ORIGINAL ARTICLE

A comprehensive method to design and assess mixed reality simulations

Agnese Brunzini[1] · Alessandra Papetti[1] · Daniele Messi[2,3] · Michele Germani[1]

Abstract

The scientific literature highlights how Mixed Reality (MR) simulations allow obtaining several benefits in healthcare education. Simulation-based training, boosted by MR, offers an exciting and immersive learning experience that helps health professionals to acquire knowledge and skills, without exposing patients to unnecessary risks. High engagement, informational overload, and unfamiliarity with virtual elements could expose students to cognitive overload and acute stress. The implementation of effective simulation design strategies able to preserve the psychological safety of learners and the investigation of the impacts and effects of simulations are two open challenges to be faced. In this context, the present study proposes a method to design a medical simulation and evaluate its effectiveness, with the final aim to achieve the learning outcomes and do not compromise the students' psychological safety. The method has been applied in the design and development of an MR application to simulate the rachicentesis procedure for diagnostic purposes in adults. The MR application has been tested by involving twenty students of the 6th year of Medicine and Surgery of Università Politecnica delle Marche. Multiple measurement techniques such as self-report, physiological indices, and observer ratings of performance, cognitive and emotional states of learners have been implemented to improve the rigour of the study. Also, a user-experience analysis has been accomplished to discriminate between two different devices: Vox Gear Plus® and Microsoft Hololens®. To compare the results with a reference, students performed the simulation also without using the MR application. The use of MR resulted in increased stress measured by physiological parameters without a high increase in perceived workload. It satisfies the objective to enhance the realism of the simulation without generating cognitive overload, which favours productive learning. The user experience (UX) has found greater benefits in involvement, immersion, and realism; however, it has emphasized the technological limitations of devices such as obstruction, loss of depth (Vox Gear Plus), and narrow FOV (Microsoft Hololens).

Keywords Mixed reality · Augmented reality · Simulation · Medical education · Stress · Cognitive load

1 Introduction

Medical errors and healthcare-related adverse events are proved in the 8% to 12% of hospitalizations; however, comprehensive systematic approaches to patient safety can prevent 50 to 70.2% of them (WHO 2017). Patient safety can

be improved by effective and structured initiatives such as teamwork, education, and training, including simulation (Rodziewicz et al. 2021).

Simulation-based training can help health professionals to develop knowledge, skills, and attitudes, without exposing patients to unnecessary risks. Whether it uses low- or high-fidelity tools, this learning model includes cognitive, technical, and behavioural skills into instructional scenarios, which represent the reality within which learners interact. It emerges as an effective teaching method that transmits a greater motivation in the learning phase (Brunzini et al. 2020). The teacher manages the scenario and controls the parameters to achieve the desired learning goals. This form of experiential learning is suitable for both standard daily activities and rare clinical scenarios. In both cases, the

✉ Agnese Brunzini
  a.brunzini@staff.univpm.it

1 Department of Industrial Engineering and Mathematical Sciences, Università Politecnica Delle Marche, Ancona, Italy

2 Faculty of Medicine and Surgery, Università Politecnica Delle Marche, Ancona, Italy

3 Azienda Ospedaliero Universitaria Ospedali Riuniti, Ancona, Italy

ability to faithfully reflect reality is a fundamental requirement (Curtis et al. 2012).

The extended reality (XR) can be a valid boost of the simulation-based approach, as demonstrated by the numerous applications in different fields (medicine, science, engineering, etc.). In general, the most reported advantage of Augmented Reality (AR) systems in educational settings is learning gains followed by motivation (Bacca et al. 2014). However, the following difficulties still need to be overcome: complexity, technical issues, and teachers' resistance (Garzón et al. 2019).

In the medical field, the AR applications range from understanding anatomy to surgery, to patient education and allow medical students to practice invasive and high-risk procedures in a risk-free environment (Campisi et al. 2020).

The literature highlights how the use of XR simulators allows obtaining several benefits in healthcare education, such as errors reduction, need for less practice, simulation time and failure rate reduction, performance accuracy improvement, shortened learning curve, increased motivation and attention, improved trainees' assessment, enhanced learning retention, and better performance on cognitive-psychomotor tasks (Zhu et al. 2014; Munzer et al. 2019; Gerup et al. 2020). Therefore, XR allows for more authentic learning, making the simulations more realistic and immersive, and providing students a more personalized and explorative learning experience. It is considered useful also in achieving core competencies, such as decision-making and teamwork.

As stated by Sarfati et al. (2019), the success of simulation-based learning in the prevention of medical errors is strongly influenced by the following key elements: scenario design, debriefing, and perception assessment. The final aim is to satisfy the real purposes of education. However, many studies focus on technical issues that are not sufficient without the pedagogical ones (Garzón et al. 2017). To incorporate educational features in XR-based simulation is one of the ongoing challenges to widespread adoption. Moreover, achieving the learning outcomes requires the implementation of effective simulation design strategies able to preserve the psychological safety of learners (Roh et al. 2021).

A learning environment based on XR and simulation could expose students to cognitive overload due to a large amount of information, unfamiliar devices to interact with, and the complex tasks to perform (Wu et al. 2013; Atalay et al. 2016). The risk exposure increases for learners with limited clinical experience (Fraser et al. 2012). Moreover, medical simulations could elicit stress, anxiety, and worry, compromising students' performance and psychological safety (Dias et al. 2018). Indeed, it is important to avoid stress and cognitive overload of students to ensure productive learning (Brunzini et al. 2021). For this reason, stress management education can be a key aspect to enhance technical performance (Goldberg et al. 2018). Nevertheless, the

investigation of the impacts and effects of simulations in education often remains a topic for further research.

Although the difficulty of interpretation, physiological parameters are a sensitive means of detecting changes in cognitive load levels during simulation-based learning (Naismith and Cavalcanti 2015). However, in their review, Scafà et al. (2020) found that they were poorly applied with respect to subjective measurements.

In this context, the present research work proposes a method to design and evaluate a medical simulation, with the final aim to improve student learning and performance, preserving their psychological safety. For its validation, a Mixed Reality (MR) simulator for the rachicentesis procedure has been developed and tested by involving twenty students of the 6th year of Medicine and Surgery. Multiple measurement techniques such as self-report, physiological indices, and observer ratings of stress and cognitive load are implemented to improve the rigour of the study.

2 Extended reality in medical training

In the last years, several studies have tried to integrate AR and manikins obtaining simulators in MR. The main use of AR for learning is to offer immersion in a scenario and provide feedback or additional information. Most of the education and training research focused on high-risk, invasive skills such as endoscopy and surgery. Already in 2014, Zhu et al. found out that 64% of the reviewed papers were within surgery, primarily laparoscopic surgery (44%). This topic continues to be of great interest (Hong et al. 2021). However, a myriad of implementations and case studies exist. Herron et al. reported that AR, among surgery, covers areas such as anatomy and forensic medicine and supports endotracheal intubation, joint injections, and local anaesthesia administration (Herron 2016). From the review of Munzer et al. concerning AR applied to emergency medicine, out of twenty-four articles, 50% focused on education and training and in particular on procedural training and clinical decision-making in a simulated environment (Munzer et al. 2019). The most recent and up-to-date review about AR and MR applications in healthcare education beyond surgery considered twenty-six studies, from January 2013 to September 2018. Authors claimed that the most frequently studied subjects were within anatomy and anaesthesia (especially on central vein catheterization). Other subject areas are radiology, ophthalmology, cardiology, dermatology, family medicine, forensic medicine, gastroenterology, neurology, orthopaedics, paediatrics. Moreover, they highlighted that these studies involved established applications in 27% of all cases, while 73% regarded prototypes (Gerup et al. 2020).

One reason to use AR is to help the user having a visual of the patient's internal body state, interactively (Sherstyuk

et al. 2011). Several AR and MR applications and prototypes involving procedural training can be found in the scientific literature. Several examples of research papers about AR and MR application in procedural training are described in Table 1, with a specific focus on hardware (HW) and software (SW) used, assessment tools, and the number of participants in the experimentation.

As result in Table 1, seven studies (over a total of nine) apply AR for helping learners in needle insertion. Even if the medical procedures are different, five of them use AR to overlay on the skill trainer the internal human anatomy, useful for that kind of procedure (mainly the circulatory system). Thus, the core of these presented studies is always the use of AR as an aid in procedural training, giving information about what learners cannot actually see. This means that the technology is used to provide additional visual material, which is not available in real practice. In this way, the simulation moves away from reality.

Moreover, as resulting from the above-mentioned scientific literature, the principal drawbacks in using AR systems in training activities are related to the technology itself. The issues of registration and overlay, tracking and alignment of virtual content, the stability of virtual elements, hand recognition, discomfort, and limited field of view of Head-Mounted Displays (HMDs) should be solved with an advancement of the technology.

Another aspect that deserves more investigation is the educational utility of this kind of system, which moves the learners away from the actual procedure, instead of enhancing the realism of the simulation. On the other hand, Liang et al. (2020) proposed a novel way to improve the realism of a stroke assessment simulation. The AR content simulates the stroke symptom animation (through 3D animated facial drooping) to be superimposed over the head of the physical manikin, and the nursing students participating in the simulation have to identify those symptoms and act appropriately to follow the procedure. Students' comments about user experience (UX) were collected (and confirmed the usefulness of this tool), but the emotional/cognitive involvement was not assessed (Liang et al. 2020).

The assessment of AR/MR applications in medical training has had a notable increase in the very last few years (2016–2020). The main metrics and parameters used to assess systems' usability, learners' performance, emotional state, and satisfaction in most of the state-of-the-art studies collected in recent reviews, can be summarized as follows:

- Demographics Information: gender, level of expertise, and previous XR experience [by questionnaires] (Margarido Mendes et al. 2020);
- Technical Assessment: voice and gesture interaction, virtual contents realism and field of view [Likert-scale questionnaire] (Chaballout et al. 2016);

- Performance: task completion time, number of errors (Margarido Mendes et al. 2020), complication rates (Gerup et al. 2020) [by author-developed Likert scale questionnaires or procedural-based checklists];
- Learning Experience and User Acceptance: easiness to identify anatomical landmarks, ease of use, learnability, usefulness, reliability, easiness for debriefing [by Likert scales + System Usability Scale (SUS) + semi-structured interview] (Gerup et al. 2020; Margarido Mendes et al. 2020);
- Emotional State: perceived workload [by NASA-TLX questionnaire + semi-structured interview] (Margarido Mendes et al. 2020), perceived cognitive load, stress response, adverse health effects, and ergonomics [by questionnaire] (Gerup et al. 2020).

However, none of these authors includes all the parameters in their simulation effectiveness assessment studies, resulting only in partial evaluations. Also, the majority of these works use only subjective assessments techniques which can reflect only a partial result. Indeed, the monitoring of physiological parameters such as heart rate (HR), heart rate variability (HRV), breathing rate (BR), electrodermal activity (EDA), etc. would enhance the reliability of the analysis related to the felt stress and cognitive load.

Moreover, the MR advanced interactive training in healthcare education still presents other weaknesses that can be summarized as follows (Zhu et al. 2014; Garzón et al. 2017; Gerup et al. 2020):

- The kind of learning theory used to guide design and application is usually not described;
- The impact of the developed MR prototypes is usually not studied;
- The analysis of ergonomics is often poor (possible cognitive overload);
- Strong evidence for improving learning is missing.

Defined metrics for the assessment of operational quality (e.g. image registration accuracy, stability, and usability) and procedural performance (e.g. checklist-based assessments) in simulated settings had to be established and validated (Kobayashi et al. 2018). It is important to verify that AR systems satisfied the real purposes of education, complementing the learning process (Garzón et al. 2017). Moreover, AR needs to be investigated more robustly (Munzer et al. 2019) since rigorous, objective measurements of clinical procedural skills and human performance metrics are very limited or absent (Linde et al. 2019). The investigation of the educational context, learner types, and learning objectives (e.g. cognitive, technical, or non-technical such as measuring situational awareness, communication, or stress coping) must be implemented (Gerup et al. 2020). There is

Table 1 Research papers about AR/MR applications in procedural training

Paper	Application	Hardware and software	Assessment	Number of participants
(Magee et al. 2007)	AR for the simulation of ultrasound-guided needle insertion procedures for interventional radiology education and training	Manikin, mock ultrasound probe, needle, a pair of magnetic 3D position sensors	Experience evaluation (questionnaire 5-points Likert scale)	60 users: 34 consultant interventional radiologists and 26 specialist registrars in radiology
(Coles et al. 2011)	AR with haptics in femoral palpation and needle insertion	Two force feedback devices, custom-built hydraulic interface, a modified Phantom Omni end effector, LCD monitor	Face and content validation. Objective feedback in a 29-point questionnaire (7-point Likert scale)	7 experts
(Kotranza et al. 2012)	MR humans for clinical breast examination	Physical breast model, a webcam, force sensors, passive manikin, a Head-Mounted Display	Study I: Efficacy study. Real-time, quantitative measures of correct pressure and pattern of search are computed from the real-time sensor data. Study II: Performance study. Coverage and pressure recorded	3 groups: 12 novice medical students, 32 novices and 25 experienced medical students, residents, and clinicians
(Gutierrez-Puerto et al. 2015)	AR central venous access training simulator	Unity3D, Vuforia SDK, Oculus VR, Webcam, Surgical tools (real and virtual)	Technical test for the recognition of the objective and its location	1 user
(Wang et al. 2017)	AR for remote procedural training as a point of care ultrasound	Hololens, Leap Motion sensor, Unity3D	Comparative study. Performance with Global Rating Scale (GRS); Utility, Simplicity and Perceived Usefulness with a short Likert survey; Cognitive Load with time to perform the task, mental effort, and task difficulty rating	25 users: 24 students and 1 mentor
(Rochlen et al. 2017)	AR for central line insertion training	Epson Moverio, BT-200® Smart Glasses, Unity3D, Vuforia SDK, Skill trainer	Performance checklist, Total time, Time to needle insertion, Survey about: level of training, previous experience, satisfaction with AR, perceptions, likes, and dislikes, potential barriers, suggestions	40 users (medical students and anaesthesiology residents)
(Kobayashi et al. 2018)	AR/MR devices for acute care procedure training	Hololens, Maya software, 3DViewer Beta, Task trainer, a wirelessly networked laptop, LCD projector	Pilot application. Future work: Educational utility and effectiveness of AR/MR-based training on live-patient clinical outcomes. Metrics: operational quality markers (holoimage registration accuracy, stability, and usability), checklist-based procedural performance assessments	40 learners
(Tai et al. 2019)	AR-driven medical simulation platform for percutaneous renal access	Personal computer, 2 Phantom Omni with one stylus	Validation of: Face and content, skills improvement, construct, and criterion. Objective metrics and Global Rating Scale questionnaire	54 professors (36 medical students and 18 urologists)

Table 1 (continued)

Paper	Application	Hardware and software	Assessment	Number of participants
(Margarido Mendes et al. 2020)	AR training assistance to pinpoint insertion of intravenous needles	Syringe, Seldinger needle, Physical Simulator of upper torso and neck, Aryzon SDK, HMD device with a slot for smartphone, Vuforia SDK, Unity3D	Comparative study. Performance through task completion time and number of needle insertion errors; Demographics questionnaire, Satisfaction questionnaire with a 6-level Likert scale; General questionnaire for face and content validity, satisfaction, and perceived workload (NASA-TLX); Semi-structured interview	18 participants (attending specialists and medical residents)

also little information about AR usability in the healthcare setting (Munzer et al. 2019). An in-depth examination of UX beyond usability, considering also the emotional fulfilment, should be considered (Cheng et al. 2013). Indeed, Salar et al. (2020) found out that emotional investment has a direct effect on focus of attention. For this reason, they suggest designing the simulation based on real-life situations, to enable the students to establish an emotional connection with the application and improving the levels of attention. Also, usability has a direct effect on students' interests. Therefore, they recommend simply designing AR applications, to increase the interest, the capture of attention, and reality perception (which generate flow in the simulation participants) (Salar et al. 2020).

Finally, a considerable shortcoming is a wide heterogeneity among research designs and outcome measurements. Establishing guidelines and standard methodologies to analyse and assess outcomes in medical education due to the use of AR technology would lead to higher-quality studies (Gerup et al. 2020).

3 Methods

3.1 How to design a medical simulation

The framework shown in Fig. 1 is proposed to support the design of medical simulations. It is suitable for different kinds of simulations, both transversal (e.g. emergency room) and specialized skills (e.g. catheter insertion, pericardiocentesis), with or without the XR integration. It consists of five interconnected modules that need to be defined according to the learning goal and the medical technique to simulate.

Before the design of the scenario, it is important to define the simulated activities that must be consistent with predetermined educational objectives. The definition of the objectives arises from the evaluation and the awareness of the real training need of the course recipients. Through the educational objectives, it is possible to plan both the training scenario and the evaluation system that will allow recognizing whether and how the participant has achieved the learning objectives.

The scenario design includes the definition of features, tasks, and feedback. Each module builds on the previous one. Three elements constitute the features module: patient characteristics, roles, and environmental characteristics. The first one refers to the profile of the patient that usually undergoes the medical treatment and the characteristics of the patient that can influence the procedure. The second one defines the active roles (clinical actors) that need to be managed in the simulation. The third one allows reproducing as faithfully as possible the hospital or outpatient environment in which the procedure is generally performed. The care of
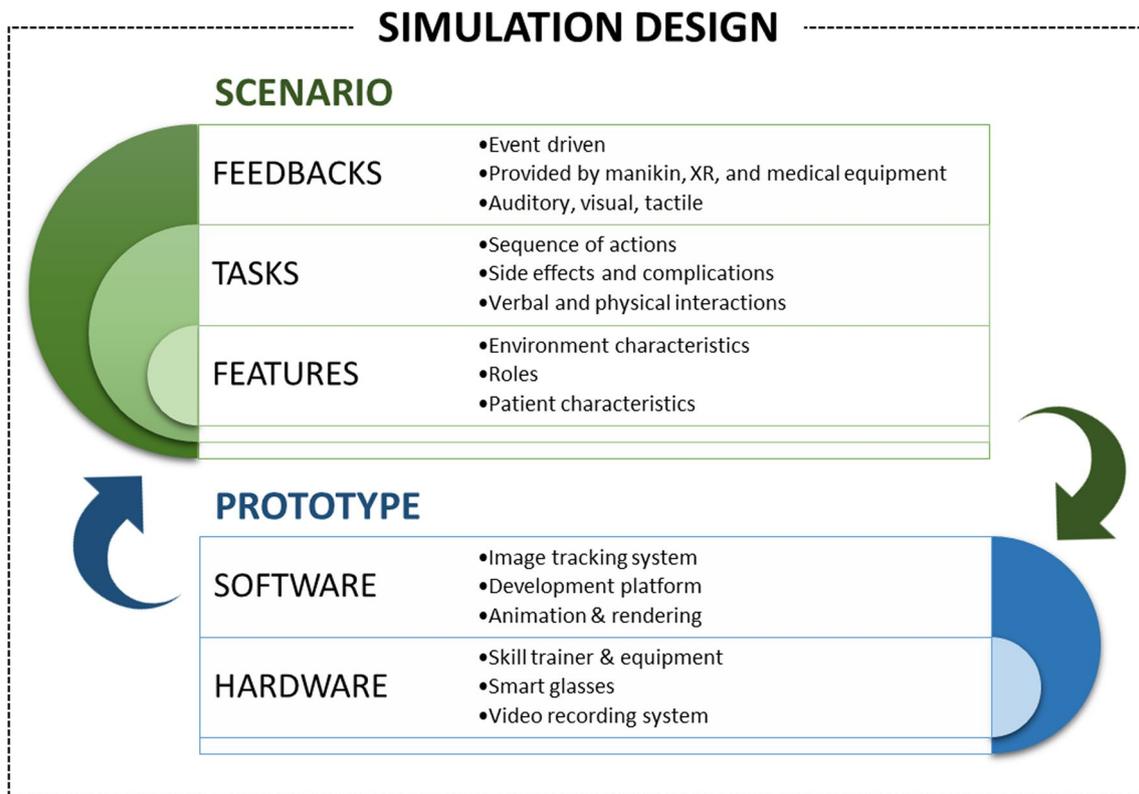
SIMULATION DESIGN

SCENARIO

FEEDBACKS	• Event driven • Provided by manikin, XR, and medical equipment • Auditory, visual, tactile
TASKS	• Sequence of actions • Side effects and complications • Verbal and physical interactions
FEATURES	• Environment characteristics • Roles • Patient characteristics

PROTOTYPE

SOFTWARE	• Image tracking system • Development platform • Animation & rendering
HARDWARE	• Skill trainer & equipment • Smart glasses • Video recording system

Fig. 1 Medical simulation design

these aspects allows guaranteeing as much as possible to reproduce highly realistic environments and situations that allow participants to carry out experiences in authentic contexts, that allow understanding how to use knowledge in real clinical settings (Onda 2012).

Then, the sequence of tasks that should be performed by clinical actors is defined, including possible side effects and complications that can arise and need to be managed. The tasks definition must accomplish specific standards based on scientific evidence and good clinical practice, with absolute consistency with the educational objective and the tool used for the performance assessment.

For each task, all kinds of interactions (human–human or human-thing) are identified, both physical and verbal. The last scenario's module aims to define all feedback triggered by events, which include task execution, complication, learner error, clinical risk, etc. For each feedback, the relative source (i.e. manikin, XR application, medical equipment) and modality (i.e. auditory, visual, tactile) need to be set. The scenario design has to define in which domain, real or virtual, each element is managed.

The prototype design, which allows simulating the designed scenario, consists of HW and SW modules. The former includes manikin and equipment suitable for the simulation purpose, the XR device, and the

video-recording system. The latter, based on the selected HW, provides the definition of the XR development platform, the tracking system, and the animation and rendering SW.

Once all the aforementioned items have been defined, it is necessary to focus on how to convey information in XR interfaces to the user. XR interfaces differ from traditional graphical user interfaces (GUI), by employing new ways of interaction (e.g. gestures, voice commands) and unconventional layouts (e.g. holograms, anchors to physical objects). However, no specific standards exist in the literature (Gattullo et al. 2020). The XR interface design implies a choice among various visual assets based on the information purpose and the characteristics of the XR device. Visual assets include text, video, photo, 2D/3D models (e.g. objects, persons), and auxiliary models (i.e. 2D or 3D graphic elements for auxiliary instructions). Identity, location, orientation, order, and notification are some information related to the visual assets that need to be defined.

The iterative approach of the framework allows the prototype design to review the scenario and so on until a satisfactory level of simulation design is achieved to pursue the learning goal.

3.2 How to assess the effectiveness of a medical simulation

The optimization of cognitive states is necessary to prevent disorders and stressful conditions and promote the best learning and performance (Pheasant 1999). For this reason, the present work proposes a structured methodology for the assessment of the effectiveness of medical training simulations, comprehensively considering performance, cognitive and emotional states of learners.

The proposed methodology has been designed to be adopted in every simulation context, from the most traditional ones (e.g. with actors simulating patients) to the most advanced ones (e.g. with high-fidelity manikin simulators or using XR applications).

To perform an overall and structured analysis of simulation effectiveness, the methodology includes different assessment methods: from the skills and performance analysis to the subjective self-assessment of perceived mental and emotional states, to the monitoring of biometric parameters to accomplish a more objective analysis of stress and cognitive load (Fig. 2).

For the performance evaluation, based on the simulation type and content, a specific checklist must be prepared to discriminate among correct, incorrect, and not performed tasks, consistently with the educational goal. Also, for each task, the completion time, the number of errors, the number of consultations, the number of attempts, and other simulation-related aspects should be evaluated. The performance assessment takes place during the simulation and debriefing. A survey to assess the improvement in acquired skills should be administered before and after the simulation session.

Concerning the self-assessment, the defined methodology requires the administration, before and after the training session, of three different kinds of surveys:

- The *Aptitude to simulation and technology* refers to the personal aptitude towards the simulation-based training and familiarity that subjects have with the use of technology, defined as the application of information technology and telematic devices (i.e. the experience that they have with technological devices from a common smartphone to more sophisticated tools such as HMDs or haptic gloves) (Brunzini et al. 2020). It is administered before and after the training to understand how the simulation and use of XR influence the learners' opinions.

- The *State-Trait Anxiety Inventory* (STAI) is a scale used for the assessment of anxiety. It consists of two modules, each one composed of 20 statements. The STAI Y-1 form allows assessing the subject's current state of anxiety, considering feelings of apprehension, tension, worry, and nervousness, in that precise moment; the STAI Y-2 form allows evaluating the individual tendency to anxiety, giving an index of how a subject generally feels, and it can be used to identify people predisposed to develop anxiety in stressful situations. For each statement, the subject must answer on a 4-point Likert scale (Spielberger et al. 1983). Both forms must be administered before the simulation session, while only the STAI Y-1 form is sufficient after the training to understand the effect of the simulation on the perceived anxiety.

- The *Numerical Analogue Scale* (NAS) is used for the assessment of the perceived level of stress. It consists of a bar or line divided into ten intervals, numbered from 0 to 10. The subject is asked to select the integer number that best reflects the intensity of his/her stress ($0=$ no stress, $10=$ very strong stress). The NAS has proved to be a valid, effective, and easy-to-implement tool for the rapid assessment of perceived stress (Lesage et al. 2012). Before the simulation, it allows defining the "baseline" of perceived stress before undergoing the activity; after the training, it highlights the stress related to the simulation.

Moreover, after the training activity, also the NASA-Task Load Index (NASA-TLX) questionnaire is administered. This is a subjective scale, developed to minimize the variability of assessments between subjects (Hart et al. 1988), consisting of questions that allow evaluating the importance assigned by the subject to different elements involved in the perceived workload. Thus, it allows the assessment of the perceived mental demand needed to perform the activity, as well as the physical demand or the emotional states related to stress such as perceived effort and frustration.
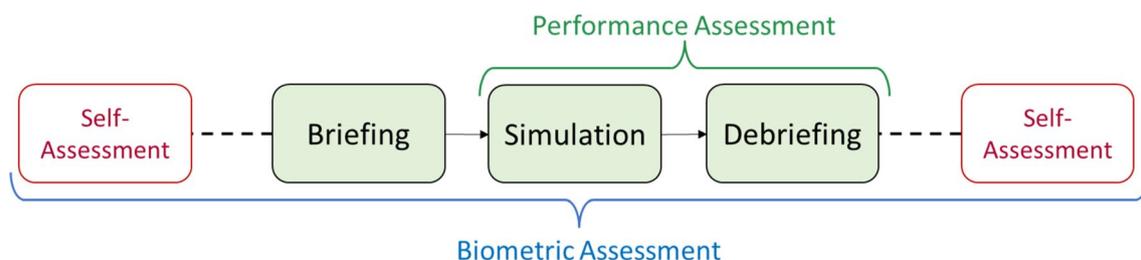

Fig. 2 Methodology to assess the effectiveness of medical simulations in terms of performance and cognitive and emotional conditions

In the case of training sessions with XR devices and applications, ad hoc usability and UX surveys could be provided to the learners at the end of the activity.

To have a more objective assessment of the cognitive conditions, even the physiological parameters (such as HR, HRV, BR, EDA, …) should be monitored and collected. Learners should wear non-invasive smart devices (e.g. chest bands, wrist bands, …) from their arrival in the training room until the end of the post-training self-assessment. In this way, it is possible to discriminate the parameters' variations among different stressful, restful, and mentally demanding situations.

A video camera for the recording of the training session should be provided because during the data analysis it could be useful to stopwatch and track events in relation to physiological variations and times.

4 Design and development of a mixed reality simulator

4.1 Scenario design

The first step of the medical simulation design is the definition of the learning goal. In this case, based on the analysis of the Core Curriculum and of the Skills approved by the Permanent Conference of Presidents of the Degree Courses in Medicine and Surgery of Università Politecnica delle Marche (Adrario et al. 2017a, b) the learning goal was to simulate a rachicentesis for diagnostic purposes in adults. Rachicentesis, or lumbar puncture, is a surgical technique to sample cerebrospinal fluid (CSF) to be analysed for diagnostic purposes or drained in case of liquor excess. The procedure involves introducing a needle into the space between the arachnoid meninx and the pia mater, at the L3/L4 level or L4/L5 level. The lumbar puncture can be performed with the patient in the lateral foetal position or sitting position.

Studies in the literature affirm that it is possible to reduce more than ten times the number of patients who have a headache after rachicentesis, the days of stay, and the costs for the health service, by changing the needle type and the way lumbar puncture is done. According to experts, this reduction in inappropriate hospitalizations can lead to 3 million euros savings in Italy, as well as to a great reduction of avoidable pain, and fear of lumbar puncture (Bertolotto et al. 2016). However, the innovative procedure proposed by Bertolotto et al. requires greater dexterity than the traditional one, and the technical skills result crucial. For this reason, it was considered appropriate to proceed gradually, in line with the basic principle that the complexity of each simulation course must be graduated on the student. The simulation training responds well to the need that invasive manoeuvring requires. In fact, through simulation courses, the student is

able to become familiar with the technique and procedure to achieve a minimum standard level of effectiveness without the related clinical risk and thus safeguarding the patient's health.

In this case study, the simulation scenario of the rachicentesis procedure is schematized in Fig. 3.

The main active roles are the following two: the anaesthetist, simulated by the learner, and the patient, simulated in a hybrid manner (real and virtual) by the manikin and the body virtual model. Different physical characteristics of the patient such as normal body mass index (BMI), obesity, and pregnancy are virtually simulated. Given its nature, the rachicentesis cannot involve actors simulating patients that is a successful strategy in medical education (e.g. blood sampling procedure). For this reason, it is an excellent candidate for XR applications. Furthermore, the same framework designed and developed for the rachicentesis can be easily adapted to similar medical procedures such as pericardiocentesis and thoracentesis.

The simulation environment is a classroom, at the Faculty of Medicine of Università Politecnica delle Marche, equipped with the abdomen skill-trainer lying on a desk and the following instruments: needle, test tube, latex gloves, sterile cloth, and sterile gauze. A typical layout is shown in Fig. 3.

The last step consists of the definition of tasks and the relative elements in the real and virtual domains. They include:

- Verbal interaction, that mainly refers to the explanation of the procedure to the patient by the learner.
- Physical interaction between the learner, the skill trainer, and the equipment.
- Tactile feedback provided by the skill trainer during the anatomical palpation.
- Visual feedback in the real and virtual domain. The former refers to the skill trainer and more precisely to the CSF leakage. The latter includes the position of the body virtual model and the simulation of patient spasm.
- Auditory feedback, which simulates the patient voice and complaints.
- Such a scenario demonstrates how the proposed MR application does not aim to facilitate the procedure execution but to "immerse" the trainee in a more realistic environment, influencing his/her emotional state as in real practice.

4.2 Prototype design and development

The development of the MR prototype consisted of six main steps (Fig. 4).

The first one is the definition of the target images, which were created using an online free tool that exploits features easily recognizable by *Vuforia*. The symbols of a patient
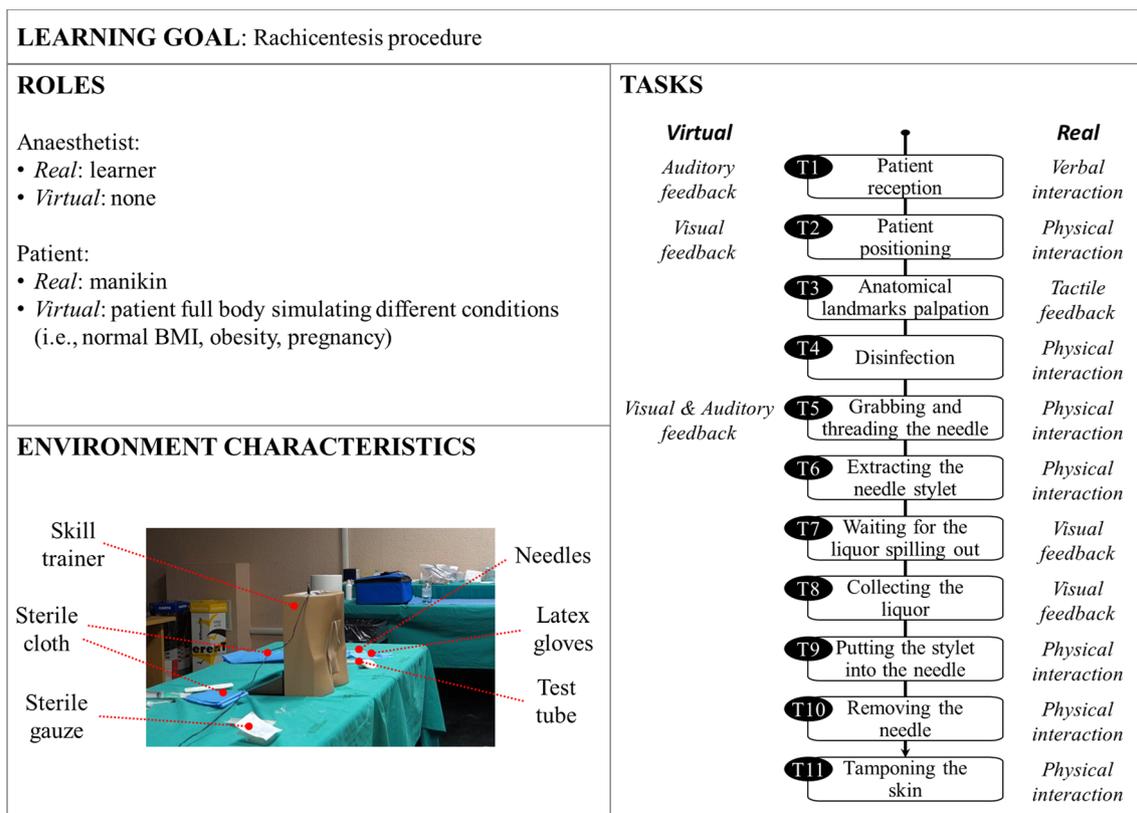
Fig. 3 Simulation scenario of rachicentesis procedure

and a doctor were superimposed on the target images, to facilitate their recognition to the user.

The images were then uploaded in *Vuforia SDK* by specifying the real dimensions of the physical targets. It allows generating the libraries to be imported into *Unity,* by *Unity Technologies,* for tracking and recognizing targets by the AR/MR glasses. A.unitypackage file and a string for the license were then downloaded and imported into *Unity*, where the targets can be associated with the 3D models.

For the development of the digital patient, a 3D model of a woman and the texture were, respectively, downloaded in .obj and .mtl file formats, from an anthropometric database (as suggested in Paul and Scataglini 2019). The virtual patient posture was modified from the traditional one (upright, open arms) to left lateral decubitus and sitting positions, which are required to perform the lumbar puncture. This repositioning was performed in the open-source *Blender* software, through the "rigging" procedure.

In the texturing phase, the 3D model surface was smoothed to increase the fidelity to reality. In particular, a "smooth" filter was used on the 3D model to round the edges of the mesh. This filter allows changing the normal of each surface as it gets closer to the edge; in this way, the 3D model surface is smoothed, and the user does not

have the impression of watching a polygonal model. Then, the model animation was created by defining the times and movements of the skeleton. It allows simulating the spasm due to needle insertion. The output of this step is a .fbx file that can be imported in *Unity*.

While using the MR application, the virtual model must accurately overlap the real manikin, without appearing either too large or too small. Indeed, in the first case, the virtual model would appear disproportionate and therefore unrealistic. In the second case, some parts of the skill trainer could remain visible, thus worsening the user's sensation of immersion and realism.

To ensure the perfect dimensioning, the skill trainer (i.e. the *Lumbar Puncture Skills Trainer S411* by *Gaumard®*), which reproduces anatomical features, provides realistic tactile feedback, and lifelike needle resistance combined with a fluid pressure system, that allows the liquor spilling out and collection), after the application of markers on its surface, was scanned and a virtual counterpart was created. For the scanning process, a non-contact 3D laser scanner (*Range 7* by *Konika Minolta®*) was used.

The previous outputs were put together in *Unity*, to create the main scene of the application that contains the following elements (Fig. 5):
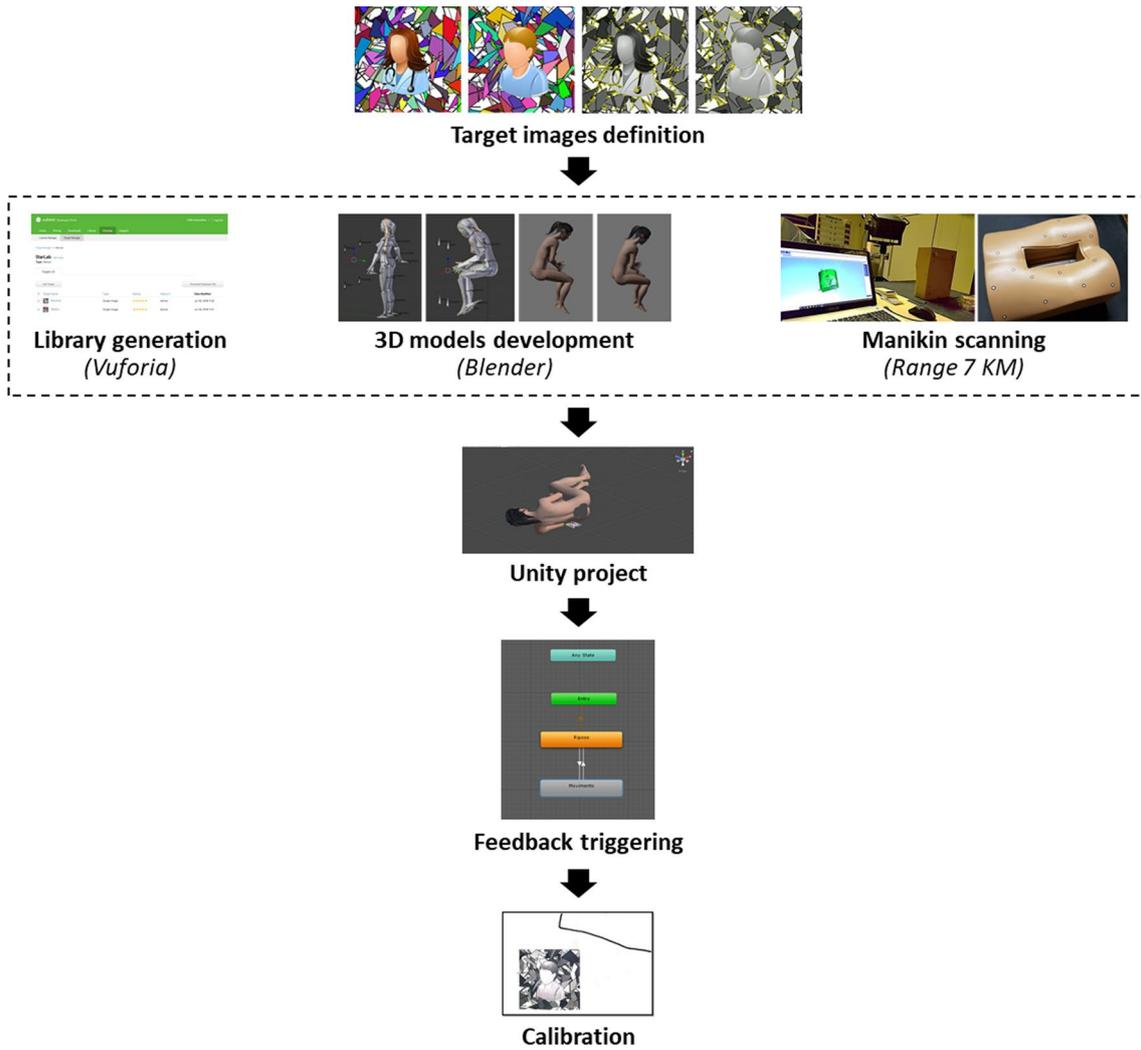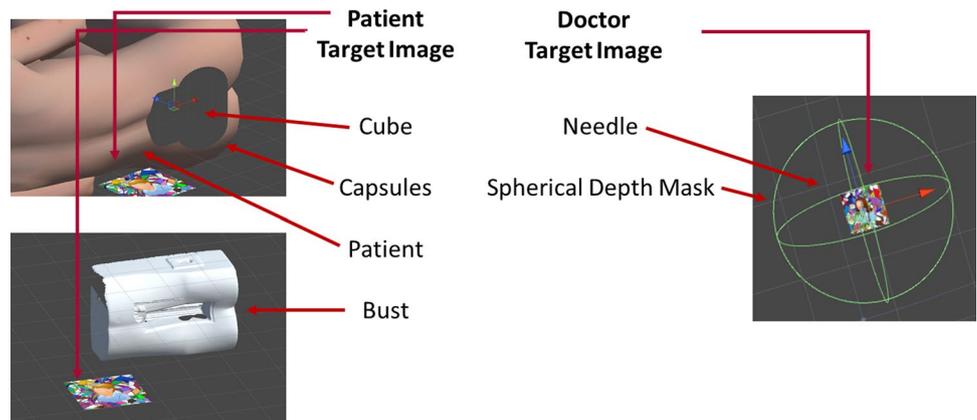
Fig. 4 Workflow

Fig. 5 Unity project

- Patient Target Image, which is the 2D target image, integral with the manikin and fixed on the desk, represents the patient. For this tracker, the "extended tracking" option is activated. This means that, even if the target image is no longer in the user's field of view, the system, based on the available sensors, hypothesises its position

and continues to overlap the virtual model to the manikin. It allows avoiding the target image goes outside the student's field of view when he/she focuses on the manikin. The Patient Target Image has the following children:

- Cube: a parallelepiped, which is equipped with its own "Box Collider", hidden within the virtual model of the patient, immediately behind the operation area. It allows enabling the movements triggering.
- Bust: the virtual model of the manikin used to correctly dimensioning and positioning the 3D patient model.
- Capsules: two capsule-shaped elements positioned in correspondence with the operation area. They are rendered as "Depth Mask" so that the user can see through them, while the application is running. Their function is to ensure that the student can always see the operation area on the real skill trainer (the rest of the manikin is not visible, since it is covered by the virtual patient).
- Patient: 3D model of the patient in the left lateral decubitus or sitting position.

- Doctor Target Image that it is the 2D target image that represents the doctor. It is smaller than the other target to be fixed to the student's hand during the procedure. It has the following children:

  - Needle: a spherical collider located in correspondence with the doctor target. A radius of 200 mm was selected because it represents the average distance between Doctor Target and the tip of the needle held in the student's hand during the rachicentesis procedure. It allows enabling the movements triggering.
  - Spherical Depth Mask: a depth mask centred on the doctor target, and therefore integral with the student's hand, that allows the student always to see his/her hand during the operation. Otherwise, the hand could be covered by the virtual image of the patient.

To enhance the sense of realism, an animation of the scene has been provided. According to a C# script developed in *Visual Studio* by *Microsoft*, the patient spasm movement is triggered by the contact between "Needle" and "Cube" colliders that occurs when the needle touches the manikin during the simulation.

The animation is activated by the "move" Boolean parameter of the animator. The animator is a file in *Unity*, which represents the sequence of states present in the scene. The used animator consists of two main states: the "Rest" state, and the "Move" state. As soon as the application is started, the "Rest" state is activated. In this state, the 3D model is stationary. The transition to the "Move" state occurs when the "move" Boolean parameter is set to true; the inverse transition occurs when "move" is set to false. "Move" is set to false by the script, when the needle comes out of the cube (i.e. when it comes out from the manikin). This condition is necessary to prevent the patient from moving in a loop when the needle has been inserted into the manikin. Moreover, a 3-s cooldown prevents the animation from re-run before the previous one is finished. It may happen with uncertain movements in inserting the needle or with several repeated attempts. In addition to the spasm of the virtual patient, audio files reproducing the patient's laments and complaints due to pain are randomly played.

To ensure that the virtual patient is precisely superimposed over the real skill trainer, the Patient Target Image must be correctly positioned with respect to the manikin. Therefore, a simple "calibration sheet" has been created starting from the Patient Target Image. This sheet allows to correctly position the manikin above the relevant profile and thus to have the correct alignment between skill trainer and virtual patient.

This MR application (Fig. 6) has been developed to be used with different HMDs.

5 Effectiveness assessment

The effectiveness of the MR prototype for rachicentesis has been assessed through a pilot study, using two different devices: *Vox Gear Plus*® and *Microsoft Hololens*®. The former is a low-cost VR headset to be used with a smartphone. Setting the camera of the smartphone in a "see-through" manner, it is possible to use the device also for AR applications. However, the loss of depth perception, due to the lack of stereoscopic vision, made it impossible to safely complete the simulation with that device.

Hololens is an advanced HMD for MR. It provides a holographic experience thanks to its see-through holographic lenses and manages different user's inputs, i.e. gaze, gesture, and voice.
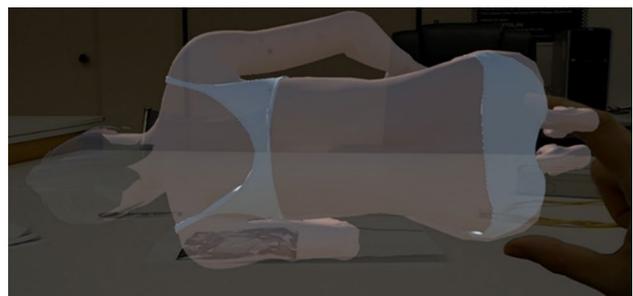
Figure 7 shows the MR simulation trial configuration.

Fig. 6 Prototype

Fig. 7 A couple of students, equipped with wearable sensors (Empatica E4 circled in blue and Zephyr BioHarness circled in red), performing MR rachicentesis simulation with Vox Gear Plus (**a**) and Hololens (**b**)
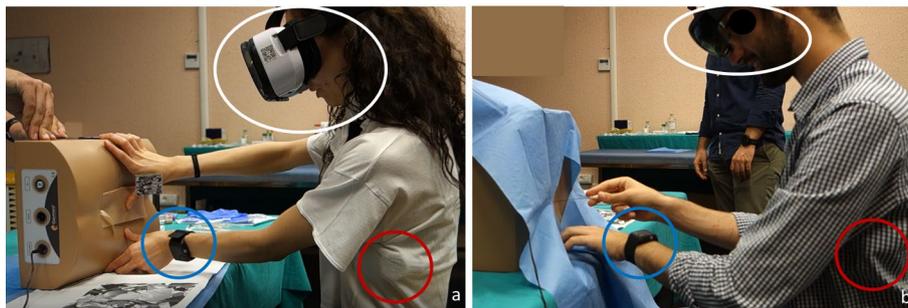
Table 2 MR simulation—Characteristics of participants

	% of Participants
Male	55.55%
Female	44.45%
Smoker	33.33%
Pertinent thesis	83.33%
Previous experience in:	
Group simulation	83.33%
Residency	50.00%
Professional training activity	94.44%
Training course (e.g. BLS, ATLS, …)	55.55%
Working experience	11.11%
Health volunteer activities	11.11%
Invasive procedure	38.89%

To compare the results with a reference, students traditionally performed the simulation also, with the partial manikin without using the AR application. This is useful to understand the effect that the greater immersion and higher realism, given by the MR, have on the cognitive conditions of participants. It is expected that the interaction with the simulated patient (i.e. the auditory and visual stimuli that the digital patient gives to the learner in response to his/her action) can increase the level of stress.

5.1 Participants

Twenty students of the 6th year of Medicine and Surgery of Università Politecnica delle Marche were randomly selected for the pilot study. After the explanation of the trial and the eventual contraindications in the use of HMDs and wearable devices for biometric signals monitoring, eighteen of them agreed in being enrolled in the trial. They signed the informed consent and the form for the processing of personal data. They were, on average, 25.6 ($\pm$0.6) years old, and their characteristics are summarized in Table 2.

From Table 2, it is important to highlight that 38.89% of them have already had previous experience with invasive procedures and 83.33% of them was working on a degree thesis pertinent to the rachicentesis (i.e. anaesthesia and

intensive care, cardiology, emergency, gastroenterology, obstetrics and gynaecology, orthopaedics, pneumology, radiology, urology and every kind of surgery).

5.2 Workflow/protocol

At the arrival in the classroom, the eighteen participants were divided into three groups of six students each. While one group performed the simulation using *Hololens,* another one used *Vox Gear Plus,* and the other one accomplished the simulation without HMD (low-fidelity). In turn, each student experienced all three kinds of simulation. In each group, while one subject was doing the rachicentesis, the other five participants were in the same room, observing the performance. The execution sequence was always different, randomly established, and not revealed, to avoid eventual influences on students' anxiety, stress, and cognitive states.

The four steps technique (Peyton's four-step) for the training of practical skill (George et al. 2019) was followed. Firstly, the teacher performed the manoeuvre without comment or explanation. Subsequently, the teacher slowly re-showed the procedure, explaining each manoeuvre and dwelling on the most complicated or decisive steps for the effectiveness of the procedure, justifying its correctness. At this stage, learners can ask the questions raised by the demonstration. In the third step, the students illustrated the manoeuvre performed by the teacher, to focus on the theoretical aspects without worrying about the practical-manual component. In the last step, the student performed the procedure, thus combining theory and practice.

For the assessment of emotional and cognitive conditions, the protocol in Fig. 8 was followed for the simulations with *Hololens* and without HMDs, but first, at the arrival in the classroom, also other questionnaires were asked to be answered. Before the beginning of the simulation sessions, the Numerical Analogue Scale (NAS 0) was administered to record the basal level of perceived stress, and the State-Trait Anxiety Inventory (STAI) Y2 was administered to analyse the anxious trait of students in their lives. The questionnaire about the aptitude to technology was administered before and after the completion of both simulations to analyse the
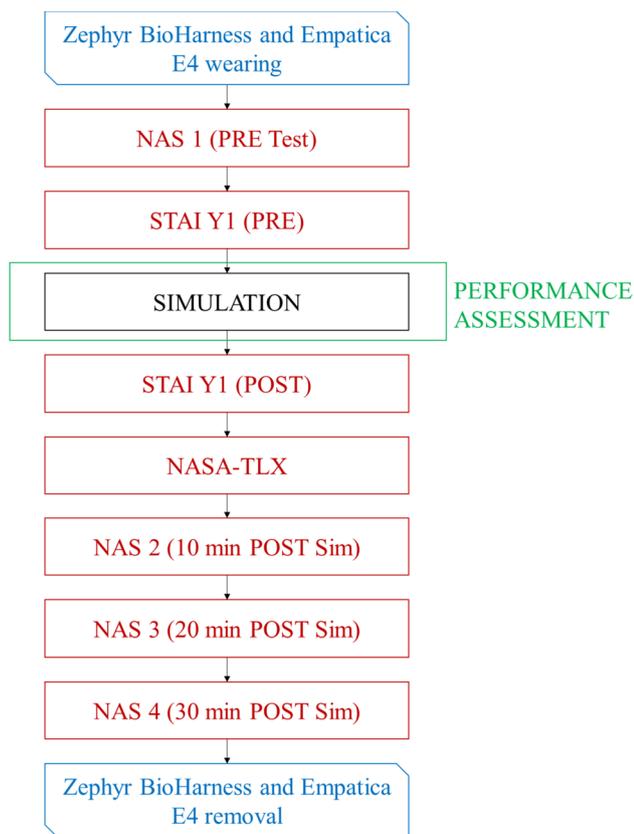
Fig. 8 Effectiveness assessment protocol

variations in students' opinions on the usefulness and appreciation of using advanced technologies.

The protocol consists of a structured combination of self-assessment questionnaires, biometric monitoring through the wristband Empatica E4® and the chest band Zephyr BioHarness®, and skills and performance evaluation.

NAS was asked before (NAS 1), and ten (NAS 2), twenty (NAS 3), and thirty minutes (NAS 4) after the end of each simulation session (the one without HMDs and the one with *Hololens*). The State-Trait Anxiety Inventory (STAI) Y1 was answered before each simulation to assess the perceived level of anxiety in that precise moment, and then, after the performance, to understand the variation in perceived anxiety, due to the simulation itself. The NASA-TLX was administered after each simulation to assess the perceived workload (i.e. physical, mental, and temporal demands, effort, performance, and frustration).

The collection of physiological parameters has been done during the entire session.

A performance checklist was used to assess each task as 'correctly performed', 'incorrectly performed', or 'not performed'. Other relevant data such as times (i.e. patient preparation time, time to succeed, total time), number of errors (i.e. touching the needle in the wrong place, do not re-insert the stylet, coming in-and-out with the needle), number of attempts to succeed, and teacher interference were also recorded and assessed. Also, to analyse the learning of skills, a 5-item questionnaire was administered before and after the simulation.

6 Results and discussion

Psychophysiological results refer to fifteen participants. The results related to three students have been discarded because they did not complete the subjective surveys and/or their physiological data were affected by errors. Since it was impossible to accomplish the simulation with *Vox Gear Plus*, the psychophysiological assessment and performance refer to the traditional simulation and the one performed with *Hololens*. Instead, the UX has been analysed also for the use of *Vox Gear Plus*.

6.1 Psychophysiological assessment

6.1.1 Self-assessment

First, the participants' familiarity with advanced technological devices has been assessed before the simulation sessions to understand their attitude towards the use of technology. More than 72% of enrolled students were familiar with the use of everyday life technological devices such as computers, tablets, and smartphones, while only 16.67% have used haptic gloves for gaming. However, more than half of participants (55.56%) were used to play with simulation videogames and/or serious games and were familiar with the use of HMDs. Regarding the use of smart wearables for healthcare and lifestyle, only 16.67% were familiar with them. However, 61.11% of students felt suitable to work in a high-tech environment, and only 22.22% of them would feel stressed to work in it.

Then, a comparative evaluation between students' points of view, before and after the MR training, has been done to analyse the perceived usefulness of AR in the simulation context.

After having performed the rachicentesis with the digital responsive patient, students' opinion about the utility of providing feedback through technological devices decreases a bit. However, it should be kept in mind that in this case the MR application was thought to improve the sense of realism and immersion and it was not designed to give informative feedbacks or to make the simulation easier. Indeed, participants continue to think that technological devices and multisensory interaction are valuable tools for learning, even after the simulation. Overall, even if they felt a little bit stressed to work in a high-tech environment, they also feel suitable to work in it, confirming the practice with physical manikin

as the simulation that would benefit the most by the use of high-tech devices, together with the understanding of human anatomy, physiology, and pathology.

Concerning the subjective self-assessment about the perceived anxiety, stress, and cognitive conditions, before and after the simulation sessions, Table 3 summarizes the main results related to the traditional simulation without virtual contents and the rachicentesis performed in MR.

On average, participants are more anxious in their life than in the classroom before the simulations (STAI Y2 (45.17) > both STAI Y1 PRE). The mean STAI Y1 POST agrees with the value expected for college students (Spielberger et al. 1983), while the mean STAI Y1 Pre is few points higher. Thus, perceived anxiety decreases after the simulation (mean STAI State Post < mean STAI State Pre) in both cases, even if the one related to the MR simulation is a bit higher, highlighting the increment of anxiety due to the interaction with the digital patient.

The mean values of NAS, on a 10-point Likert scale, reflect the perceived stress at the arrival in the simulation room (NAS 0), ten minutes later, and ten, twenty, and thirty minutes after the end of each simulation (respectively NAS 2, NAS 3, and NAS 4). In both cases, the peak in NAS 2 corresponds to the perceived stress during the rachicentesis training. However, in the MR simulation, even if the feeling of stress constantly decreases from NAS 2 to NAS 4, thirty minutes after the end of the simulation the perception of stress is still high. Indeed NAS 4 is higher than NAS 0 (2.34). A debriefing should be suggested, to keep the stress perceived by the students under the basal level (NAS 0). Moreover, mean NAS 2, NAS 3, and NAS 4 in MR simulation are higher than in the traditional one. This could be an index of the enhanced realism and immersion

in the simulation that causes an increment of perceived stress through the interaction with the digital patient, who responds to the learner's actions.

Also, the perceived workload is higher for the MR simulation than for the traditional one. According to literature's score interpretation (Sugarindra et al. 2017), the mean perceived workload for the simulation in MR is on the inferior boundary of the high-range (high for 50–79 total score), while it is a bit lower for the traditional training. Thus, the use of high-tech devices seems to not cause high increments of perceived workload. However, the use of the HMD incremented the perceived physical demand, effort, and frustration (maybe due to the interaction with the digital patient who complains and has spasms during the procedure). In both cases, the performance received the greatest weight.

Lastly, t-test analysis has been performed to assess if the differences between the mean values of perceived anxiety, stress, and workload, in the simulations with and without MR, depend on the used MR application. Unfortunately, only the differences in the perceived effort resulted statistically significative with *p-value < 0.05*. However, this analysis would benefit of a greater sample of participants.

6.1.2 Biometric assessment

The physiological parameters have been analysed through the proprietary algorithm and SW module for cognitive states detection by *Phasya s.a. (Seraing, Belgium)*. This allows the discrimination of stress and cognitive load (CL) on six different levels. Information about the *Phasya*'s algorithm for stress and CL identification is protected by a non-disclosure agreement; their core SW module for the drowsiness assessment is described in François et al. 2016 and Stawarczy et al. 2020.

Stress and CL levels were computed for simulation phases. Figure 9 shows the trends in stress and CL levels during the simulation phases both for the traditional and the MR simulations. In the MR simulation, the cognitive load is higher than the stress during the puncture execution. It is worth noting that the cognitive load increases from preparation to puncture phase and then decreases towards the end of the simulation until the rest after the training. Conversely, the stress increases from the basal level before the simulation to the preparation phase and then constantly decreases during the following phases. The stress and CL levels after the training on average turn back to the basal levels (even if debriefing is not performed). On average, stress and CL during the execution of the procedure in MR can be considered medium, not high (on the 6 levels scale).

Preliminary observations can be done on the difference in stress and CL relationship between traditional and MR simulations. While for the rest period before the simulation and the preparation phase stress is higher than CL in

Table 3 Mean values of perceived anxiety (STAI), stress (NAS), and workload (NASA-TLX + 6 items) before and after the traditional and MR rachicentesis simulation

	Traditional simulation	Mixed reality simulation
STAI Y1 PRE	41.06 ± 10.98	40.81 ± 10.14
STAI Y1 POST	36.66 ± 10.17	37.53 ± 9.69
NAS 1	2.79 ± 1.27	2.41 ± 1.23
NAS 2	3.31 ± 1.98	3.41 ± 1.84
NAS 3	2.62 ± 1.14	3.06 ± 1.92
NAS 4	2.33 ± 1.08	2.71 ± 1.93
NASA-TLX	48.91 ± 21.15	50.41 ± 19.52
Mental demand	201.65 ± 128.19	188.50 ± 126.82
Physical demand	45.88 ± 53.40	59.61 ± 68.51
Temporal demand	103.12 ± 107.26	108.28 ± 116.26
Performance	221.18 ± 128.25	203.50 ± 120.56
Effort	110.24 ± 79.90	136.67 ± 88.30
Frustration	51.53 ± 100.93	59.67 ± 103.81

Fig. 9 Stress and CL per simulation phases during traditional (left) and MR (right) simulations

both simulations, from the lumbar puncture phase to the rest after the simulation, their relationship is inverted. Indeed, in MR simulation, conversely to the traditional one, during the last phase of the simulation and the rest period after the simulation, stress is higher than CL. This could be an index of the fact that the use of MR, and the enhanced realism, causes an increment of stress (also confirmed by the NAS). However, even with the use of AR, it remains stable on a medium level.

6.2 Performance

The assessment of the skills acquisition has been done through the five-question survey administered before and after the rachicentesis simulation. The mean number of correct answers increased from 3.68 ($\pm$0.98) to 4.70 ($\pm$0.54) after the training.

Performance in traditional and MR simulation was similar. The main results are summarized in Table 4.

More than 90% of participants were able to correctly perform the puncture and make the liquor pouring out, in both cases, with less than three attempts. The most committed error was to completely come in-and-out with the needle during the procedure (for hygienic reasons, the needle should be extracted only once at the end of the procedure; even if it has been inserted in the wrong position, the learner should try to improve the needle incline without extracting it). Also, some students touched the needle where it is not allowed but no one extracted the needle without re-inserting the stylet.

In both cases, most students (more than 84%) correctly performed the tasks. The percentage of students who incorrectly executed the tasks is very low (under 5%), while the percentage of participants who did not perform some tasks was between 7 and 10%. Thus, students understand how to execute the tasks but sometimes they forget some steps.

However, even if results between the two different kinds of simulation are similar (indeed AR is not used to give informative feedback (e.g. tips, suggestions, etc.) and help

Table 4 Performance results		Traditional	Mixed reality
	SUCCESS	93.84%	94.45%
	Correctly performed	84.92%	88.77%
	Incorrectly performed	4.87%	3.98%
	Not performed	10.21%	7.25%
	Number of attempts	2.91 ($\pm$3.98)	2.21 ($\pm$2.03)
	ERRORS		
	Came in-and-out with the needle during the procedure	30.23%	27.78%
	Touched the needle where it is not allowed	4.97%	5.56%
	Extracted the needle without re-inserting the stylet	0%	0%
	TIMES		
	Preparation time	1.08$\pm$0.94 min	1.16$\pm$0.50 min
	Time to success	3.20$\pm$5.05 min	3.15$\pm$2.87 min
	Time to finish	1.04$\pm$3.04 min	0.92$\pm$1.79 min
	Total time	5.26$\pm$4.81 min	5.92$\pm$3.09 min

the students, but only to improve the realism), the t-test shows statistical significance (*p-value < 0.01*) for the performance improvement from the traditional to the MR simulation.

6.3 User experience

The observation of all kinds of simulations performed by students allowed identifying several usability issues related to the MR system to deal with. The use of *Vox Gear Plus* makes impossible the execution of the procedure because of the loss of depth. The smartphone is not equipped with stereoscopic 3D cameras; therefore, users saw the same duplicated images in their eyes, losing the sense of depth.

Users also experienced some problems with patient tracking. When the patient's target went out of the user's field of view for long periods, the 3D digital model tended to misalign or disappear. It is mainly due to the limited accuracy of the smartphone tracking technique, which estimates the digital patient's position based on the movements of the user's head through accelerometers.

The use of *Hololens* allowed solving these problems but has highlighted new ones. They are equipped with depth sensors that offer stereoscopic vision and allow reconstructing a detailed mesh of the surrounding environment, which ensures an optimal tracking of fixed elements in space. At the same time, the following limitations were observed:

- Narrow Field of View (FOV), which is only $30° \times 17.5°$. To see the complete 3D digital model of the patient, users often forced their heads away from the manikin reducing the sense of realism and immersion.
- Needle Tracking, which was slower and less responsive than with *Vox Gear Plus*.

At the end of the three kinds of simulation (without HMDs, with *Hololens*, and with *Vox Gear Plus*), a semi-structured survey for the assessment of UX was administered

to the students. Figure 10 shows the main results of closed-ended questions by comparing the score related to both devices. In particular, the immersion level and field of view were evaluated according to a 5-point Likert scale, whereas for the other items a boolean option (yes or no) was provided. In any case, the students were asked to take the low-fidelity simulation (without virtual contents, only with the manikin) as a reference. For example, the first statement was "AR use has increased the feeling of immersion and involvement compared to the use of the manikin alone". From the immersion point of view, which is one of the main objectives of the proposed MR application, the results obtained are satisfactory, especially with the *Vox Gear Plus*. Considering the procedure execution, few students perceived benefits in terms of extra checks and actions allowed by MR or errors reduction. On the contrary, MR seems to make the operation more difficult to perform. It is mainly due to the technological limits mentioned above. Regardless of the device, visual feedback has been very successful (more than 90% of the students found it effective) compared to auditory feedback.

The second part of the survey consisted of open-ended comments on all three simulations. The results were elaborated to identify the main advantages and drawbacks of the MR application and devices.

Table 5 summarizes the main pros and cons of *Vox Gear Plus*. Most of the pros declared by users come from the impression of being truly immersed in the digital world while being able to actively interact with the real surrounding environment. On the other hand, users' comments confirmed the technical limitations related to depth, coordination, and tracking that have been observed during the usability evaluation. Moreover, they highlighted discomfort due to the feeling of nausea and the perception of the viewer as an obstacle. The former is a typical problem of all the XR systems and is due to the delay (latency gap) between what our eyes see and our movements.

The pros and cons of *Hololens* are summarized in Table 6. The perception of three-dimensionality was the main
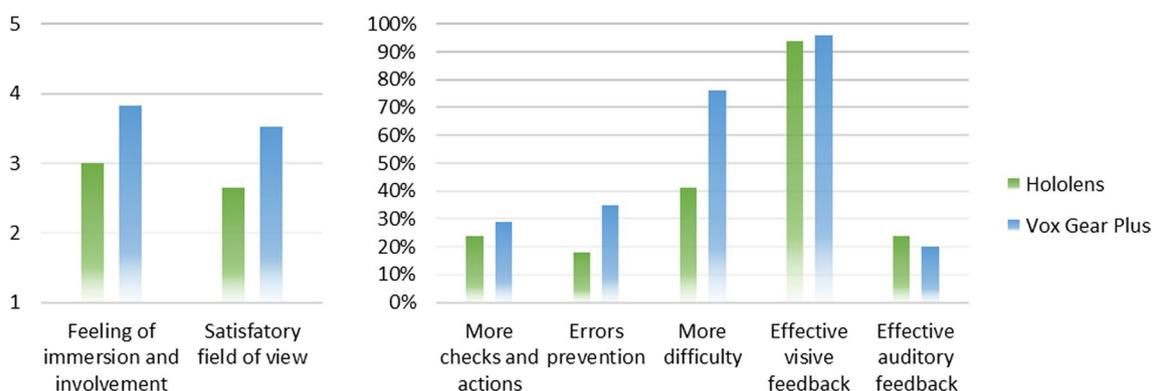

Fig. 10 User experience about the use of MR in simulation training

Table 5 Pros and cons of Vox Gear Plus

Vox Gear Plus pros	Vox Gear Plus cons
Feeling of operating on a real patient	Loss of depth
Very immersive and realistic	Difficult eye-hand coordination
Actively interaction with the surrounding real environment	Limited accuracy for patient tracking
Attention also shifts to the patient with whom interact	Device obstruction
Light, ergonomic, and comfortable device	Nausea and dizziness

Table 6 Pros and cons of Hololens

Hololens pros	Hololens cons
The vision of the whole body helps to orientate the user	Attention was sometimes caught more by MR itself than by the procedure
Feeling of three-dimensionality	It is initially difficult to get used to
Realistic patient engagement	Narrow FOV
Ergonomic device	Slow needle tracking Device physical bulk

advantage mentioned by users. It also positively affected the quality of whole-body vision that improved orientation. At the same time, all users agreed on the narrow FOV, which focused the procedure on a specific area, forcing the user to move his/her head to see the rest of the digital patient's body. Even in this case, despite the lightness and ergonomics of the device, it is perceived as an obstacle.

7 Conclusions

The paper proposes a structured approach to design an MR-based medical simulation and evaluate its effectiveness in terms of learning experience and performance. The application aims at increasing the students' level of immersion during the simulation of the rachicentesis procedure, preserving their psychological safety.

Considering the shortage in the literature of rigorous and objective measurements of clinical procedural skills and human performance metrics, this pilot study is one of the first attempts in this direction. It includes psychophysiological assessment (self-assessment questionnaires and biometric monitoring), performance assessment, and UX assessment.

Satisfactory levels of performance and improvement of skills have been achieved. A right compromise between realism and overload has been reached in terms of stress and cognitive load. The results highlight their correct trend during the simulation, which also provide useful insights to improve the learning experience such as the debriefing introduction for correct stress management. However, the MR may bias the students' emotional state and cognitive load. The anxiety and stress perceived by students during the simulations could certainly be different compared to the real clinical situation. The lack of quantification of these differences and understanding how the interaction with unfamiliar technologies affects the training is a limitation of the proposed approach and will be addressed in future works, also enrolling a larger sample of students. Practitioners with different training paths will also be involved to be aware of the real impact of MR simulation-based training on medical education.

The UX highlighted the achievement of the desired level of engagement, immersion, and realism but has been negatively affected by technological limitations, as already shown by the literature (Zhu et al. 2014). Despite them, the practitioner instructor recommended the use of the MR simulation for the degree of immersion that it provides. Technological development in the XR sector is constantly growing, therefore updated and more ergonomic HMDs will be tested to solve these open issues.

Given that redundant training and feedback are critical to the successful acquisition of skills in simulation courses (Bosse et al. 2015), the feedback value, content, frequency, modality, etc., will be investigated in more detail to maximize its effect.

The high number of subjective and objective measures is very interesting to characterize the MR solution comprehensively. However, its use in practice probably will require some adaptations. Once a robust validation of the algorithms for the analysis of vital parameters and the full awareness of the correlation between the various evaluation methods have been obtained, the possibility of reducing the number of questionnaires and devices will be investigated.

Availability of data and material The authors confirm that the data supporting the findings of this study are available within the article. Code availability Not applicable.

Declarations

Conflict of interest No conflicts of interest to disclose.

References

Adrario E, Messi D, Diambrini G (2017a) Attivita' Formativa Professionalizzante AFP, Prima parte: La normativa e le novità previste. Lettere Della Facoltà Bollettino Della Facoltà Di Medicina e Chirurgia dell'UNIVPM 2:5–6

Adrario E, Messi D, Diambrini G (2017b) Attività Formativa Professionalizzante Seconda Parte - Attività formativa professionalizzante integrata. Lettere Della Facoltà Bollettino Della Facoltà Di Medicina e Chirurgia dell'UNIVPM 3:4–5

Atalay KD, Can GF, Erdem SR, Muderrisoglu IH (2016) Assessment of mental workload and academic motivation in medical students. J Pak Med Assoc 66(5):574

Bacca J, Fabregat R, Baldiris S, Graf S, Kinshuk, (2014) Augmented reality trends in education: a systematic review of research and applications. Edu Technol Soc 17:133–149

Bertolotto A, Malentacchi M, Capobianco M, di Sapio A, Malucchi S, Motuzova Y, Pulizzi A, Berchialla P, Sperli F (2016) The use of the 25 Sprotte needle markedly reduces post-dural puncture headache in routine neurological practice. Cephalalgia 36(2):131–138

Bosse HM, Mohr J, Buss B, Krautter M, Weyrich P, Herzog W, Jünger J, Nikende C (2015) The benefit of repetitive skills training and frequency of expert feedback in the early acquisition of procedural skills. BMC Med Edu. https://doi.org/10.1186/s12909-015-0286-5

Brunzini A, Papetti A, Serrani EB, Scafà M, Germani M (2020) How to improve medical simulation training: a new methodology based on ergonomic evaluation. In: Karwowski W, Ahram T, Nazir S (eds) Advances in human factors in training, education, and learning sciences. AHFE 2019. Advances in intelligent systems and computing, vol 963. Springer, Cham. https://doi.org/10.1007/978-3-030-20135-7_14

Brunzini A et al (2021) Cognitive load and stress assessment of medical high-fidelity simulations for emergency management. In: Cassenti D, Scataglini S, Rajulu S, Wright J (eds) Advances in simulation and digital human modeling. AHFE 2020. Advances in intelligent systems and computing, vol 1206. Springer, Cham. https://doi.org/10.1007/978-3-030-51064-0_44

Campisi CA, Li EH, Jimenez DE, Milanaik RL (2020) Augmented reality in medical education and training: from physicians to patients. In: Geroimenko V (ed) Augmented reality in education. Springer, Cham. https://doi.org/10.1007/978-3-030-42156-4_7

Chaballout B, Molloy M, Vaughn J, Brisson Iii R, Shaw R (2016) Feasibility of augmented reality in clinical simulations: using google glass with manikins. JMIR Med Edu 2(1):e2. https://doi.org/10.2196/mededu.5159

Cheng KH, Tsai CC (2013) Affordances of augmented reality in science learning: suggestions for future research. J Sci Edu Technol 22:449–462. https://doi.org/10.1007/s10956-012-9405-9

Coles TR, John NW, Gould DA (2011) Integrating haptics with augmented reality in a femoral palpation and needle insertion training simulation. In IEEE Trans Haptics 4:199–209

Cook DA, Hamstra SJ, Brydges R, Zendejas B, Szostek JH, Wang AT, Erwin PJ, Hatala R (2013) Comparative effectiveness of instructional design features in simulation-based education: systematic review and meta-analysis. Med Teach 35(1):e867–e898. https://doi.org/10.3109/0142159X.2012.714886

Curtis MT, DiazGranados D, Feldman M (2012) Judicious use of simulation technology in continuing medical education. J Contin Edu Heal Prof 32:255–260

Dias RD, Ngo-Howard MC, Boskovski MT, Zenati MA, Yule SJ (2018) Systematic review of measurement tools to assess surgeons' intraoperative cognitive workload. Br J Surg 105:491–501

François C, Hoyoux T, Langohr T, Wertz J, Verly JG (2016) Tests of a new drowsiness characterization and monitoring system based on ocular parameters. Int J Environ Res Public Health 13(2):174. https://doi.org/10.3390/ijerph13020174

Fraser K, Ma I, Teteris E, Baxter H, Wright B, McLaughlin K (2012) Emotion, cognitive load and learning outcomes during simulation training. Med Educ 46:1055–1062

Garzón J, Pavón J, Baldiris S (2017) Augmented reality applications for education: five directions for future research. Lect Notes Comput Sci 10324:402–414. https://doi.org/10.1007/978-3-319-60922-5_31

Garzón J, Pavón J, Baldiris S (2019) Systematic review and meta-analysis of augmented reality in educational settings. Virtual Real 23:447–459. https://doi.org/10.1007/s10055-019-00379-9

Gattullo M, Dammacco L, Ruospo F, Evangelista A, Fiorentino M, Schmitt J, Uva AE (2020) Design preferences on industrial augmented reality: a survey with potential technical writers. Adjunct Proceedings of the 2020 IEEE International Symposium on Mixed and Augmented Reality, ISMAR-Adjunct 2020 9288426, 172–177

George A, Blaauw D, Green-Thompson L et al (2019) Comparison of video demonstrations and bedside tutorials for teaching paediatric clinical skills to large groups of medical students in resource-constrained settings. Int J Edu Technol High Edu 16:34. https://doi.org/10.1186/s41239-019-0164-z

Gerup J, Soerensen CB, Dieckmann P (2020) Augmented reality and mixed reality for healthcare education beyond surgery: an integrative review. Int J Med Edu 11:1–18. https://doi.org/10.5116/ijme.5e01.eb1a

Goldberg MB, Mazzei M, Maher Z, Fish JH, Milner R, Yu D, Goldberg AJ (2018) Optimizing performance through stress training - An educational strategy for surgical residents. Am J Surg 216:618–623

Gutierrez-Puerto E, Vega-Medina L, Tibamoso G, Uribe-Quevedo A, Perez-Gutierrez B (2015) Augmented reality central venous access training simulator in: HCI international 2015 - posters' extended abstracts. Commun Comput Inform Sci 528:174–179. https://doi.org/10.1007/978-3-319-21380-4_31

Hart S, Staveland L (1988) Development of NASA-TLX (Task Load Index): results of empirical and theoretical research. Adv Psychol 52:139–183

Herron J (2016) Augmented reality in medical education and training. J Electron Resour Med Libr 13:51–55

Hong M, Rozenblit JW, Hamilton AJ (2021) Simulation-based surgical training systems in laparoscopic surgery: a current review. Virtual Real 25:491–510. https://doi.org/10.1007/s10055-020-00469-z

Kobayashi L, Zhang XC, Collins SA, Karim N, Merck DL (2018) Exploratory application of augmented reality/mixed reality

devices for acute care procedure training. West J Emerg Med 19(1):158–164. https://doi.org/10.5811/westjem.2017.10.35026

Kotranza A, Scott Lind D, Lok B (2012) Real-time evaluation and visualization of learner performance in a mixed-reality environment for clinical breast examination. IEEE Trans Visual Comput Gr 18:1101–1114

Lesage F, Berjot S, Deschamps F (2012) Clinical stress assessment using a visual analogue scale. Occup Med 62:600–605

Liang CJ, Start C, Boley H, Kamat VR, Menassa CC, Aebersold M (2020) Enhancing stroke assessment simulation experience in clinical training using augmented reality. Virtual Real. https://doi.org/10.1007/s10055-020-00475-1

Linde AS, Geoffrey TM (2019) Applications of future technologies to detect skill decay and improve procedural performance. Mil Med 184:72–77

Magee D, Zhu Y, Ratnalingam R (2007) An augmented reality simulator for ultrasound guided needle placement training. Med Biol Eng Compu 45:957–967

Mendes HCM, Costa CIAB, da Silva NA, Leite FP, Esteves A, Lopes DS (2020) PIÑATA: pinpoint insertion of intravenous needles via augmented reality training assistance. Comput Med Imag Graph 82:101731. https://doi.org/10.1016/j.compmedimag.2020.101731

Munzer BW, Khan MM, Shipman B, Mahajan P (2019) Augmented reality in emergency medicine: a scoping review. J Med Internet Res 21(4):e12368. https://doi.org/10.2196/12368

Naismith LM, Cavalcanti RB (2015) Validity of cognitive load measures in simulation-based training: a systematic review. Acad Med 90:S24–S35

Onda EL (2012) Situated cognition: its relationship to simulation in nursing education. Clin Simul Nurs 8(7):e273–e280. https://doi.org/10.1016/j.ecns.2010.11.004

Paul G, Scataglini S (2019) Open-source software to create a kinematic model in digital human modeling. DHM and Posturography, Academic Press, pp 201–213. https://doi.org/10.1016/B978-0-12-816713-7.00017-9

Pheasant S (1999) Body space: anthropometry, ergonomics and the design of work. Taylor & Francis, UK

Rochlen LR, Levine R, Tait AR (2017) First-person point-of-view-augmented reality for central line insertion training: a usability and feasibility study. Simul Healthc 12:57–62

Rodziewicz TL, Houseman B, Hipskind JE (2021) Medical error reduction and prevention. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing

Roh YS, Jang KI, Issenberg SB (2021) Nursing students' perceptions of simulation design features and learning outcomes: the mediating effect of psychological safety. Collegian 28(2):184–189. https://doi.org/10.1016/j.colegn.2020.06.007

Salar R, Arici F, Caliklar S, Yilmaz RM (2020) A model for augmented reality immersion experiences of university students studying in science education. J Sci Educ Technol 29:257–271. https://doi.org/10.1007/s10956-019-09810-x

Sarfati L, Ranchon F, Vantard N, Schwiertz V, Larbre V, Parat S, Faudel A, Rioufol C (2019) Human-simulation-based learning to prevent medication error: a systematic review. J Eval Clin Pract 25(1):11–20. https://doi.org/10.1111/jep.12883

Scafà M, Serrani EB, Papetti A, Brunzini A, Germani M (2020) Assessment of students' cognitive conditions in medical simulation training: a review study. In: Cassenti D (ed) Advances in human factors and simulation AHFE 2019 advances in intelligent systems and computing. Springer, Cham

Sherstyuk A, Vincent D, Berg B, Treskunov A (2011) Mixed reality manikins for medical education. In: Furht B (ed) Handbook of augmented reality. Springer, New York. https://doi.org/10.1007/978-1-4614-0064-6_23

Spielberger CD, Gorsuch RL (1983) State-trait anxiety inventory for adults : sampler set: manual, test, scoring key. Mind Garden, Redwood City, Calif.

Stawarczyk D, François C, Wertz J, D'Argembeau A (2020) Drowsiness or mind-wandering? Fluctuations in ocular parameters during attentional lapses. Biol Psychol 156:107950. https://doi.org/10.1016/j.biopsycho.2020.107950

Sugarindra M, Suryoputro MR, Permana AI (2017) Mental workload measurement in operator control room using NASA-TLX. IOP Conf. Series: Materials Science and Engineering 277

Tai Y, Wei L, Zhou H (2019) Augmented-reality-driven medical simulation platform for percutaneous nephrolithotomy with cybersecurity awareness. Int J Distrib Sens Netw. https://doi.org/10.1177/1550147719840173

Wang S, Parsons M, Stone-McLean J, Rogers P, Boyd S, Hoover K, Meruvia-Pastor O, Gong M, Smith A (2017) Augmented reality as a telemedicine platform for remote procedural training. Sensors 17(10):2294. https://doi.org/10.3390/s17102294

WHO World Health Organization Data and Statistics (2017) Regional Office for Europe. http://www.euro.who.int/en/health-topics/Health-systems/patient-safety/data-and-statistics.

Wu HK, Lee WY, Chang HY, Liang JC (2013) Current status, opportunities and challenges of augmented reality in education. Comput Educ 62:41–49

Zhu E, Hadadgar A, Masiello I, Zary N (2014) Augmented reality in healthcare education: an integrative review. PeerJ 2:e469. https://doi.org/10.7717/peerj.469