants (19%).

Positive comments emphasized subjective knowledge gain and fun during the simulation. Furthermore, the students appreciated the quality of skin lesions and the discussions with a dermatologist on the findings as follow-up to the simulation.

“It was a lot of fun. A great change from traditional learning” (male, 23 years).

“The skin findings were very good, in the outpatient clinic you would not be able to see the nevi so well.” (female, 23 years).

“Really great! Thank you for organizing the seminar! Definitely learnt a lot and it was also a lot of fun.” (female, 24 years).

Neutral responses contained both praise and criticism or requests.

“Very long introduction, but I wouldn’t be able to cope well with the simulation without an introduction.” (female, 23 years).

“It was fun but also time consuming.” (female, 25 years).

“I think the learning format with the VR glasses is very good, but I think it’s better to learn a screening on a real patient. However, I think it’s an extremely good medium for other learning content.” (female, 24 years).

Negative comments mainly covered technical limitations, such as bumpy vision, annoying cable or the lack of certain examination possibilities.

“Mucous membranes [could] not be seen.” (female, 23 years).

“Slightly shaky image.” (female, 23 years).

“[Please] not as a compulsory event.” (male, 31 years).

Some students gave recommendations, e.g., using other game engines, the ability to directly interact with the patients’ body parts, less guided history taking or more body regions to be examinable, e.g., mouth cavity or interdigits.

“Is the Unreal Engine 5 perhaps better in terms of graphics [than Unity]?” (male, 34 years).

“Define and communicate the position of other people in the room beforehand as an orientation aid.” (female, 24 years).

“Fewer predefined answers. Unlock more examination locations.” (female, 24 years).

4 Discussion

Due to the lack of available practical courses for teaching and practicing skin cancer screening, our research design aimed to capture the initial reactions and immediate outcomes of using a VR simulation to fill this gap in dermatological education. This study represents an evaluation of a self-developed VR scenario created exclusively with academic partners. Consequently, assessing usability, immersion, task load, user acceptance and experience was crucial for the developers to understand and explore the potential and limitations of this novel approach.

4.1 VR-Experience, suitability and self-efficacy

Our study revealed that around two-thirds of participants had some prior HMD-based VR experience, but none of them regularly. A significant association was observed between overall prior VR experience and participation in VR educational courses. In a previous study (Mergen et al., 2023b) only 32.5% had reported prior VR experience. The increase is partly explainable by the integration of a VR course in the preceding semester that is part of the neurosurgical curriculum, in which 32.8% of our cohort had also participated. While being VR-naïve can impede personal learning experience (Hamilton et al., 2021), this increase is one example of VR being utilized more frequently in medical education, which ultimately enhances familiarity with its controls and overarching concepts, thereby shifting the focus of interventional studies towards the accurate measurement of the actual constructs.

A majority of our cohort reported having only theoretical knowledge about skin cancer screening. Only few have observed or conducted the procedure under supervision, which highlights the practical gap concerning this important clinical skill. Furthermore, this is also supported by the initial self-assessment which indicated that about half of the students lacked confidence in performing skin cancer screenings. A relevant number of melanomas is diagnosed accidentally, which is why all physicians should be aware of the basic rules to recognize and differentiate between benign and malignant skin lesions (Omara et al., 2020).

Before conducting the simulation, around 70% found VR suitable for learning skin cancer screening which slightly increased after the simulation, indicating the usefulness of VR as an educational tool even at this early development stage. As part of a preceding survey conducted at our medical faculty, students were asked on the potential of integrating VR into the curricula of clinical subjects (Mergen et al., 2023b) of which dermatology in contrast was voted for by only 20%. The students' perceptions might be biased by currently available VR programs and respective research, which mainly focus on surgical procedures (Co et al., 2023; Ntakakis et al., 2023), missing other use cases like training skin cancer screening.

The participants of our study were at a late stage of their medical training but assessed their general self-efficacy rather low. Compared to the reference values described by Beierlein et al. (2012), the average self-efficacy score is similar to those of a cohort with a low degree of education. This indicates rather inferior self-confidence and trust in one's own abilities of our cohort of prospective medical doctors who are about to graduate. Moreover, such low self-assessment does not coincide with the Dunning Kruger effect that rather expects overestimations and is often mentioned in critics against studies examining self-assessment (Dunning et al., 2003; Kruger & Dunning, 1999). However, the low scores might be biased by having the following skin cancer screening in mind while responding to the ASKU items, although the questionnaire's items focus on handling difficult tasks and situations in general without referring to a specific topic. Nevertheless, these results indicate that medical students do not feel well trained and prepared, supporting the fact that medical school is still focusing too much on theoretical knowledge and too little on practical clinical skills, such as decision-making. As one initiative accounting for this circumstance, the German stakeholders agreed

in 2017 on the “Masterplan for medical schools 2020” (Wissenschaftsrat, 2018) aiming to enhance competence-based concepts for education of health professionals with revised learning objectives (Medizinischer Fakultätentag, 2021) that emphasizes training of soft and social skills among other crucial changes to the curricula.

4.2 Usability, task load, immersion and presence

When implementing a new application for teaching important skills, it is crucial to not limit its potential and acceptance by usability problems. With an average SUS score of 67.8 ($SD=14.0$), our simulation received an already acceptable rating comparable with that reported by Junga & Schmidle et al. ($M=70.9$, $SD=15.5$) (Junga, Schmidle et al., 2024), still leaving room for improvement. The main concerns covered lack of confidence using the system and the need for help of technical support during the simulation (see Supplement 1). Considering this application is a first prototype and after collecting valuable feedback, we will re-evaluate usability in follow-up studies.

Cognitive task load has been rather low with a mean SIM-TLX score of 26.2 ($SD=10.5$) in this study, which is comparable with the value reported by Junga & Schmidle et al. ($M=27.7$, $SD=13.4$) (Junga, Schmidle et al., 2024).

Literature shows that learning suffers, when task load is too high (Favre-Félix et al., 2022). Still, it is important to challenge the students adequately to generate a long-term influence on skill development (Guadagnoli et al., 2012; Nelson & Elias, 2023). The issue of determining the so-called Optimal-Challenge Point (OCP) is part of current discussions in literature (Nkulu-Ily, 2023).

Oversimplifications of actual reality are thought to inadequately prepare trainees for real life scenarios (Bui et al., 2021). However, VR simulations today often focus on single tasks leaving out many other components that contribute to a real clinical case (Robinson et al., 2017). Currently, e.g., virtual patients’ expressions are not complex enough to resemble real human behavior, which makes creating adequately stimulating VR scenarios very challenging.

Furthermore, we found a strong negative correlation between the SIM-TLX and SUS scores indicating that bad usability could lead to high task load and vice versa, emphasizing the need for a good user experience as baseline for providing real added value. In addition, our results reveal that a higher appreciation of exploring new technologies and a higher perceived confidence in using the application were linked to smaller task load accounting once more for the obligatory familiarization of medical students with VR to enhance its usefulness. Surprisingly, no significant differences were detected between VR-experienced and VR-naïve participants regarding SUS and SIM-TLX score. This could be at least partially explained by the fact that even VR-experienced users in our cohort have not used VR on a regular basis. However, the average SIM-TLX score of VR-experienced participants was still lower ($M=24.4$, $SD=9.8$) than the VR-naïve participant’s score, respectively ($M=29.6$, $SD=11.2$).

In general, we argue that VR courses might reduce extraneous cognitive load by providing information visually and interactively and therefore leading to better preparing the students for their clinical practice.

Considering the sense of immersion and realism (RJPQ), our simulation was valued slightly above average on a rating from 0 (*not at all*) to 10 (*absolutely*). There is clear evidence that the level of presence in a VR environment influences behavior (Slater, 2009). Furthermore, the sense of immersion and presence increases as an outcome of training in VR (Gamito et al., 2010), which benefits the Situated Learning Theory (Lave & Wenger, 1991). Reference values for the RJPQ scores would be of high interest for further evaluations. Considering that our skin cancer screening course was for many participants as well as our faculty the first course with VR components, chances are that RJPQ scores will increase with further similar courses alongside experience. Junga & Schmidle et al. reported the mean values for the ordinal variables reality judgment ($M=4.82$, $SD=2.01$), internal/external correspondence ($M=5.27$, $SD=1.70$), and attention/ absorption ($M=5.76$, $SD=1.95$) (Junga, Schmidle et al., 2024), which are not directly comparable with the median scores we computed, i.e. $Mdn=4.94$ (IQR 4.13–6.13), $Mdn=6.42$ (IQR 4.96–7.67), and $Mdn=5.75$ (IQR 4.69–7.06) respectively. A significant difference was found between VR-experienced and VR-naïve participants regarding external and internal correspondence, which refers to how realistically actions and movements in the real world are represented in VR. Fast progress and new advancements in AI technology have been shown to allow for augmentation of educational scenarios contributing to more realistic experiences (Chheang et al., 2024; Nagi et al., 2023). The potential of AI is demonstrated by the fact that our AI-based depictions of melanomas were rated as very realistic according to the study participants. Nonetheless, despite qualitative feedback iterations with dermatologists during the development phase of the application, the discussed quantitative evaluation is only based on the participating medical students. Quantitative experts' evaluations of the skin lesions will be included in prospective studies and are recommendable for similar VR-based training scenarios. Junga & Schmidle et al. reported a mean value for the ordinal variable realism of the self-developed skin lesions ($M=6.48$, $SD=2.26$), whereas we reported a median score of 8 (IQR = 6–9), thus results are not directly comparable (Junga, Schmidle et al., 2024).

4.3 Subjective learning outcome

Constructivist Learning has been shown to improve clinical skills (Thampinathan, 2022). With our simulation we followed this approach by immersing the students in the role of a physician that is responsible for a virtual patient in a face-to-face interaction. In addition, gamification concepts can benefit skill acquisitions (Seymour et al., 2023) and are already used in the context of dermatological training (Szeto et al., 2021). Ratings on the subjects' fun and usefulness of the VR course were high, and a considerable number of students felt more competent in skin cancer screening after the simulation, which is demonstrated by a significant difference of large effect considering the respective pre- and post-VR response. Junga & Schmidle

et al. reported a respective significant difference as well before and after training (Junga, Schmidle et al., 2024). According to a study of Gulbrandsen et al. (2013), such results can positively correlate with actual performance, even years after the intervention (Gulbrandsen et al., 2013). As with general self-efficacy, the initially low self-assessment of competences does not indicate the presence of the Dunning-Kruger effect (Dunning et al., 2003; Kruger & Dunning, 1999). As the evaluation of long-time retention of acquired knowledge and skills is crucial for the appropriateness of teaching methods, this will be subject of prospective investigations.

4.4 Enhanced learning through observation

Using the elucidated concept of mutual observation during the examination, we ensured three aspects: Firstly, we could increase learning perspectives by guaranteeing that the students had to engage in three different ways in the practical course—from an audience perspective during the introduction to skin cancer screening through the dermatologist, from a first-person perspective as examiner and once from an observer's perspective reviewing the examiner's performance.

The questions of the observer's checklist followed a systematically structured order on the steps of a complete skin cancer screening. The student that took the observer role first could therefore immediately bring their newly gained knowledge to practice while the other student could benefit from the repetition after the simulation. Secondly, this questionnaire provided a form of performance measurement that resembles aspects of an OSCE. Lastly, we compensated for the limited amounts of simultaneously usable HMDs. Since every HMD needed a separate room, the number of parallel sessions was limited to three in our study. This strategy is further supported by the Experiential Learning Theory (Kolb, 1984), as it provides a concrete experience in combination with observation and experimentation. In this aspect, our course design differs from that of Junga & Schmidle et al., where participants conducted the simulation alone, receiving instructions and technical assistance via an intercom system (Junga, Schmidle et al., 2024). In contrast to the latter, we provided technical help directly in the room through the presence of study assistants.

In future studies, to extend the results of the external evaluation in a more detailed fashion, Direct Observation of Procedural Skills (DOPES) could be used as demonstrated by Yi-Ping Chao et al. (Chao et al., 2023). when examining the effectiveness of 360° videos in history and physical examination training.

Interestingly, the most frequently forgotten examination steps were not part of the actual skin cancer screening, but rather related to general principles of patient management, such as closing the door to ensure privacy or hygiene, i.e., disinfect hands/wear gloves. Interpretation of these results are two-fold. On the one hand, students might lack these skills, emphasizing the need for training them, e.g., through VR courses. On the other hand, it remains unclear if students had performed these steps in a real-world clinical setting due to situational realism. To account for this, we tested for significant differences between participants who remembered or omitted these steps in terms of their RJPQ scores as indicators of perceived immersion: remarkably, none was detected, supporting the first hypothesis.

4.5 Cybersickness

One major possible flaw of VR simulations is the phenomenon of cybersickness. Its occurrence needs to be assessed because it can significantly reduce user experience and therefore overall course-quality (Chandra et al., 2022). In agreement with another study, the intensity of cybersickness strongly correlated with experienced cognitive task load, which therefore had a negative impact (Hsin et al., 2023). Additionally, increased cybersickness was significantly associated with a poorer usability rating. Large parts of our cohort did not report any respective issues, whereas a still considerable number reported dizziness, nausea, or headache, which are most common symptoms of cybersickness reported in literature (Caserman et al., 2021). Junga & Schmidle et al. reported the mean value for the ordinal variable intensity of cybersickness ($M=2.0$, $SD=1.3$), whereas we reported the median and IQR ($Mdn=2$, IQR 1–3), thus we could not directly compare our results with theirs (Junga, Schmidle et al., 2024). Remarkably, perceived intensity of cybersickness seemed to increase with age in our cohort while participants with prior VR experience described less severe intensity compared to VR-naïve students. The role of age indeed is of interest for further investigations, as negative correlations with system usability, immersion, the enjoyment and positive assessment of the appropriateness of VR were observed among our cohort, as described earlier.

Even though these sensations can create discomfort in few individuals, research shows that the benefits of immersive VR usage outweigh the risk of cybersickness (Drazich et al., 2023). Furthermore, there are concrete approaches to reduce these symptoms, such as minimizing latency, increasing framerate, or avoiding flickering (Chandra et al., 2022).

4.6 Feedback comments

The received feedback turned out to be very appreciative, which is in line with a previous conducted survey showing a strong affinity towards VR integration (Mergen, et al., 2023b). Nonetheless, there is room for improvement, especially considering technical aspects. Improving graphics, interaction and expanding the range of diagnostic possibilities will further enhance user experience. Even though evaluating free text answers is not standardized like quantitative data, it adds a lot of value to the overall insight of the students' opinions, especially in this early development stage. Therefore, we recommend including this aspect in form of free text answers, structured interviews or focus groups for all courses and educational applications using VR. Corresponding recommendations for analyzing qualitative research data should be taken into account when developing questions for structured interviews or focus groups and analyzing responses (Bryda & Costa, 2023).

4.7 Strengths and limitations of this study

This study was conducted at the Medical Faculty of Saarland University implemented in the mandatory dermatological practical course. Therefore, selection bias, e.g., in the form of only selecting participants who are interested in VR, was avoided. In addition, the cohort resembled the current demographic statistics of medical students in Germany, making the results reliable and representative.

Throughout the pre- and post-VR simulation surveys, we focused on using established questionnaires to provide valid outcomes. Partly, shortened versions of original questionnaire versions were used (e.g., RJPQ, SIM-TLX, CSQ-VR) in agreement with either outcomes from corresponding publications (Baños et al., 2000; Harris et al., 2020) or due to overlapping aspects covered by the items considering the total length of post-simulation questionnaire items. The original rating scale of SIM-TLX was reduced from 0–20 to 0–10 due to unification in our study. However, we argue that this range is likely to appropriately capture the self-assessments in a differentiated way as well. Furthermore, we designed our own self-assessment items partially based on already existing questionnaires as well (LEQ, TEI).

Different concepts, such as the Technology Acceptance Model (TAM) by Davis et al. (Davis, 1989), which consists of the aspects “Perceived Usefulness (PU)” and “Perceived Ease of Use (PEU)” could be of further interest for future evaluations.

While we used Spearman’s correlation coefficient to examine monotonic relationships among variables, we acknowledge that this approach does not allow for the exploration of more complex interrelationships, such as those involving latent variables or indirect effects, which could be examined, e.g., using Structural Equation Modeling (SEM). However, since SEM typically requires a larger sample size to produce reliable and stable estimates, future research with a larger cohort should consider them to explore these complex relationships for deeper insights into the dynamics among the variables studied. Since the 5 examination cases were randomly assigned to the students at the beginning of each simulation with the same probability, a recall error was avoided because the student who first observed and the student who first examined were confronted with malignant melanomas and nevi at different sites.

We acknowledge the subjective nature of self-reported data. Nonetheless, we argue that the students had no incentive for not answering to the best of their estimation, since all the data were pseudonymized and no grade was tied to the students’ performance or answers. Following the scope of a feasibility study that primarily addressed subjective learning outcomes and users’ opinions, we consider the used questionnaires as appropriate for answering our research questions. To objectify assessments, future studies could compare the students’ exam grades with their performance in VR or a short test about theoretical knowledge could be done prior to and after the simulation. Especially the examination of long-term retention of the acquired knowledge will be of high interest in prospective study designs, as well as clinical transferability of the skills acquired in VR.

Skin cancer screening has so far only been taught in theoretical form as subject of corresponding dermatology lectures. However, the examination process of a skin cancer screening itself could not be addressed by traditional teaching methods.

Thus, despite being an interesting research question, direct comparisons between traditional teaching and VR-based learning are not feasible at this stage. This VR course represents a pioneering approach, offering a novel and essential addition to the dermatology curriculum.

The simulation itself has been developed by the medical training - consortium, which consists of experienced 3D designers, AI experts and medical didactic experts. To the best of our knowledge this is the only interdisciplinary VR and AI project focusing on medical education that follows a small iteration strategy by early implementation of the prototypes into medical curricula. With this strategy we ensure valuable feedback by the actual users and clinical experts to further improve the simulation.

Standardized guidelines for the evaluation of VR applications for medical education are still lacking and highly demanded according to literature (Kim & Kim, 2023; Mergen et al., 2024; Walter et al., 2021; Wu et al., 2022), they could not be considered for this work. However, this feasibility study provides crucial outcomes for the follow-up developments and extensions to this new approach and offers insights for similar VR-based training concepts.

5 Conclusions

Despite the benefits for patients, medical students hardly have the possibility to actively practice a complete skin cancer screening in medical school during their studies to date. This highlights the great potential of supplementing training with VR simulations, also considering patient comfort and privacy which are natural and ethical limits during shadowing and simulations with actors. We conducted a feasibility study to investigate the students' acceptance, evaluation of usability and self-assessments considering the suitability of VR for training essential medical skills, after having them perform a complete skin cancer screening in VR. The study results demonstrate a strong positive attitude with a significant increase in perceived self-competence in performing a skin cancer screening. This emphasizes the high potential of benefit of such practical VR-courses in medical education, while simultaneously pointing out possible ways of improvement to be addressed accordingly, e.g., considering usability and cybersickness.

Our findings suggest that prior experience with skin cancer screening positively influences self-assessed competence and knowledge on this subject. Age emerged as a critical factor, with older participants assessing lower usability scores, reduced enjoyment and appropriateness ratings for the simulation. Gender differences were also noted considering need for technical help during the simulation and openness to new technologies. A negative correlation between task load and usability underscores the importance of optimizing for a better user experience. Higher intensity of cybersickness was associated with higher cognitive task load and lower usability ratings. Since this study indicated the feasibility of integrating VR into medical curricula, further dissemination is planned after the implementation of improvements in the application. Follow-up studies are planned for validation and verification of our

findings. This work can serve as an example for how to close gaps in medical education successfully with VR-assisted courses and how to evaluate crucial aspects of feasibility in this context.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s10639-024-13019-w>.

Acknowledgements We would like to thank Thomas Jany, Joe Miebach and Gaspar Eltze for helping with the technical supervision and data collection during the VR course. Especially, we would like to thank Joe Miebach for his additional help in formatting this article and compiling the bibliography. Furthermore we thank all consortium members of the medical tr.AI.ning—project for the cooperative work: University of Münster (CVMLS, IfAS)—Benjamin Risse, Pascal Kockwelp, Leon Pielage, Valentin Brosch, Bernhard Marschall, Anna Junga, Henriette Schulze, Ole Hätscher, Niklas Tiefenbach, Julia Fedrau; University of Applied Sciences Münster (FHMS)—Tina Glückselig, Kathrin Ungru, Philipp Bozdere, Julia Leuer, Mariel Kruithoff; Saarland University (UMTL, CHELM)—Antonio Krüger, Dimitar Valkov, Tim Düwel, André Zenner, Florian Daiber, Erum Manzoor, Dennis Gudea, Norbert Graf, Marvin Mergen, Marcel Meyerheim, Thomas Jany, Joe Miebach, Gaspar Eltze; University of Fine Art Saar (HBKsaar)—Michael Schmitz, Mert Akbal, Corbin Sassen, Melisa Kujevic, Frederieke Wagner, Jennifer Schwarz.

Authors' contributions All authors contributed to the study conception and design in collaboration with the University of Münster (IfAS). Data collection was performed by Marvin Mergen and Marcel Meyerheim. Statistical analysis was conducted by Marcel Meyerheim. The first draft of the manuscript was written by Marvin Mergen and Marcel Meyerheim. All authors commented on previous versions of the manuscript, revised, and approved the final manuscript.

Funding Open Access funding enabled and organized by Projekt DEAL. This study received funding from the German Federal Ministry of Education and Research under the BMBF project number 16DHBKI080 within the project “medical tr.AI.ning”.

Availability of data and other materials Collected data and statistical analysis from the included study are available from the authors upon reasonable request.

Statements and Declarations

Ethics approval and consent to participate Final ethical approval for studies related to the medical tr.AI.ning project was obtained from the “Ärztchamber des Saarlandes” on November 8, 2021 (Reference: 244/21). Written informed consent was obtained from all study participants at the beginning of the respective practical course dates with the VR simulation.

Conflicts of interest The authors declare that they have no competing interests.

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

- Abbas, J. R., Chu, M. M. H., Jeyarajah, C., Isba, R., Payton, A., McGrath, B., Tolley, N., & Bruce, I. (2023). Virtual reality in simulation-based emergency skills training: A systematic review with a narrative synthesis. *Resuscitation plus*, 16, 100484. <https://doi.org/10.1016/j.resplu.2023.100484>
- Abundez Toledo, M., Ghanem, G., Fine, S., Weisman, D., Huang, Y. M., & Rouhani, A. A. (2024). Exploring the promise of virtual reality in enhancing anatomy education: A focus group study with medical students. *Frontiers in Virtual Reality*, 5. <https://doi.org/10.3389/frvir.2024.1369794>
- Bangor, A., Kortum, P., & Miller, J. (2009). Determining what individual SUS scores mean: Adding an adjective rating scale. *J. Usability Studies*, 4(3), 114–123. <https://doi.org/10.5555/2835587.2835589>
- Baños, R. M., Botella, C., Garcia-Palacios, A., Villa, H., Perpiña, C., & Alcañiz, M. (2000). Presence and reality judgment in virtual environments: A unitary construct? *Cyberpsychology and Behavior*, 3(3), 327–335. <https://doi.org/10.1089/10949310050078760>
- Beierlein, C., Kovaleva, A., Kemper, C. J., & Rammstedt, B. (2012). *Ein Messinstrument zur Erfassung subjektiver Kompetenzerwartungen: Allgemeine Selbstwirksamkeit Kurzskala (ASKU)*. GESIS-Working Papers, 2012/17, 24.
- Bonmarin, M., Läubli, S., & Navarini, A. (2022). Augmented and Virtual Reality in Dermatology—Where Do We Stand and What Comes Next? *Dermato.*, 2(1), Article 1. <https://doi.org/10.3390/dermato2010001>
- Brooke, J. (1996). *SUS – a quick and dirty usability scale* (pp. 189–194)
- Bryda, G., & Costa, A. P. (2023). Qualitative Research in Digital Era: Innovations, Methodologies and Collaborations. *Social Sciences.*, 12(10), Article 10. <https://doi.org/10.3390/socsci12100570>
- Bui, I., Bhattacharya, A., Wong, S. H., Singh, H. R., & Agarwal, A. (2021). Role of Three-Dimensional Visualization Modalities in Medical Education. *Frontiers in Pediatrics*, 9. <https://doi.org/10.3389/fped.2021.760363>
- Caserman, P., Garcia-Agundez, A., Zerban, A. G., & Göbel, S. (2021). Cybersickness in current-generation virtual reality head-mounted displays: Systematic review and outlook. *Virtual Reality*, 25(4), 1153–1170. <https://doi.org/10.1007/s10055-021-00513-6>
- Chandra, A. N. R., Jamiy, F. E., & Reza, H. (2022). A systematic survey on cybersickness in virtual environments. *Computers*, 11(4), 51. <https://doi.org/10.3390/computers11040051>
- Chao, Y.-P., Kang, C.-J., Chuang, H.-H., Hsieh, M.-J., Chang, Y.-C., Kuo, T. B. J., Yang, C. C. H., Huang, C.-G., Fang, T.-J., Li, H.-Y., & Lee, L.-A. (2023). Comparison of the effect of 360° versus two-dimensional virtual reality video on history taking and physical examination skills learning among undergraduate medical students: A randomized controlled trial. *Virtual Reality*, 27(2), 637–650. <https://doi.org/10.1007/s10055-022-00664-0>
- Chheang, V., Sharmin, S., Marquez-Hernandez, R., Patel, M., Rajasekaran, D., Caulfield, G., Kiafar, B., Li, J., Kullu, P., & Barmaki, R. L. (2024). *Towards Anatomy Education with Generative AI-based Virtual Assistants in Immersive Virtual Reality Environments* (arXiv:2306.17278). arXiv. <https://doi.org/10.48550/arXiv.2306.17278>
- Co, M., Chiu, S., & Cheung, H. H. B. (2023). Extended reality in surgical education: A systematic review. *Surgery*, 174(5), 1175–1183. <https://doi.org/10.1016/j.surg.2023.07.015>
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Routledge. <https://doi.org/10.4324/9780203771587>
- Das, K. (2024). *Futuristic Trends in Dermatology* (Vol. 2). <https://www.doi.org/https://doi.org/10.58532/V3BBMS2P5CH1>
- Davis, F. D. (1989). Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. *MIS Quarterly*, 13(3), 319–340. <https://doi.org/10.2307/249008>
- Drazich, B. F., McPherson, R., Gorman, E. F., Chan, T., Teleb, J., Galik, E., & Resnick, B. (2023). In too deep? A systematic literature review of fully-immersive virtual reality and cybersickness among older adults. *Journal of the American Geriatrics Society*. <https://doi.org/10.1111/jgs.18553>
- Duarte, A. F., Sousa-Pinto, B., Azevedo, L. F., Barros, A. M., Puig, S., Malvehy, J., Haneke, E., & Correia, O. (2021). Clinical ABCDE rule for early melanoma detection. *European Journal of Dermatology : EJD*, 31(6), 771–778. <https://doi.org/10.1684/ejd.2021.4171>
- Dunning, D., Johnson, K., Ehrlinger, J., & Kruger, J. (2003). Why People Fail to Recognize Their Own Incompetence. *Current Directions in Psychological Science*, 12(3), 83–87. <https://doi.org/10.1111/1467-8721.01235>

- Favre-Félix, J., Dziadzko, M., Bauer, C., Duclos, A., Lehot, J.-J., Rimmelé, T., & Lilot, M. (2022). High-fidelity simulation to assess task load index and performance: A prospective observational study. *Turkish Journal of Anaesthesiology and Reanimation*, 50(4), 282–287. <https://doi.org/10.5152/TJAR.2022.21234>
- Friedman, R.J., Rigel, D.S., & Kopf, A.W. (1985). Early detection of malignant melanoma: The role of physician examination and self-examination of the skin. *CA: A Cancer Journal for Clinicians*, 35(3):130–151. <https://doi.org/10.3322/canjclin.35.3.130>
- Fritz, C. O., Morris, P. E., & Richler, J. J. (2012). Effect size estimates: Current use, calculations, and interpretation. *Journal of Experimental Psychology. General*, 141(1), 2–18. <https://doi.org/10.1037/a0024338>
- Gamito, P., Oliveira, J., Morais, D., Baptista, A., Santos, N., Soares, F., Saraiva, T., & Rosa, P. (2010). Training presence: The importance of virtual reality experience on the “sense of being there.” *Studies in Health Technology and Informatics*, 154, 128–133. <https://doi.org/10.3233/978-1-60750-561-7-128>
- García-Robles, P., Cortés-Pérez, I., Nieto-Escámez, F. A., García-López, H., Obrero-Gaitán, E., & Osuna-Pérez, M. C. (2024). Immersive virtual reality and augmented reality in anatomy education: A systematic review and meta-analysis. *Anatomical Sciences Education*, 17(3), 514–528. <https://doi.org/10.1002/ase.2397>
- Gençoğlu, Ş. (2024). Enhancing dermatology: The current landscape and future prospects of augmented and virtual reality technologies. *Journal of Health Sciences and Medicine*. 7(1): Article 1. <https://doi.org/10.32322/jhsm.1358284>
- Gruber, M., & Strömsten, L. M. J. (2021). *Psychometric evaluation of the Learning Experience Questionnaire (LEQ) course evaluation instrument* (pp. 22–22)
- Guadagnoli, M., Morin, M.-P., & Dubrowski, A. (2012). The application of the challenge point framework in medical education: Application of the challenge point framework. *Medical Education*, 46(5), 447–453. <https://doi.org/10.1111/j.1365-2923.2011.04210.x>
- Gulbrandsen, P., Jensen, B. F., Finset, A., & Blanch-Hartigan, D. (2013). Long-term effect of communication training on the relationship between physicians' self-efficacy and performance. *Patient Education and Counseling*, 91(2), 180–185. <https://doi.org/10.1016/j.pec.2012.11.015>
- Hamilton, D., McKechnie, J., Edgerton, E., & Wilson, C. (2021). Immersive virtual reality as a pedagogical tool in education: A systematic literature review of quantitative learning outcomes and experimental design. *Journal of Computers in Education*, 8(1), 1–32. <https://doi.org/10.1007/s40692-020-00169-2>
- Harden, R. M., & Gleeson, F. A. (1979). Assessment of clinical competence using an objective structured clinical examination (OSCE). *Medical Education*, 13(1), 39–54. <https://doi.org/10.1111/j.1365-2923.1979.tb00918.x>
- Harris, D., Wilson, M., & Vine, S. (2020). Development and validation of a simulation workload measure: The simulation task load index (SIM-TLX). *Virtual Reality*, 24(4), 557–566. <https://doi.org/10.1007/s10055-019-00422-9>
- Hsin, L.-J., Chao, Y.-P., Chuang, H.-H., Kuo, T. B. J., Yang, C. C. H., Huang, C.-G., Kang, C.-J., Lin, W.-N., Fang, T.-J., Li, H.-Y., & Lee, L.-A. (2023). Mild simulator sickness can alter heart rate variability, mental workload, and learning outcomes in a 360° virtual reality application for medical education: A post hoc analysis of a randomized controlled trial. *Virtual Reality*, 27(4), 3345–3361. <https://doi.org/10.1007/s10055-022-00688-6>
- Junga, A., Schmidle, P., Pielage, L., Schulze, H., Hätscher, O., Ständer, S., Marschall, B., Braun, S. A., & Consortium, the medicaltr A. ning. (2024). New horizons in dermatological education: Skin cancer screening with virtual reality. *Journal of the European Academy of Dermatology and Venereology*, n/a(n/a). <https://doi.org/10.1111/jdv.19960>
- Kolb, D. (1984). *Experiential Learning: Experience as The Source of Learning and Development*". Prentice Hall, Inc.
- Kim, H.-Y., & Kim, E.-Y. (2023). Effects of medical education program using virtual reality: A systematic review and meta-analysis. *International Journal of Environmental Research and Public Health*, 20(5), 3895. <https://doi.org/10.3390/ijerph20053895>
- Kourtesis, P., Linnell, J., Amir, R., Argelaguet, F., & MacPherson, S. E. (2023). Cybersickness in virtual reality questionnaire (CSQ-VR): A validation and comparison against SSQ and VRSQ. *Virtual Worlds*, 2(1), 16–35. <https://doi.org/10.3390/virtualworlds2010002>

- Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology*, 77(6), 1121–1134. <https://doi.org/10.1037/0022-3514.77.6.1121>
- Lave, J., & Wenger, E. (1991, September 27). *Situated Learning: Legitimate Peripheral Participation*. Higher Education from Cambridge University Press; Cambridge University Press. <https://doi.org/10.1017/CBO9780511815355>
- Medizinischer Fakultätentag. (2021). *NKLM 2.0*. <https://nkml.de/zend/objective/list/orderBy/@objectivePosition/studiengang/PF2/zeitsemester/2021/fachsemester/VIII.2.%20%20%C3%84rztliche%20Gespr%C3%A4chs%C3%BChrung>
- Mergen, M., Junga, A., Risse, B., Valkov, D., Graf, N., Marschall, B., & medical training consortium. (2023a). Immersive training of clinical decision making with AI driven virtual patients – a new VR platform called medical training. *GMS Journal for Medical Education*, 40(2), Doc18. <https://doi.org/10.3205/zma001600>
- Mergen, M., Meyerheim, M., & Graf, N. (2023b). Towards integrating virtual reality into medical curricula: A single center student survey. *Education Sciences*, 13(5), 477. <https://doi.org/10.3390/educsci13050477>
- Mergen, M., Graf, N., & Meyerheim, M. (2024). Reviewing the current state of virtual reality integration in medical education—A scoping review. *BMC Medical Education*, 24(1), 788. <https://doi.org/10.1186/s12909-024-05777-5>
- Mishra, N. R. (2023). Constructivist Approach to Learning: An Analysis of Pedagogical Models of Social Constructivist Learning Theory. *Journal of Research and Development*, 6(01), Article 01. <https://doi.org/10.3126/jrdn.v6i01.55227>
- Mühling, T., Späth, I., Backhaus, J., Milke, N., Oberdörfer, S., Meining, A., Latoschik, M. E., & König, S. (2023). Virtual reality in medical emergencies training: Benefits, perceived stress, and learning success. *Multimedia Systems*, 29(4), 2239–2252. <https://doi.org/10.1007/s00530-023-01102-0>
- Muralidharan, V., Tran, M. M., Barrios, L., Beams, B., Ko, J. M., Siegel, D. H., & Bailenson, J. (2024). Best Practices for Research in Virtual and Augmented Reality in Dermatology. *Journal of Investigative Dermatology*, 144(1), 17–23. <https://doi.org/10.1016/j.jid.2023.10.014>
- Nagi, F., Salih, R., Alzubaidi, M., Shah, H., Alam, T., Shah, Z., & Househ, M. (2023). Applications of Artificial Intelligence (AI) in Medical Education: A Scoping Review. In *Healthcare Transformation with Informatics and Artificial Intelligence* (pp. 648–651). IOS Press. <https://doi.org/10.3233/SHTI230581>
- Nelson, A., & Elias, K. L. (2023). Desirable Difficulty: Theory and application of intentionally challenging learning. *Medical Education*, 57(2), 123–130. <https://doi.org/10.1111/medu.14916>
- Nkulu-Ily, Y. S. (2023). Combining XR and AI for integrating the best pedagogical approach to providing feedback in surgical medical distance education (pp. 452–466). https://doi.org/10.1007/978-3-031-32883-1_41
- Noll, C., von Jan, U., Raap, U., & Albrecht, U.-V. (2017). Mobile augmented reality as a feature for self-oriented, blended learning in medicine: Randomized controlled trial. *JMIR mHealth and uHealth*, 5(9), e139. <https://doi.org/10.2196/mhealth.7943>
- Ntakakis, G., Plomariti, C., Frantzidis, C., Antoniou, P. E., Bamidis, P. D., & Tsoulfas, G. (2023). Exploring the use of virtual reality in surgical education. *World Journal of Transplantation*, 13(2), 36–43. <https://doi.org/10.5500/wjt.v13.i2.36>
- Obagi, Z. A., Rundle, C. W., & Dellavalle, R. P. (2020). Widening the scope of virtual reality and augmented reality in dermatology. *Dermatology Online Journal*, 26(1). <https://doi.org/10.5070/D3261047183>
- Omara, S., Wen, D., Ng, B., Anand, R., Matin, R. N., Taghipour, K., & Esdaile, B. (2020). Identification of incidental skin cancers among adults referred to dermatologists for suspicious skin lesions. *JAMA Network Open*, 3(12), e2030107. <https://doi.org/10.1001/jamanetworkopen.2020.30107>
- Papadatou-Pastou, M., Ntolka, E., Schmitz, J., Martin, M., Munafo, M. R., Ocklenburg, S., & Paracchini, S. (2020). Human handedness: A meta-analysis. *Psychological Bulletin*, 146(6), 481–524. <https://doi.org/10.1037/bul0000229>
- Plass, J. L., Moreno, R., & Brünken, R. (Eds.). (2010). *Cognitive Load Theory*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511844744>
- Ranpariya, V. K., Huang, W. W., & Feldman, S. R. (2022). Virtual reality memory palace: An innovative dermatology education modality. *Journal of the American Academy of Dermatology*, 86(6), 1435–1437. <https://doi.org/10.1016/j.jaad.2021.06.859>

- Rashid, S., Shaughnessy, M., & Tsao, H. (2023). Melanoma classification and management in the era of molecular medicine. *Dermatologic Clinics*, 41(1), 49–63. <https://doi.org/10.1016/j.det.2022.07.017>
- Rebenitsch, L., & Owen, C. (2016). Review on cybersickness in applications and visual displays. *Virtual Reality*, 20(2), 101–125. <https://doi.org/10.1007/s10055-016-0285-9>
- Ritzmann, S., Hagemann, V., & Kluge, A. (2014). The training evaluation inventory (TEI)—Evaluation of training design and measurement of training outcomes for predicting training success. *Vocations and Learning*, 7(1), 41–73. <https://doi.org/10.1007/s12186-013-9106-4>
- Robinson, A. J., Miller, G., & Rukin, N. (2017). Simulation in urological training: Where are we in 2017? *Journal of Clinical Urology*, 10(6), 548–556. <https://doi.org/10.1177/2051415817722933>
- Studierende der Medizin nach Geschlecht bis 2022/2023. (2023, August). Statista. <https://de.statista.com/statistik/daten/studie/200758/umfrage/entwicklung-der-anzahl-der-medizinstudenten/>
- Seymour, A., Borggren, M., & Baker, R. (2023). Escape the monotony: Gamification enhances nursing education. *Journal of Emergency Nursing*. <https://doi.org/10.1016/j.jen.2023.06.004>
- Slater, M. (2009). Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 364(1535), 3549–3557. <https://doi.org/10.1098/rstb.2009.0138>
- Spoto-Cannons, A. C., Isom, D. M., Feldman, M., Zwygart, K. K., Mhaskar, R., & Greenberg, M. R. (2019). Differences in medical student self-evaluations of clinical and professional skills. *Advances in Medical Education and Practice*, 10, 835–840. <https://doi.org/10.2147/AMEP.S222774>
- Szeto, M. D., Strock, D., Anderson, J., Sivesind, T. E., Vorwald, V. M., Rietcheck, H. R., Weintraub, G. S., & Dellavalle, R. P. (2021). Gamification and game-based strategies for dermatology education: Narrative review. *JMIR Dermatology*, 4(2), e30325. <https://doi.org/10.2196/30325>
- ten Cate, O. (2005). Entrustability of professional activities and competency-based training. *Medical Education*, 39(12), 1176–1177. <https://doi.org/10.1111/j.1365-2929.2005.02341.x>
- ten Cate, O. (2013). Nuts and Bolts of Entrustable Professional Activities. *Journal of Graduate Medical Education*, 5(1), 157–158. <https://doi.org/10.4300/JGME-D-12-00380.1>
- Thampinathan, S. (2022). The Application of the Constructivism Learning Theory to Physician Assistant Students in Primary Care. *Education for Health*, 35(1), 26. https://doi.org/10.4103/efh.Efh_333_20
- Walls, R., Nageswaran, P., Cowell, A., Sehgal, T., White, T., McVeigh, J., Staykov, S., Basett, P., Mitelpunkt, D., & Sam, A. H. (2024). Virtual reality as an engaging and enjoyable method for delivering emergency clinical simulation training: A prospective, interventional study of medical undergraduates. *BMC Medicine*, 22(1), 222. <https://doi.org/10.1186/s12916-024-03433-9>
- Walter, S., Speidel, R., & Leitner, J. (2021). Skepticism towards advancing VR technology-student acceptance of VR as a teaching and assessment tool in medicine. *GMS Journal for Medical Education*, 38(6). <https://doi.org/10.3205/zma001496>
- Wissenschaftsrat. (2018). *Neustrukturierung des Medizinstudiums und Änderung der Approbationsordnung für Ärzte | Empfehlungen der Expertenkommission zum Masterplan Medizinstudium 2020*. <https://www.wissenschaftsrat.de/download/archiv/7271-18.html>
- Wu, Q., Wang, Y., Lu, L., Chen, Y., Long, H., & Wang, J. (2022). Virtual simulation in undergraduate medical education: A scoping review of recent practice. *Frontiers in Medicine*, 9. <https://doi.org/10.3389/fmed.2022.855403>

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