POLITECNICO DI TORINO
Repository ISTITUZIONALE

Enhancing Difficult Airway Management Training: The Role of Virtual Reality and Adaptive Learning

Original
Enhancing Difficult Airway Management Training: The Role of Virtual Reality and Adaptive Learning / Battegazzorre, E., Gino, B., Strada, F., Kapralos, B., Dubrowski, A., Lamberti, F., Bottino, A.. - In: VIRTUAL REALITY. - ISSN 1434-9957. - ELETTRONICO. - 29:1(2025). [10.1007/s10055-025-01105-4]

Availability:
This version is available at: 11583/2996580 since: 2025-01-23T15:45:06Z

Publisher:
Springer

Published
DOI:10.1007/s10055-025-01105-4

Terms of use:

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

Publisher copyright

(Article begins on next page)

25 November 2026

Virtual Reality (2025) 29:22
https://doi.org/10.1007/s10055-025-01105-4

ORIGINAL ARTICLE

Enhancing difficult airway management training: the role of virtual reality and adaptive learning

Edoardo Battegazzorre[1] · Bruno Gino[2] · Francesco Strada[1] · Bill Kapralos[3] · Adam Dubrowski[2] · Fabrizio Lamberti[1] · Andrea Bottino[1]

Received: 7 June 2024 / Accepted: 9 January 2025
© The Author(s) 2025

Abstract
Emergency physicians play a central role in healthcare. They often must make quick and accurate decisions to save patients' lives. Among the critical procedures they have to master is difficult airway management (DAM), a procedure required to establish and maintain a patient's airway for adequate ventilation and oxygenation. To ensure optimal proficiency in DAM, the clinical skills that comprise this procedure must be regularly practiced and updated. However, traditional training approaches present significant organizational challenges in terms of time and cost. In response to these issues, we have developed an innovative education and training application employing immersive Virtual Reality (VR) for teaching basic to advanced DAM procedures, supported by an Adaptive Learning system. To evaluate the effectiveness of our DAM training system, we conducted experiments with a control group trained using traditional methods and two VR subgroups, one with and one without the Adaptive Learning component. Our results show that simulating the DAM procedure in VR is effective in improving students' knowledge and produces comparable learning outcomes to traditional teaching methods. Interestingly, our study did not provide conclusive evidence that the adaptive design was superior to the non-adaptive one in terms of knowledge and acquisition of skills. However, it demonstrated greater efficiency, particularly in reducing training time.

Keywords Virtual reality · Medical training · Difficult airway management · Adaptive learning

1 Introduction

Emergency physicians are often faced with situations in which they must manage patients' airways, a fundamental aspect of critical care. A difficult airway occurs when even a trained healthcare professional has difficulty with mask ventilation, orotracheal intubation, or both (Apfelbaum et al. 2013). In extreme cases, surgical intervention such as a cricothyrotomy may be required (Rourke 2006). Dealing effectively with these scenarios requires rapid and informed decision-making. Physicians need to assess the patient's condition, medical history and available medications, which

✉ Edoardo Battegazzorre
edoardo.battegazzorre@polito.it

Bruno Gino
bruno.gino@ontariotechu.ca

Francesco Strada
francesco.strada@polito.it

Bill Kapralos
bill.kapralos@ontariotechu.ca

Adam Dubrowski
adam.dubrowski@ontariotechu.ca

Fabrizio Lamberti
fabrizio.lamberti@polito.it

Andrea Bottino
andrea.bottino@polito.it

[1] Department of Control and Computer Engineering (DAUIN), Politecnico di Torino, Corso Duca degli Abruzzi 24, 10129 Turin, Piemonte, Italy

[2] Faculty of Health Sciences, Ontario Tech University, 2000 Simcoe St N, Oshawa, Ontario L1G 0C5, Canada

[3] Faculty of Business and Information Technology, Ontario Tech University, 2000 Simcoe St N, Oshawa, Ontario L1G 0C5, Canada

Published online: 21 January 2025

emphasizes the need for comprehensive training (Yang et al. 2016).

Traditional Difficult Airway Management (DAM) training primarily relies on face-to-face, instructor-led methods. These include didactic lectures, clinical exercises, mentoring, and simulations with medical manikins. Instructors often take on multiple roles, such as describing clinical scenarios, presenting patient conditions, and providing feedback. Although this approach is comprehensive, it is also highly resource-intensive. It requires a significant time and financial commitment from both trainees and instructors. In addition, the requirement for physical presence can lead to logistical challenges (Yang et al. 2016).

In response to these limitations, Virtual Reality (VR)-based training offers a promising alternative to traditional approaches. VR applications help to reduce logistical and financial burdens (Haerling 2018; Mergen et al. 2023) while delivering high-quality learning, practice, and skills assessment sessions (Liu et al. 2023). These sessions take place in a safe and controlled environment, making them particularly valuable for high-risk procedures such as those often encountered in the medical field (Carruth 2017). VR also supports self-directed and self-paced training and offers flexibility in terms of location, time and repeatibility (Yang et al. 2016; Xie et al. 2021). Furthermore, VR aligns with the principles of Adaptive Learning (AL) (Zahabi and Abdul Razak 2020) and Adaptive Training (AT) (Kelley 1969). AL focuses on the personalization of educational resources and activities based on learners' individual needs, preferences, and progress. In contrast, AT adapts training programs by adjusting content and difficulty levels to the learner's performance on practical tasks. Through the use of computer algorithms and artificial intelligence, VR applications can seamlessly integrate both AL and AT. These personalized approaches maximize the effectiveness and efficiency of the educational process, increasing engagement and improving learning outcomes (Kaplan 2021).

Another point to consider is that training in emergency medicine involves two key elements: psychomotor skills and cognitive skills. Psychomotor skills include the ability to perform physical tasks that require coordination and precise movements, while cognitive skills relate to knowledge of the steps, requirements, and constraints of medical procedures (Rourke 2006). Traditional DAM training methods promote the development of psychomotor skills through hands-on practice with lifelike manikins and simulators. Cognitive skills training, however, often relies on textbooks, lectures and video-based instruction. While these approaches provide foundational knowledge, they lack the interactive, scenario-based learning required to foster deep cognitive processing and decision-making skills. These skills are essential for addressing the unique challenges of DAM and require a thorough understanding of procedural steps, potential complications, and strategies tailored to patient-specific factors.

Another strength of VR applications lies in their ability to adopt a "learn by doing" approach (Radianti et al. 2020; Makransky and Petersen 2021). This is particularly effective for teaching procedures that are difficult to teach in traditional lectures. However, a thorough review of the relevant literature (Sect. 2) revealed several unexplored aspects of procedural learning for DAM in VR, particularly in patient assessment and decision-making. Our work addresses this gap by developing and evaluating a VR-based application that leverages the immersive and interactive capabilities of VR. This approach allows users to gain a deeper understanding of the DAM procedure and explore not only the "how" but also the "why" behind each step. In this way, important skills such as rapid decision-making and critical thinking are fostered, which are crucial in DAM training (Dabija et al. 2019; Rosen et al. 2006) as understanding the nuances of the procedure and anticipating potential complications are just as important as executing the steps themselves.

Furthermore, our work leverages AL and introduces an innovative approach by applying the principles of the Expertise Reversal Effect (ERE) (Sweller et al. 2003). ERE suggests that instructional strategies that are effective for novice learners may be less appropriate for experienced learners and vice versa (Kalyuga 2009). In line with this principle, our application adapts the level of detail of the instructions to the learner's familiarity with the topic. Novices receive comprehensive instructions to address gaps in foundational knowledge and avoid confusion. Conversely, experts receive concise and focused information to avoid redundancy and maintain engagement. This approach optimizes learning efficiency at different levels of expertise.

The contributions of this work can be summarized as follows:

- We have developed the first immersive VR application dedicated to the comprehensive procedural knowledge required for DAM, covering all clinical interventions for adult patients, including decision-making following patient assessment.
- We have introduced an AL framework based on the ERE, designed to dynamically adapt the instructor's feedback during the learning process.To evaluate the effectiveness of our approach, we conducted an experimental study with 46 participants (physicians). The aim of this study was to validate two primary research hypotheses (RH):

- Our VR application effectively improves the participants' knowledge of the DAM procedure compared to a control group trained using traditional methods (**RH1**).

- Our adaptive framework is more effective for teaching compared to a non-adaptive version of the same application (**RH2**).Our results suggest that the proposed VR application is more effective in terms of learning outcomes than traditional methods. Although our study does not provide conclusive evidence for the added advantage of the adaptive framework over its non-adaptive counterpart, it highlights opportunities for further investigation and suggests that AL principles under different conditions or with additional development may indeed contribute to further improve educational outcomes.

The remainder of the article is structured as follows. Sect. 2 presents the current state of the art of simulator-based training, implementations of DAM in immersive VR, AL in medical applications and ERE-based frameworks for VR. Sect 3 presents details of the developed simulation system. Sect 4 introduces the experimental protocols, and Sect. 5 presents and discusses the experimental results as well as the limitations and possible future developments of this study. Finally, Sect. 6 concludes the paper.

2 State of the art

2.1 Simulator-based training for difficult airways

Numerous studies have investigated both simulator-based (SBT) and non-simulator-based (NSBT) training methods for DAM, where a simulator in this context is a digital or physical system designed to replicate real-world scenarios in a controlled environment. A recent review (Sun et al. 2017) discusses 19 works comparing SBT and NSBT, seven of which are identified as VR-based. However, it is important to point out that many of these VR-based studies (i.e., (Morgan et al. 2002; Multak et al. 2002; Hall et al. 2005; Kory et al. 2007; Wenk et al. 2009; Hallikainen et al. 2009; Modell 2002)) predominantly involve computerized Human Patient Simulators (HPS) rather than immersive VR simulations with computer-generated environments and virtual humans. HPS are advanced simulation manikins that mimic human physiological responses to medical interventions by integrating specialized software. In contrast, VR simulations can create a fully immersive, interactive learning experience that transcends the physical limitations of HPS. The work presented in Yang et al. (2016) further explores the strengths and weaknesses of SBT and NSBT. It highlights the high cost and logistical complexity of delivering face-to-face training, the inherent limitations of physical simulators in replicating the full range of clinical scenarios, and the need for "refresher courses" to maintain and update skills and knowledge. Finally, in Demirel et al. (2016), the authors conduct a study aimed at refining the simulation design for the cricothyrotomy intervention by applying hierarchical task analysis to describe the crucial tasks, parameters and common errors associated with this technique. However, it is important to emphasize that while this study provides valuable guidelines for simulation design, it didn't present a concrete implementation.

2.2 Airway simulators in immersive VR

Given the growing interest in immersive VR in SBT, particularly for its ability to enable learners to experience realistic scenarios with greater flexibility and to simulate a wider range of emergency situations, several studies have investigated its application in the teaching of DAM procedures. These studies have primarily focused on isolated clinical interventions such as endotracheal intubation and cricothyrotomy (Demirel et al. 2016; Rajeswaran et al. 2019a, b; Samosorn et al. 2020). However, this limited focus neglects a crucial aspect for holistic DAM training, which is the development of decision-making skills that are essential for assessing a patient's condition and determining the appropriate intervention.

For example, Demirel et al. (2016) presents an immersive DAM simulation that includes assessments and interventions such as Mallampati evaluation, endotracheal intubation, and cricothyrotomy using a Head-Mounted Display (HMD) and haptic devices. However, this application does not include a wide range of patient assessments such as the Glasgow Coma Scale (Teasdale and Jennett 1974) nor scenarios requiring non-invasive techniques such as external ventilation. This selective approach to patient condition assessment suggests that the critical decision-making processes that are part of comprehensive DAM training are only partially covered. In Rajeswaran et al. (2018, 2019a, 2019b), the authors present Airway VR, a VR learning tool focused on endotracheal intubation, and conduct a limited exploratory user test with nine participants using a self-assessment knowledge questionnaire along with a non-standard usability questionnaire. Their approach, while promising, neglects some critical aspects of DAM (e.g., external ventilation, cricothyrotomy and patient assessment) and divides the experience into discrete lessons on specific topics, limiting the continuity and immersion that are critical for simulating real-world emergency scenarios. Another work (Samosorn et al. 2020) conducts a pilot study of a DAM training tool for nursing students that presents various topics as narrated VR lessons with some interactive elements, although segmenting again the curriculum without creating a consistent and continuous training flow. However, the paper lacks detail about the nature of these interactions and their impact on the learning experience.

Our work differs from the ones discussed so far by presenting the DAM procedure as a unified and seamless experience that begins with a comprehensive patient assessment and leads to contextualized interventions. This design ensures that the decision on which intervention to perform is not predetermined, but emerges from a thorough assessment of the clinical scenario, which may include randomized elements to reflect the unpredictability and complexity of real-world medical emergencies where decision-making is a central aspect of saving a patient's life.

Finally, we point out a major limitation of current immersive VR systems: the lack of detailed haptic feedback, which limits their ability to fully replicate the rich tactile experience essential for mastering fine motor skills provided by physical manikins. Indeed, VR-based training systems use a variety of interaction methods and devices, ranging from handheld controllers that provide simple vibration feedback to confirm actions (Rajeswaran et al. 2018, 2019b), to more advanced haptic devices such as the Geomagic Touch, which simulate tissue resistance and tool dynamics during airway procedures (Demirel et al. 2016). However, even the Geomagic Touch has significant limitations, such as the small active interaction area and a maximum force output that is not sufficient to realistically replicate the mechanical resistance that occurs during interventions.

2.3 Adaptive learning in VR

While immersive VR provides a robust platform for simulating complex medical scenarios addressing the limitations of traditional training methods, its potential can be further enhanced by integrating AL and AT techniques. These approaches have evolved considerably since their inception in the 1960 s (Kelley 1969). However, despite their increasing adoption in various educational contexts, their application in immersive VR is relatively new. The review of self-adaptive technologies in VR presented in Vaughan et al. (2016) highlights the lack of a unified framework for the implementation of AL in VR and points to the critical need for an adaptability that meets the challenges, strengths, and learning preferences of individual learners. The recent systematic literature review in Zahabi and Abdul Razak (2020) further elaborates on the categorization of VR-based AT applications and identifies key components such as performance measurement, adaptive variables and logic. The authors emphasize the importance of adaptive feedback in terms of timing, content and delivery method (Zahabi and Abdul Razak 2020).

A limited number of studies combine AL or AT with immersive VR (HMDs or cave automatic virtual environments (CAVEs)). Some approaches use biosignals (electroencephalography, electrodermal activity and eye tracking) to modulate difficulty based on stress level in contexts ranging from driving simulations (Ben Abdessalem and Frasson 2017; Dey et al. 2019) to interpersonal communication management (Blankendaal and Bosse 2018). Other studies adapt scenarios based on user habits, such as driving behavior in a virtual city (Lang et al. 2018), or provide specific feedback aimed at overcoming certain weaknesses of the learner (Jeelani et al. 2017). For example, Lin and Wang (2019) shows how adapting to individual learning styles can improve motivation and educational outcomes in engineering disciplines. However, such applications often tend towards AT which focuses on practical skills, rather than AL which emphasizes a broader understanding of procedural knowledge.

Similar limitations can be highlighted for the use of VR in learning complex medical procedures. While efforts have been made to use adaptive technologies in medical education (e.g., for laparoscopic surgery Pham et al. 2005; Mariani et al. 2018 and central venous catheterization Yovanoff et al. 2018), the proposed adaptation often revolves around training aimed at improving the practical and technical skills required to perform specific tasks. They therefore often neglect procedural skills, which include competencies such as decision-making and problem-solving that are critical to performing complex medical procedures (Chan et al. 2024). Moreover, these applications typically validate their effectiveness using simplified and controlled environments ("toy" scenarios) designed to demonstrate the feasibility of the adaptive features rather than simulations represnting the compexity of real-world scenarios.

2.4 Expertise reversal effect and (Desktop) VR

While ERE expands the potential for personalized learning experiences and provides a compelling framework for tailored feedback based on the learner's evolving expertise, its application in VR is still relatively limited. The authors of Billings (2012) and Serge et al. (2013) have conducted experiments to analyze the effects of ERE-based adaptive systems in specific contexts such as search and rescue scenarios. Their results show that the group that received detailed feedback performed best, followed by the adaptive group that switched from detailed to general feedback. However, the focus of this research was primarily on desktop-based VR applications. Although no experimental support data were presented, the authors of Dalgarno and Lee (2010) argue that immersive VR technologies should make a significant contribution to the effectiveness of the learning application and enhance the positive effects of ERE-based adaptive learning strategies.

Taking all the previous observations into account, our research aims to integrate immersive VR, AL, AT and ERE

into a unified training platform for DAM, creating a holistic and adaptive learning environment that targets both procedural knowledge and critical decision-making skills. This approach seeks to overcome the limitations of existing tools by providing a seamless, immersive experience tailored to the dynamic needs of learners.

3 Methods

The main objective of our work is to design and develop an interactive learning environment tailored to improve the procedural understanding and decision-making skills of emergency physicians in DAM. Our application integrates immersive VR technologies with adaptive learning principles to provide an educational experience that adapts to the individual needs and performance level of the users.

The remainder of this Section illustrates the clinical scenarios presented to the users, Sect. 3.1, followed by a description of the virtual environment (VE) in which the simulation takes place, Sect. 3.2. Finally, in Sect. 3.3, the application architecture is detailed.

3.1 Implemented procedure

The learning experience begins in the emergency room of a hospital. The users are first familiarized with the clinical scenario of a patient exhibiting certain symptoms. The initial tasks include preliminary operations such as undressing the patient, connecting monitors, positioning the patient correctly and gaining intravenous (IV) access, which are essential for the subsequent medical interventions. In our application, these four operations are carried out by a nurse upon request by the user.

Evaluating the patient's condition is a crucial part of the training. This includes examinations to detect obstructions or fluids in the patient's mouth and determining the Mallampati class (Mallampati et al. 1985), which is a predictive measure of intubation difficulty. Additionally, the Glasgow Coma Scale (Teasdale and Jennett 1974), a widely used clinical tool for assessing a patient's level of consciousness based on verbal, motor, and eye-opening response, is employed to assess the severity of the patient's condition. Based on this assessment, a choice is made between three progressively more complex interventions: external ventilation, orotracheal intubation and cricothyrotomy (Fig. 1). It is important to clarify that each subsequent intervention encompasses the previous ones and adapts to the increasing severity of the patient's condition. In clinical cases that are not critical, external ventilation is sufficient. In more severe cases, such as comatose patients or when maintaining airway access for a prolonged period is necessary, orotracheal intubation is required. In critical situations where conventional methods prove ineffective, a cricothyrotomy is required. The application guides the user through the sequence of these interventions and presents them with common challenges such as unilateral bronchial and esophageal intubations, providing guidance on how to handle such complications. The steps of our procedure (which are not described in detail for the sake of brevity) follow the structure described in Rourke (2006) and can range in number from 20 to 50, depending on the severity of the patient's condition.

To avoid confusion, in the following, we will use the term *DAM procedure* to refer to the comprehensive sequence of steps and decision-making processes in the management of difficult airways, while the term *clinical intervention* will refer to the macro steps of the DAM (i.e., external ventilation, orotracheal intubation, and cricothyrotomy).

3.2 Virtual environment

The VE (Fig. 2) recreates the most important elements of an emergency room and includes a table with essential medical instruments, a hospital bed, and an equipped monitoring area. Above the bed is a large screen that displays supplementary information for the user, such as text instructions, additional media, symptoms, and feedback on their actions. The environment is populated by three virtual agents: the patient, a nurse, and the virtual instructor (VI), all of whom play a crucial role in the simulation.

The dynamic nature of the VR experience allows the position and appearance of the patient to adapt in real time to the interactions initiated by the user or the nurse. The nurse engages in actions solely at the request of the user.

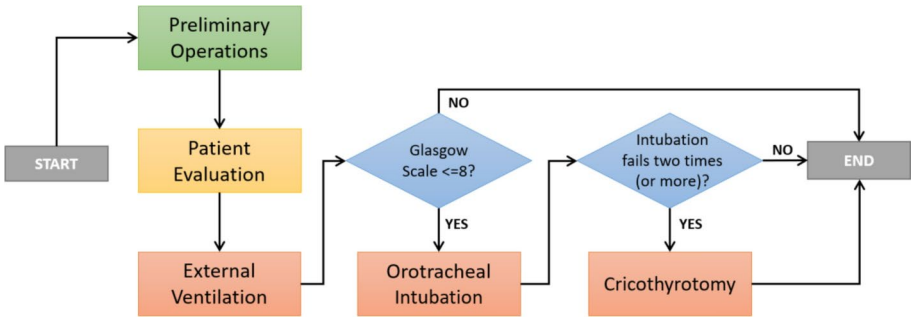


Fig. 1 A high-level schema of the Difficult Airway procedure that can be experienced in the VR application. In this workflow, the orange blocks are the clinical interventions, and the blue blocks are the decision points



Fig. 2 Two views of the virtual environment. Left: the VI and tools table. Right: the patient, nurse, and screen for displaying videos, symptoms, and complementary media

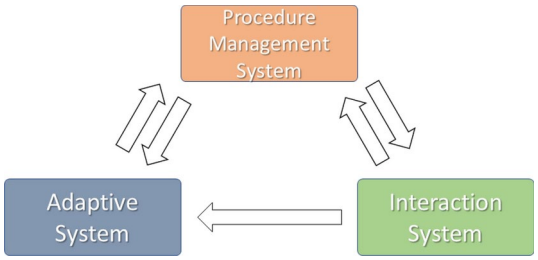


Fig. 3 General schema of the VR Application

Meanwhile, the VI plays a central role in guiding the user through the process by providing instructions and cues without directly interacting with the patient or the medical equipment. All text lines pronounced by the VI are also simultaneously displayed on a large panel next to it (Fig. 2, left), to ensure that the instructions are clear and understandable. In addition, any object on the table, including syringes, drug vials, and other medical tools, can be interacted with. This feature allows the users to practice handling devices that are important for emergency medical interventions.

3.3 Application architecture

The architecture of our application can be divided into three high-level core systems that work in close synergy to ensure a seamless and effective training experience. These systems (Fig. 3) include the *Procedure Management System*, which orchestrates the sequence of procedural steps and decision points, the *Interaction System*, which manages the interaction within the VE, and the *Adaptive System*, which is responsible for personalizing the learning experience by adapting the content and difficulty level based on the users' performance and progress.

3.3.1 Interaction system

In the VE, the users can seamlessly move, interact with objects, and converse with virtual agents. The navigation leverages the "real-walking" metaphor, where movements in the real world are mapped one-to-one in the VE. The application supports multimodal interactions, including body movements and voice communication.

Handheld controllers are used for object selection and manipulation, allowing the users to easily pick up and activate objects and interact with the application's user interfaces (UIs). In the VR environment, users interact with objects via two interaction metaphors: hand-grasping and raycasting. The two metaphors differ primarily in how the object is selected. In the hand-grasping metaphor, the selection is performed by positioning the virtual hand near the object (Fig. 4, left), making it ideal for objects within reach. In contrast, the raycasting metaphor uses a visual ray extending from the virtual hand to select objects at a distance (Fig. 4, center). Once an object is selected using either method, the grabbing action is completed by pressing the controller's grip button. Regardless of the selection method, the final result is the same: the object snaps to the virtual hand, which adapts its shape to match the grasped object, creating a seamless and realistic interaction

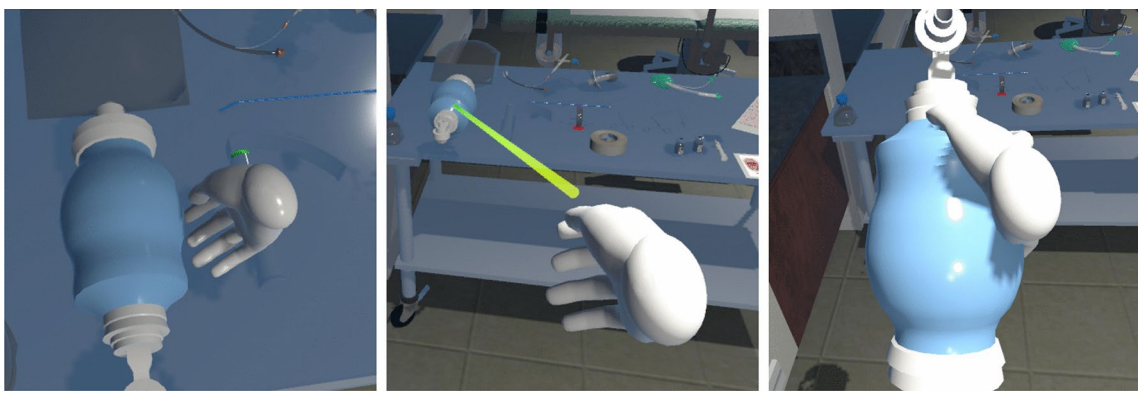


Fig. 4 Details of the grasping interaction. Left: An object within direct reach is selected by touching it with the virtual hand. Center: A distant object is selected using the raycasting metaphor. Right: In both cases, the selected object is held by the virtual hand after being grabbed

(Fig. 4, right). By combining these interaction mechanisms, the system provides a learning experience that allows learners to practice procedural tasks in a way that closely resembles real-life interactions.

For complex tasks that require precise hand and finger movements, such as manipulating a syringe or using a laryngoscope, an *activation* UI is displayed. This interface, identified by a distinct icon (Fig. 5, left), highlights the interaction zone around the object and prompts the user to position the controller correctly and press a button to complete the interaction. Visual, auditory and haptic feedback is provided to confirm the action. As reported by Johnson et al. (2022), this type of interaction is generally considered intuitive and less cognitively demanding than other alternatives.

The application does not rely on an external full-body tracking system to keep the hardware setup simple and portable. Thus, only the user's hands, represented as white gloves (Figs. 4 and 5, right), are visible in the VR application. Although no specific user data was collected to confirm this assertion, our observations during usability testing suggest that this design choice does not significantly impact the immersive experience, as the hand animations dynamically adapt to mimic actions such as grasping various objects to enhance the sense of presence within the VE.

In addition, the interaction system, which utilizes a commercial Natural Language Processing (NLP) system, includes text-to-speech, speech-to-text, and intent recognition capabilities, allowing the users to interact with virtual agents via natural language. These verbal interactions are an integral part of the DAM procedure and allow the users to provide commands to the nurse or inquire about the virtual patient's condition, as well as ask the VI for advice. Voice interaction can significantly improve immersion as interaction metaphors are no longer needed when communicating with the agents (Monteiro et al. 2021). However, it is important to highlight that our implementation only recognizes a predefined set of "intents" that are relevant to the procedure and does not support open-ended conversation. For example, the user may instruct the nurse to "Hold the patient's head" or "Position the patient correctly", with such requests being interpreted as the same intent.

Since interacting with a VR application can be a daunting experience for first-time users, which is the most likely condition of our potential users, we have included an introductory tutorial that familiarizes them with the interaction paradigms, including hand gestures, verbal commands, and visual cues. The users are encouraged to practice each interaction at least twice to familiarize themselves with how the VR environment works. Our goal is to ensure that the users can focus solely on the educational content without being distracted by the novelty associated with the use of VR. This approach is crucial to ensure that the users can work efficiently and with ease on their learning journey.

3.3.2 Procedure management system

The execution of the DAM procedure requires certain actions to be performed in a specific order. This procedure is inherently complex and contains multiple decision points that require the users to evaluate the current scenario and make informed decisions to navigate through different paths or branches of the procedure.

To cope with this complexity, our application adopts the structured approach proposed in Strada et al. (2019), which models the dependencies between actions using a directed graph. In this framework, the individual procedure steps are represented as nodes, and the graph's edges denote the dependency requirements. Each procedure step is implemented as an independent software module whose execution flow is determined by the user's interactions in the VE and by internal and external events. The graph design includes composite nodes to group sub-nodes under different execution algorithms such as sequential, parallel or loop control to increase the flexibility in handling procedural dependencies. The procedure management system is responsible for scheduling procedure steps, managing step completions, checking whether the user's actions comply with the rules and constraints of the DAM procedure and providing feedback on the user's performance.

The learning journey supported by the application distinguishes between *Learning* and *Trainingmodes* to accommodate different learning phases.

The *Learning mode* adopts an explanatory strategy. In this approach, the users are guided by the VI step-by-step through the activities, rules and guidelines of the procedure, and the information is conveyed in a direct and structured



Fig. 5 Left: The icon highlighting that an activation is requested to proceed. In this case, the procedure step corresponds to the application of the chest pain stimulus to the patient's sternum, to assess their motor response. Right: The user's virtual hand, holding the laryngoscope

manner. Also, a video of an actual doctor performing the operation on a manikin is shown in any node where it is relevant.

In *Training mode*, the VI refrains from active guidance, no visual cues are provided to indicate the necessary interactions, and the application only provides auditory feedback for correctly executed steps or errors. In the event of a major error, i.e., a critical error that endangers the patient's health or life, such as failure to ventilate the patient or to initiate intubation without prior sedation, the VI immediately alerts the user of the grave error, then the scenario is interrupted and the user returns to the main menu. This dire consequence was introduced to encourage the users to learn from their mistakes and improve their decision-making skills. Finally, every minor and major error during training is recorded by the application and taken into account when assessing the user's proficiency at the different procedural stages, as explained in more detail in Sect. 3.3.3.

This architecture fits seamlessly into the overarching goal of adaptability and accommodates the varying educational needs of emergency physicians. By navigating between *Learning* and *Trainingmodes*, the users can deepen their understanding, review procedural knowledge, and assess their skills in different scenarios, tailoring their experience to their specific knowledge and skills.

3.3.3 Adaptive system

The adaptive learning framework used in this study builds on the foundations of previous research (Bloom and Loftin 2020; Nour et al. 1995; Yaghmaie and Bahreininejad 2011). Its main goal is to improve the learning outcomes of the VR application by dynamically adapting the learning materials' content to the users' performance and preferences. To achieve this goal, the framework combines user modeling techniques and adaptive algorithms to analyze the users' behavior and performance. The overall design of our adaptive framework (Fig. 6) consists of six interconnected modules:

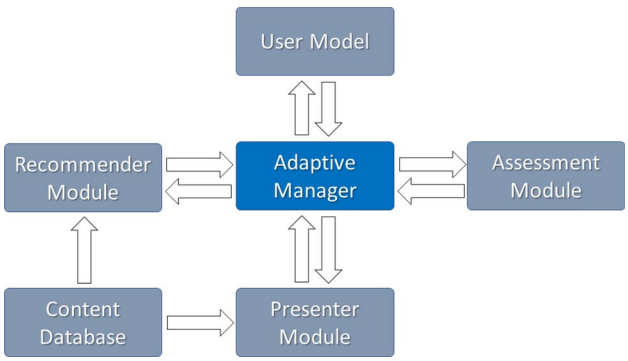


Fig. 6 The general scheme of the Adaptive System

- The Content Database, also known as the Expert Module or Expertise Module (Brusilovskiy 1994), contains a meta-description of all learning content. This includes descriptions of the individual procedure steps, available in different levels of detail, and supplementary media, such as images and videos, to enrich the learning experience.
- The User Model collects data about the user's behavior and performance in the VE. In particular, it stores details about the user's history in *Learning* and *Training mode*, the number of repetitions for each phase, minor and major errors made by the user when performing each step of the procedure, and their interactions with virtual agents.
- The Presenter Module is responsible for presenting the learning material to the user in an engaging and effective manner. It uses an algorithm to determine the user's current proficiency level, which is categorized into three levels (Beginner, Intermediate, and Expert) based on the information from the User Model. It then selects the optimal level of detail of instructions for the VI and decides on the need for additional media to support the user's understanding. A detailed description of the operation of this algorithm is provided below.
- The Recommender Module provides the user with guidance on the optimal learning path based on their performance in previous sessions. Further details are discussed below.
- The Assessment Module assesses the user during the *Training mode* and tracks execution times and errors for each step. It updates the User Model data and informs the Recommender Module about the most appropriate learning path for the user. It also supports the Presenter Module with the assessment data required to determine the user's expertise level for each step of the procedure.
- The Adaptive Manager serves as the central element responsible for orchestrating messages and requests between all modules of the Adaptive System as well as managing external communication with the other systems.The Recommender Module helps the users select the most appropriate procedure and execution mode at the beginning of each session to optimize their learning experience. The three possible clinical interventions in DAM (external ventilation, orotracheal intubation and cricothyrotomy) have an increasing level of complexity. Therefore, the users are encouraged to engage with all interventions to gain a holistic understanding. For instance, they might start in *Learning mode* with full support before moving to *Training mode* where guidance is reduced although feedback is still provided. If challenges arise during a scenarios's *Training mode*, resulting in unsuccessful completion or numerous errors,

the Recommender Module advises returning to *Learning mode* for that scenario to improve mastery and understanding.

The Presenter Module is responsible for determining the expertise level of the users at each procedure step *i*, which in turn adjusts the verbosity of the VI and the amount of additional information provided during training. To accomplish this task, the Presenter Module uses data collected during previous *Training* sessions. This data includes the number of repetitions (*r*) and the minor (*e*) and major (*E*) errors made at step *i*. These parameters are weighted with constants (k_r, k_e and k_E) to calculate the Expertise Score for step *i* (S_i) as follows:

$$S_i = (r_i \cdot k_r) - (e_i \cdot k_e) - (E_i \cdot k_E) \tag{1}$$

Based on this score, the transitions between user's Expertise Levels for each step (L_i) are determined as follows:

$$L_i = \begin{cases} Beginner, & S_i < t_{int} \\ Intermediate, & t_{int} \leq S_i < t_{exp} \\ Expert, & S_i \geq t_{exp} \end{cases} \tag{2}$$

where t_{int} and t_{exp} represent the threshold values for advancement to the Intermediate and Expert levels respectively.

To best support users with different levels of expertise, the Presenter Module provides level-specific instructions for each step of the procedure. This ensures that Beginners receive detailed instructions, while Intermediate and Expert users receive increasingly concise instructions. Table 1 shows as an example the three verbosity levels in one of the procedue steps, namely the one involving drug administration.

Table 1 Example of descriptions provided by the VI for the drug administration step, based on the user's expertise level

Expertise level	Virtual instructor description
Beginner	"Before intubation, you always need to sedate the patient. There are several drugs used for sedation prior to orotracheal intubation. In this specific situation, we will use Fentanyl and low-dose Midazolam, which are proven to result in significantly faster patient recovery. Use a dosage of Fentanyl of 1–3 mcg/kg. Midazolam is administered with an initial dosage of 2 to 2.5 mg 5 to 10 min before intubation, then increased with a rate of 1 mg per minute to a maximum of 7 mgs per course. You can find both drugs and the syringe on the table. Fill the syringe and administer both drugs intravenously."
Intermediate	"To sedate the patient, administer the patient both fentanyl and midazolam. You can find both drugs and the syringe on the table. Administer drugs intravenously."
Expert	"Sedate the patient."

The values of the hyperparameters (k_r, k_e, k_E, t_{int} and t_{exp}) within the ERE framework can be adjusted to modulate the adaptive behavior of the Presenter Module and ensure that specific operational requirements are met. Ideally, these values should be optimized through systematic validation involving multiple iterations of experiments to determine the combination that maximizes learning efficiency while ensuring a balance between sensitivity and stability in transitions between expertise levels.

However, finding and validating the optimal values for these hyperparameters was beyond the scope of this study, mainly due to time and organizational constraints that limited the number of training iterations available for experimentation. Therefore, we chose a set of arbitrary values (see Sect. 4.1) designed to ensure that participants experienced noticeable transitions between expertise levels during the experiments so that we could evaluate the potential of the adaptive framework within the constraints of the study.

3.4 Implementation details

The application was developed with the Unity game engine leveraging its "write once, deploy anywhere" approach, which enables seamless deployment on multiple platforms. In our experiments, the Meta Quest 2 HMD was used and Microsoft Azure's text-to-speech API was integrated into the application.

4 Experimental protocol

We performed a user study to validate the following RHs:

- **RH1:** Our VR application effectively improves the participants' knowledge of the DAM procedure compared to a control group trained using traditional methods.
- **RH2:** Our adaptive framework is more effective for teaching compared to a non-adaptive version of the same application.Our study involved 46 physicians in their first or second year of specialization in emergency medicine from the Molinette Hospital in Turin, Italy. Recruitment was on a voluntary basis, without any form of compensation, and was motivated by the opportunity to contribute to the validation of an innovative VR-based training tool. All participants gave informed consent before the start of the study. Physicians were randomly assigned to one of the two main groups: the VR group (28 subjects) and the Control group (18 subjects). To validate our second hypothesis, the VR group was randomly divided into two subgroups, the Adaptive

and Static subgroups, each with 14 subjects who experienced the VR application with and without the adaptive system, respectively.

Detailed demographic information about the subjects are reported in Table 2. It is worth noting that the majority of the subjects in the VR group were unfamiliar with VR technology. Overall, 26 subjects (92.86%) had no prior experience with VR and only two (7.14%) had used it before but only to a limited extent. In terms of experience with video games, 21 subjects (75.00%) had never played, five (17.86%) rarely played, one (3.57%) played somewhat frequently and only one (3.57%) played regularly.

All experimental subjects were native Italian speakers, except one, who was a native Spanish speaker. Although knowledge of English was expected, some older subjects had difficulty understanding and communicating with the VI. This problem was only identified after the experiment, as language proficiency was only assessed by self-evaluation and no formal assessment had been carried out beforehand. It is important to emphasize this point as it might have influenced the VR group results, as discussed later in Sect. 5.1.

Participants in the VR group received training with the VR application, while the Control group received a traditional lecture with identical content, delivered by medical experts using oral explanations, videos, and slides. The experiment followed slightly different protocols for each group, which are detailed in the following Sections.

In designing the experimental protocol, we limited the time spent in VR to less than one hour and scheduled breaks between each block to reduce potential overexertion and cybersickness associated with VR use. This approach took into account the participants' time availability and ensured comparable duration of the experimental protocol between the Control and VR groups. However, it is important to note that these experimental conditions did not allow for a comprehensive evaluation of the Recommender Module, as its proper assessment requires data collection across multiple iterations. The two experimental blocks conducted in this study did not provide sufficient data for such an evaluation, which will therefore be addressed in future work.

Under this premise, and after consultation with the experts involved in the study, we decided to focus the experimental learning activities on the intubation intervention (see Fig. 1), which was identified as crucial in the DAM scenarios. This focus enabled a targeted investigation of common intubation errors and included essential skills such as external ventilation. The more complex cricothyrotomy intervention was excluded in this experimentation due to time constraints and its relative rarity in medical practice (it is required in less than 1.1% of cases according to Sakles et al. (1998)).

4.1 Protocol-VR group

The protocol for the VR group is illustrated in Fig. 7. A confederate acted as a facilitator and introduced each subject to the VR scenario and the hardware used (i.e., the HMD and the handheld controllers). Upon arrival, the subjects provided informed consent as well as basic personal information, and completed a (custom) Pre-Knowledge Test with ten open-ended questions to assess their prior knowledge of DAM. This questionnaire was developed by the authors based on the main procedural concepts and clinical guidelines for DAM as described in the relevant literature. Although the questionnaire has not been formally validated in previous studies, it was reviewed by domain experts to ensure content relevance and consistency with the study objectives.

Afterwards, they completed the mandatory tutorial described in Sect. 3.3.1, which could be repeated until they were familiar with the VR application and its interactions.

The VR experiment consisted of two blocks, each containing a Learning (L) phase and a Training (T) phase. Each block's L and T phases will be distinguished with the number 1 or 2 (i.e., L1 and T1 for block 1). In block 1, the participants must intubate a patient and deal with unilateral bronchial intubation (a common and dangerous intubation error). The procedure in block 1 comprises 32 individual steps. In block 2, the patient has different symptoms than in block 1 and a foreign body obstruction. The participants this time have to handle esophageal intubation (another common intubation error). The procedure to follow in block 2 comprises 37 steps, 30 of which are in common with that of block 1. We recall that a session in *Training mode* is marked as "failed" and then interrupted when the participant makes a major error during the procedure i.e., a mistake that puts the patient's life in immediate danger.

For the Adaptive subgroup, the expertise level of the participants is initialized to "Beginner" for each procedure step

Table 2 Experimental subjects' demographic data

Group	Subgroup	Number	Male	Female	Age (range)	Age (mean)
VR	Adaptive	14	8	6	26–54	31.71
VR	Static	14	2	12	26–45	30.43
Global (VR)	–	28	10	18	26–54	31.07
Control	–	18	11	7	26–40	30.17
Global (All)	–	46	21	25	26–54	30.72

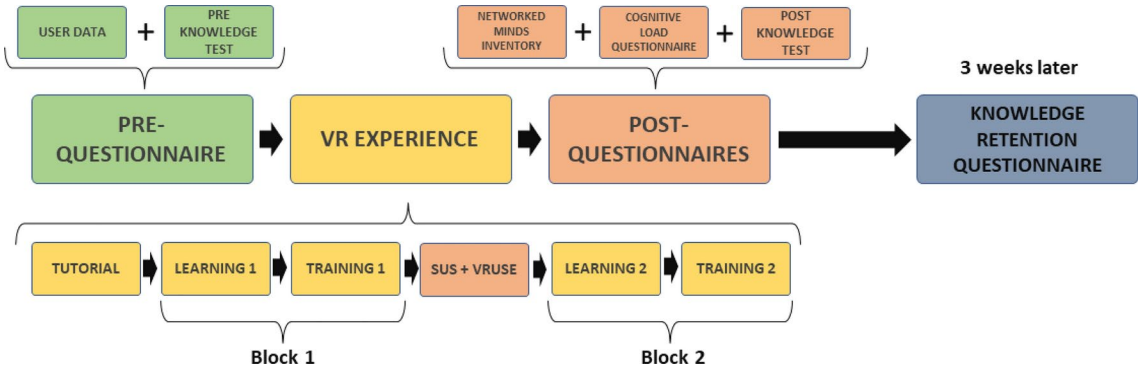


Fig. 7 The experimental protocol used for the VR group

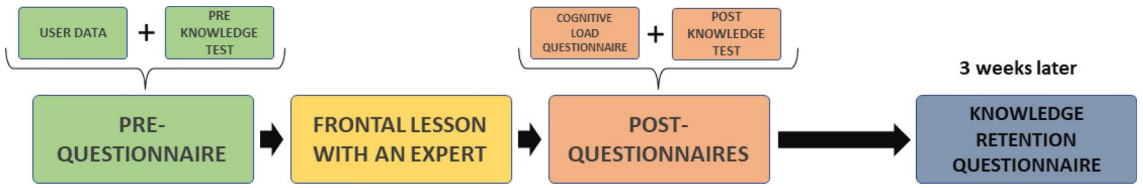


Fig. 8 The experimental protocol used for the Control group

and the following constant values were used for expertise level computation (Eq. 1): k_r=1, k_e = 0.5, k_E = 1, t_{int} = 1, t_{exp} = 3. This configuration enabled the participants to progress to the "Intermediate" level unless they made more than three minor errors or one major error during that step. If the participant made a major error in step *j* (leading to the interruption of the *Training mode* in *Block 1*), the expertise level remains "Beginner" from step *j* onwards. In this way, we ensured that the participants could experience the changes the adaptive system made to the explanations in just two blocks.

To assess the usability of the application, the participants were provided a short break between the two blocks and were asked to complete a questionnaire consisting of the System Usability Scale (SUS, Brooke (1996)) and selected sections of the VRUSE (Kalawsky 1999) (*input*, *fidelity* and *immersion/presence*). After completing the second block, we administered a cognitive load questionnaire taken from Leppink et al. (2013) and Leppink et al. (2014). The questionnaire includes four introductory questions (scored on a 9-point Likert scale), and 13 other questions to assess three distinct categories of cognitive load: *intrinsic* (the inherent difficulty of the task), *extraneous* (external factors such as the material and environment that may influence learning) and *germane* (the ability of working memory to combine new information with existing knowledge), all scored on a 5-point Likert scale. Additionally, we administered the Networked Minds Social Presence Inventory (Biocca et al. 2001) (NWM) to assess the effectiveness of interactions with virtual agents. The questions related to Perceived Emotional Contagion were intentionally excluded as they were irrelevant to our research objectives.

Finally, all participants completed post-experiment assessments (Post-Knowledge Test, with the same eight questions as the Pre-Knowledge Test) to measure the participants' improvement, and a follow-up questionnaire that was administered after three weeks to assess long-term knowledge retention. The entire experimental session lasted 1.5 to 2 h, with *Block 1–2* lasting an average of 50.75 min and the introductory explanations, tutorial mode and all questionnaires taking up the remaining time.

In addition to the subjective data, quantitative data was collected throughout the experiment to analyze participant performance, interaction patterns, and error rates.

4.2 Protocol-control group

The protocol for the Control group, shown in Fig. 8, was similar to that of the VR group. Instead of participating in a VR experience, the Control group attended a lecture given by a domain expert. This lecture covered the same topics as the VR application. The Control group was administered the Pre-Post Knowledge and the Cognitive Load questionnaires. The lecture lasted approximately 45 min and was followed by 20 to 30 min for completing the questionnaires. Finally, all participants took part in the post-experiment assessment and the long-term follow-up.

4.3 Data analysis

Pairwise comparisons of questionnaire scores and qualitative data were conducted between the VR and Control group to verify RH1, and between the Adaptive and Static subgroups to verify RH2. Pairwise comparisons were performed with Student's t-text, Welch's t-test or Mann-Whiteny U rank test depending on data normality (checked with Shapiro-Wilk test) and variance equality (checked with F-test). For knowledge improvement, 2-way ANOVA was used with group and administration time as independent variables. Also, correlation analysis between selected variables was carried out using Pearson's correlation coefficient.

5 Results

In the following, we discuss the experimental results, starting with the study of usability and user experience as fundamental aspects, since the effective adoption of VR technologies depends on their accessibility and intuitiveness. By analyzing the feedback via questionnaires and participant comments, we create a baseline usability assessment that serves as a foundation for the following detailed analysis of the educational impact of the application and for the validation of our key research hypotheses.

5.1 Usability and user experience

In this Section, we examine the usability and user experience (UX) of our VR application from four perspectives: **agent interaction**, **usability**, **VR UX**, and **user comments**, which provide direct qualitative insights from the participants.

5.1.1 Agent interactions

The number and type of vocal interactions with the virtual agents played a crucial role in our design and likely influenced both usability and UX. In the following, we discuss separately the interactions that are mandatory in the procedure and those that are optional.

Mandatory Interactions. Table 3 summarizes the average number of interactions per phase. While the number of mandatory interactions to progress in the procedure was seven (four with the virtual nurse and three with the virtual patient; interactions with the VI are not mandatory), the average number of actual interactions in *Block 1* was 12.39 for L1 and 9.71 for T1, whereas in *Block 2* was 11.46 for L1 and 8.82 for T2. Thus, it is clear that the participants engaged with the virtual agents more than required (with no significant difference in the number of interactions between the Adaptive and Static subgroups). By analyzing the application logs, we found that the main issues faced by the participants were misassignments of messages and unrecognized intents.

Misassignments relate to a problem in our design of voice interactions, which forwards the recognized intents to the agent the participant is currently looking at. However, a challenge arises from the delay between the participant's spoken communication and the system's recognition of intents. If the participant directs their gaze to another agent while the system is still processing the voice input, the communication could be incorrectly attributed to the unintended agent. One possible solution to mitigate this problem is to incorporate a naming convention into the interaction design. By requiring participants to name the avatar they want to communicate with before they deliver their message (e.g. by saying "*Nurse, attach the electrodes*" or "*Patient, can you hear me?*"), the system can route the communication to the correct avatar.

Unrecognized intents refer to cases where the participant's message is incorrectly interpreted by the intent recognizer. The high standard deviation of the number of agent interactions suggests that the cause of these errors lies in the variability of the participants' linguistic interaction abilities rather than in the limitations of the intent recognizer. In support of this hypothesis, we found a significant correlation between the number of agent interactions and the participants' age (Pearson's correlation coefficient, $r = 0.63$, $p < 0.05$), suggesting that younger participants were more proficient in English or more familiar with conversational interfaces. Nevertheless, the participants appeared to become accustomed to the agent interaction system over time, as shown by the average decrease in mandatory interactions between the two blocks.

Optional Interactions. During the *Learning mode*, the participants had the option of communicating verbally with the VI, e.g., asking for repetitions of steps, inquiring about the correct tool to use, or requesting additional details. These options were presented during the tutorial and made easily accessible within the application via a diegetic interface, i.e., a virtual sheet on which the available verbal

Table 3 Mean number and standard deviation of mandatory agent interactions, per *mode* (L/T) and *block* (1/2)

	L1		T1		L2		T2	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Global	12.39	5.10	9.71	3.82	11.46	3.77	8.82	2.88
Adaptive	12.14	5.24	9.64	4.05	10.86	3.87	8.86	2.88
Static	12.64	5.14	9.79	3.72	12.07	3.69	8.79	2.99

Table 4 NWM categories, scores and standard deviations

Category	Global		Adaptive		Static	
	Mean	SD	Mean	SD	Mean	SD
Co-Presence	4.83	1.11	4.71	1.27	4.95	1.27
Perceived Attentional Engagement	4.88	1.27	4.54	1.28	5.21	1.28
Perceived Comprehension	4.80	1.35	4.56	1.21	5.05	1.21
Perceived Behavioural Interdependence	4.29	1.04	4.40	1.03	4.18	1.03

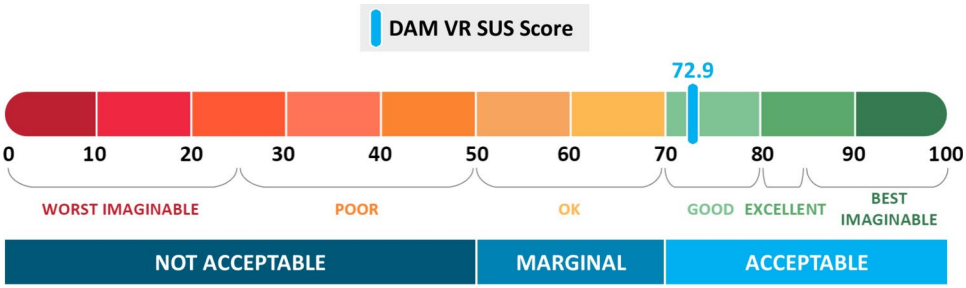


Fig. 9 SUS Questionnaire result, the blue bar indicates the usability score of our application

interactions are listed. However, interactions with the VI were rare. The "*Repeat*" command was used 20 times (14 times in L1, six times in L2) by nine participants, "*Ask for tool*" was invoked four times (three times in L1, one time in L2) by two participants, and the "*Details*" request was made only once (in L1).

The limited interactions with the VI could be attributed to the design of the scenario. The spoken step instructions were also displayed in a text box so that the participants could read them at their convenience. The "*Ask for tool*" command may have been overlooked since the explanations of the tools were often embedded in the video demonstrations, so selecting and using the tools was relatively intuitive. The limited "*Details*" requests probably indicate a suitable clarity of the instructional contents for each step.

Social Presence. The results of the NWM (Table 4) show that the mean scores for the four categories (Co-Presence, Perceived Attentional Engagement, Perceived Comprehension, and Perceived Behavioral Interdependence) ranged from 4.29 to 4.88 on a seven-point Likert scale. These results suggest that the virtual agents could foster a certain degree of social presence in the participants. However, the outcomes were not optimal, pointing out an area where our software can still be improved.

We believe these limitations may be attributed to the interaction and communication issues previously described, i.e., misassignments and unrecognized intents. The implemented NLP system required that the participants looked at a specific agent before speaking to them, and the agents could process only a limited number of conversation intents. Such a system was probably unable to live up to the level of naturalness expected by the participants, to the point that one participant suggested using keywords instead of natural language. Also, the fact that the participants were non-native English speakers may have played a role in the inability of the participants to communicate naturally, as their intents were not always recognized.

The analysis revealed no significant correlation between the NWM results and other variables and no statistically significant difference between the Adaptive and Static subgroups.

5.1.2 System usability assessment

After Z-scoring the SUS results and removing outliers (using the threshold of ± 1.96), the average SUS score was 72.9, which is only slightly above the benchmark score of 68 for good usability (Bangor et al. 2008, 2009). We believe that various elements contributed to this somewhat less than satisfactory result.

First, we believe that the SUS score not only reflects the suboptimal usability of the application, but may also be due to the complexity of the procedure (which comprises up to 37 individual steps). Another factor that influenced the results was the timing of the administration of the SUS questionnaire, i.e., during the break between *Block 1* and *Block 2*. The reason for this decision in the design of the experimental protocol was to obtain objective assessments of the initial challenges faced by the participants that were not influenced by the intensive use of the application. This information was intended to be used for gathering insights to further improve the usability of the application for first-time participants. However, our observations during the experiments showed that the more participants used the application, the more confident and comfortable they became with the procedure and technology. Therefore, it is plausible that SUS scores would have been higher if the questionnaire had been administered at the end of the experience.

Finally, we found a negative correlation between the number of errors made by the participants and SUS scores

($r = -0.46$ and a $p < 0.05$). In other words, participants with lower proficiency tended to rate the application lower than participants with higher proficiency. This finding suggests that the application should be improved to better support the participants who struggle with the learning curve. One possible option to address this issue in the future is to integrate AI-based intelligent feedback to provide real-time guidance during the VR practice. This will provide the participants with immediate insights into their performance, with the AI highlighting areas for improvement and providing timely advice.

5.1.3 VR user experience (VRUSE)

To evaluate the UX, the participants were asked to complete selected sections (*input*, *fidelity* and *immersion/presence*) of the VRUSE questionnaire evaluating items on a five-point Likert scale. For the analysis, we computed the overall scores of each usability factor and the diagnostic factors, as in Kalawsky (1999). Results are detailed in Table 5. In general, the VRUSE results were positive, indicating that the VR application was well received by the participants. Most factors scored high, indicating that the application was user-friendly (Ease of Use: 3.79), suitable for the task (Appropriateness: 3.75), and performed effectively (System Performance: 3.83). The scores for Immersion (4.21), Presence (4.03), Global Fidelity (4.11), and Global Immersion/Presence (4.39) were remarkably high, suggesting that the participants were effectively engaged with the virtual environment.

However, there were also some areas of concern. Intuitiveness received moderate scores (3.03). This result could indicate that while the participants generally found the application easy to use and suitable for the task, they required a more complex learning curve than expected. In addition, the standard deviation in the Disorientation factor is higher than in the others, implying that some participants may have felt more disoriented than others. This could be partly due to the participants' lack of experience with VR technologies.

These results may help reveal some factors contributing to the moderate SUS rating. In fact, we found a very high correlation coefficient between SUS scores and intuitiveness ($r = 0.77$, $p < 0.05$) and disorientation ($r = -0.74$, $p < 0.05$). This is because an application that is not sufficiently intuitive may require the participants to spend more time learning or adapting to it, which may negatively affect their perception of overall usability (Kalawsky 1999). In addition, the feeling of disorientation experienced by some participants when using the application has a negative impact on the overall usability.

Table 5 Usability and diagnostic factor scores and Standard Deviations from the VRUSE questionnaire (1–5 Likert scale)

Category	Mean	SD
Ease of Use	3.79	0.50
Appropriateness	3.75	0.37
System Performance	3.83	0.48
Input Sensitivity	2.00	0.94
Functionality	4.14	0.59
Intuitiveness	3.03	0.99
Disorientation	2.32	1.22
Immersion	4.21	0.79
Presence	4.03	0.63
Global Input	4.36	0.56
Global Fidelity	4.11	0.78
Global Immersion/Presence	4.39	0.68

5.1.4 User comments

To gain a deeper understanding of the participants' experience with the VR application, we collected unstructured feedback in the form of open comments. Of the 28 participants in the VR group, 17 provided comments, which were analyzed qualitatively to identify general sentiment and recurring themes and keywords.

Of the comments received, 15 were positive, one was neutral and one was negative. Eight participants described the application as "useful", while five described it as a "good summary of the topic". Some of the most enthusiastic feedback includes the following:

- *"Amazing and highly interactive application. Perfect for beginners who are not familiar with emergency medicine procedures. It is also a great tool for refreshing topics for more experienced users. Thank you, and keep up the good work with this incredible project. I had a lot of fun and found it extremely useful."*
- *"This was my first experience with VR. Nonetheless, I found the simulation to be an excellent review of topics I had already covered in the past. This VR simulation could be a valuable tool for learning, in addition to traditional studies."*
- *"A very interesting and useful tool for learning. The simulation teaches practical techniques that are used daily in emergency medicine. I recommend everyone to experience this type of teaching at least once. Thank you very much."*We also received constructive feedback and suggestions for improvement from five respondents. The points of criticism were as follows:

- *"Make the Glasgow answers clearer."*
- *"The labels on the drugs were too big."*
- *"Maybe you should use keywords instead of full sentences when interacting with the avatars."*
- *"The patient should be more dynamic and interactive."*

- *"It is not easy to realize that you have made a mistake."*
- *"The video of the real physician performing the operation should be shown in isolation and separately, it is not easy to focus on it while everything else is going on."*
- *"The tools should be closer to the bed."*
- *"I could not concentrate on the instructor because he was speaking in English, I was just reading the text."*This feedback provides valuable insights for improving the UX in future iterations of our VR application.

5.2 RH1: effectiveness of the VR application in teaching DAM procedures

In this Section, we examine the results that can support the validation of RH1. Specifically, we compare the knowledge improvement and cognitive load between the Control group and the VR group, without distinguishing between its Adaptive and Static subgroups. It is worth recalling that the duration of the experiences was intentionally aligned to ensure a balanced comparison between the groups. The participants in the VR group took an average of 50.75 min to complete all *Learning* and *Training modes* of the program, which is very similar to the duration of the Control group's traditional lecture, lasting approximately 45 min.

Knowledge Improvement. The results of the knowledge questionnaires (i.e., pre-test, post-test, and long-term) for the VR and Control groups are shown in Table 6. Both groups showed a significant improvement in their knowledge. For the Control group, the mean scores improved from 4.94 (pre-test) to 6.06 (post-test) ($p < 0.05$). For the VR group, the mean scores improved from 4.04 (pre-test) to 6.39 (post-test) ($p < 0.05$). The improvement in knowledge between the pre-test and post-test for the VR group (+29.46%) was significantly higher than that of the Control group (+13.89%) ($p < 0.05$).

Table 6 Results of the knowledge questionnaires.

Category	Pre-Post	Imp	Post-LT	Imp
VR	4.04–6.39	29.46% (89.28%)	6.39–5.86	−6.7%(−42.85%)
Control	4.94–6.06	13.89% (72.22%)	6.06–5.78	−3.5% (−33.33%)
VR - Adaptive	4.21–5.93	21.43% (85.71%)	5.93–5.43	−6.2% (−42.85%)
VR - Static	3.86–6.86	37.50% (92.85%)	6.86–6.29	−7.1% (−42.85%)

Pre-Post: mean score of all users between the preand post questionnaire. Post-LT: mean score of all users between the post and the long-term questionnaire. **Imp:** improvement for the two pairs under test (Pre-Post and Post-LT), expressed as percentage increase in mean score and in () as the percentages of users who exhibited an improvement. A negative number means that the score has decreased from one questionnaire to another

We note that the observed differences in learning outcomes are not only influenced by the VR training itself, but also by the differences in the experimental settings of the two groups. Nevertheless, the results underline the potential of VR training to enhance declarative and procedural knowledge compared to traditional, lecture-based methods. Both modalities target cognitive learning and focus on understanding the "what" and "how" of DAM procedures. While psychomotor skills are essential for mastering DAM, they require hands-on practice with physical simulators regardless of whether the initial cognitive training is delivered via VR or traditional lectures, which is outside the scope of this study. This distinction illustrates that VR and lecture-based training are comparable in terms of cognitive outcomes. Future studies could compare the practical skills of participants trained through VR or lectures by assessing their performance on manikins, providing deeper insights into how VR complements hands-on training.

Regarding knowledge retention, the questionnaires showed a slight decrease over time of knowledge in both the Control group and the VR group, although there was no significant difference between the two groups ($p > 0.05$). This result is consistent with other works that have examined the long-term efficacy in knowledge retention after training with high-fidelity simulators (Boet et al. 2011). Similar to our approach, these simulators enable the implementation of realistic and engaging simulation environments that mirror the clinical setting and allow for a deeper and lasting learning effect.

Cognitive Load. The cognitive load questionnaire showed a marked distinction between the VR and Control groups in the perceived complexity of the instructional module and the mental effort invested (Table 7). These differences were particularly evident in the answers to three specific questions, namely Q2: *In the instructional module just completed, I invested a mental effort/load that I would define as:* [Very Low - Very High], 7.25 vs. 4.56, $p < 0.05$, Q3: *The instructional module just completed was:* [Very Simple - Very Complicated], 5.75 vs. 3.72, $p < 0.05$, and Q4: *Learning with this instructional system was:* [Very Simple - Very Complicated], 4.21 vs. 6.89, $p < 0.05$. In summary, the VR instructional system was perceived as easier (Q4), while the Control group perceived the topic as less complicated and requiring a lower concentration level (Q2, Q3).

Several reasons can be given for these results. The intrinsic cognitive load, which indicates the inherent difficulty of the task, was similar in the two groups. This result suggests that the perceived differences were not due to the complexity of the contents, but rather to how they were taught. The Control group may have perceived the topic as less complicated because they were familiar with traditional learning

Table 7 Scores and standard deviations from the cognitive load questionnaire

Question	Scale	VR		Adaptive		Static		Control	
		Mean	SD	Mean	SD	Mean	SD	Mean	SD
Q1	1–9	5.75	1.86	5.71	1.27	5.79	2.36	4.78	1.52
Q2	1–9	7.25	1.76	7.14	1.86	7.36	1.86	4.56	1.34
Q3	1–9	5.75	1.58	5.57	1.77	5.93	1.77	3.72	1.53
Q4	1–9	4.21	2.27	4.00	2.44	4.43	2.44	6.89	1.64
Q5–8 (Intrinsic)	1–5	2.29	0.77	2.29	0.71	2.29	0.71	2.18	0.67
Q9–12 (Extraneous)	1–5	1.51	0.62	1.34	0.68	1.68	0.68	1.25	0.39
Q13–17 (Germane)	1–5	3.61	0.86	3.67	0.83	3.54	0.83	3.51	0.76

The first four introductory questions are scored separately. Scores from questions 5 to 17 are grouped and averaged based on the type of cognitive load

Table 8 Completion times and percentage of steps completed per session

	Completion Times (mins)				Steps Completed			
	L1	T1	L2	T2	L1	T1	L2	T2
Global	20.41	9.45	13.60	7.30	100%	66.63%	100%	69.21%
Adaptive	19.25	8.92	11.50	6.89	100%	67.63%	100%	69.88%
Static	21.56	9.98	15.70	7.72	100%	65.63%	100%	68.53%

methods. Although there is no significant difference, the lower extraneous cognitive load reported by the Control group could also indicate that the traditional learning environment contained fewer distracting elements, which resulted in the complexity being perceived as lower. In contrast, it is possible that the VR group perceived the instructional system as simpler because VR is more engaging and interactive than traditional learning methods. This perception, in turn, facilitated a more intuitive understanding of the learning material.

The analysis of the combined score of all three types of cognitive load (intrinsic, extraneous and germane) shows no significant statistical difference between the VR and Control groups. Considering that all participants in the VR group were novices with VR technology, the lack of a significant effect in the extraneous dimension, which includes factors peripheral to the learning content that could potentially hinder the learning process, suggests that the VR device itself did not interfere with the learning experience (which could also be an indication of the effectiveness of the initial tutorial session). Furthermore, no statistically significant differences were found between the Adaptive and Static subgroups in terms of cognitive load.

5.3 RH2: efficacy of the adaptive framework

In this Section, we discuss RH2 by comparing the results of the two VR subgroups (Adaptive and Static) in terms of completion time, knowledge improvement, number of errors and cognitive load.

Completion Time. The results shown in Table 8 show that the average completion time decreased significantly ($p < 0.05$) between the first and second *Learning mode* in the VR group (L1: 20.41 min, L2: 13.60 min), although *Block 2* included more steps (37 compared to 32), indicating an improvement in the participants' knowledge and skills. Specifically, L1 completion times were similar for the Adaptive and Static subgroups, but L2 completion times were significantly different (Static: 15.70 min, Adaptive: 11.50 min, $p < 0.05$). This result was expected as the VI's explanations for the Adaptive subgroup were generally much shorter in *Block 2* due to the participants' improved expertise levels. Interestingly, we also found a direct correlation between the age of the participants and the duration of the *Learning mode* sessions (L1: $r = 0.66$, $p < 0.05$ and L2: $r = 0.55$, $p < 0.05$). One possible explanation is that younger participants were more comfortable with VR technology, thus shortening interaction times.

Regarding the duration of sessions in *Training mode*, although major errors could result in each participant performing a different number of steps in each block, we found similar average percentages of completed steps in T1 and T2 for the Adaptive and Static subgroups, allowing for a fair comparison. Under this premise, we found a decrease in average completion time between T1 and T2 for the VR group (from 10.68 to 7.30 min, $p < 0.05$). The higher number of steps in T2 again suggests that the participants' skills improved between the two blocks. When comparing the two subgroups, we found that the Adaptive subgroup was faster in both *Training mode* sessions, but without a significant difference when compared to the Static subgroup ($p > 0.05$ for both T1 and T2).

Knowledge Improvement. A comparison between the knowledge improvement (pre-post test comparison) of the two VR subgroups, Static and Adaptive, showed greater knowledge improvement in the Static subgroup (37.5%)

Table 9 Training mode session completion data for T1 and T2

Training		Number of Users			Percentage of Users		
T1	T2	Global	Adaptive	Static	Global	Adaptive	Static
✓	✓	5	1	4	17.9%	7.1%	28.6%
✗	✗	10	3	7	35.7%	21.4%	50.0%
✗	✓	7	5	2	25.0%	35.7%	14.3%
✓	✗	6	5	1	21.4%	35.7%	7.1%

A (✗) indicates a *Training mode* session where a major error was made. A ✓ indicates a *Training mode* session completed with no errors

Table 10 Mean number and Standard Deviation of minor errors for each *Training mode* session

Group	T1		T2		
	Mean	SD	Mean	SD	Δ
Global	7.86	6.47	15.07	10.74	+7.21
Adaptive	5.86	5.02	13.50	8.89	+7.64
Static	9.86	7.28	16.64	12.46	+6.78

The Δ column reports the increase in mean errors from T1 to T2

compared to the Adaptive subgroup (21.43%). Although the T-test did not show statistical significance ($p > 0.05$), the result is consistent with expectations, as the Static subgroup received a more detailed explanation twice, whereas the Adaptive subgroup received (in most cases) a concise version in L2. This outcome suggests that more than two repetitions may be needed for novice users to move from the "Beginner" to the "Intermediate" level.

As stated in Sect. 5.2, the results in Table 6 show no significant difference between the two VR subgroups in terms of knowledge retention.

Cognitive Load. No significant difference in cognitive load was observed between the Adaptive and Static subgroups (see Sect. 5.2).

Proficiency of the VR Users. In addition to the improvement in knowledge, the overall performance of the VR group can also be assessed based on the results of the *Training mode* sessions and, in particular, based on the participants' major and minor errors. These results can be found in Tables 9 and 10, respectively.

Major errors were defined as procedural errors that led to the termination of the session. Of the 28 participants, only five successfully completed both *Training mode* sessions (four Static, one Adaptive), while 10 (seven Static, three Adaptive) failed both sessions, with no significant difference between the different VR subgroups in terms of failure and success rates. Interestingly, only two of these 10 participants repeated the same serious error twice (both forgot to ventilate the patient before sedation), suggesting that emphasizing serious errors effectively discouraged their repetition.

It is important to point out that major errors were specifically related to the participants' understanding of the material and procedural knowledge, rather than interaction-related issues. Interaction-related actions, such as selecting the wrong tool, were not categorized as major errors unless they directly led to incorrect procedure execution, i.e., using the selected tool to perform an incorrect action in the current procedural step. This distinction suggests that the observed major errors reflect participants' cognitive and procedural learning rather than difficulties with the user interface.

The analysis of minor errors revealed that the Static subgroup committed more minor errors in both sessions, with the increase in error rates between the two blocks being similar for the two VR subgroups (+6.78 for Static and +7.64 for Adaptive). Overall, the average number of minor errors increased from 7.86 in T1 to 15.07 in T2.

A detailed breakdown shows that most of the minor errors were related to the labeling of symptoms on the user interface (60.20% in T1 and 54.32% in T2). We hypothesize that this increase is likely due to a "trial-and-error" approach, as participants assumed that these errors had no real consequences after the first session and were not related to the interaction (as the marked symptoms had to be explicitly confirmed by the user before being sent to the system). Importantly, the pre- and post-session knowledge assessments showed significant improvements in symptom identification, such as the Glasgow Coma Scale and Mallampati, suggesting that participants developed effective procedural knowledge despite these errors.

Other notable minor error clusters were related to oxygen mask positioning, which had to be performed several times during the procedure and accounted for 15.00% of errors in T1 and 15.88% in T2, and errors related to communication with the virtual nurse, which accounted for 13.65% in T1 and 6.64% in T2 (with a mean of 1.07 errors in T1 and 1.0 errors in T2). The remaining errors were distributed across all other steps. Although the Static subgroup generally committed more errors, our analysis revealed no significant difference between the two subgroups in each *Training mode* session.

5.4 Discussion

The results of our study and their implications for our VR-based learning application can be described as follows.

RH1 states that VR users would learn the DAM procedure more effectively (and with a similar cognitive load)

than the Control group using traditional learning methods. The experimental results support this hypothesis and show that despite the differences in the training settings, the VR group achieved a greater increase in knowledge, 29.46% compared to 13.89% in the Control group. In terms of knowledge retention, both groups showed similar rates of knowledge decay, suggesting that VR training does not necessarily provide better long-term knowledge retention compared to traditional methods. The same results show that the application was well received, indicating a positive user experience, with an overall SUS rating of 72.9. VRUSE measure ratings for Input (4.36), Fidelity (4.11), and Immersion/Presence (4.39) were also positive. However, the data also revealed areas for improvement. The lack of debriefing, the limited feedback mechanisms and the problems in communicating with the virtual agents are aspects that, if improved, could potentially increase the educational impact of our application.

RH2 assumes that the VR Adaptive subgroup would outperform the Static one in terms of efficiency and cognitive load, based on the ERE model adopted in the Adaptive module. Contrary to expectations, there was no significant difference between the Adaptive and Static subgroups regarding error rate or knowledge acquisition. These results challenge the second hypothesis and suggest that adaptability in instructional design did not play a significant role in learning outcomes, at least not within the experimental design of this study. Indeed, as described in the literature (Billings 2012; Serge et al. 2013), it is difficult to test the usefulness of adaptive systems with such a short experiment and the actual value can only be assessed "in the long run". This observation suggests that the ERE may not have had an obvious effect due to the novice status of our participants, implying that a gradual decrease in instructional support rather than an immediate one, as in our experiments, may be more effective for users new to the medical procedure. We also believe that novice users would benefit from different iterations of the *Learning mode* (with slight variations of key parameters to avoid repetition) before venturing into the *Training mode* to better familiarize themselves with the medical procedure and the VR application.

5.5 Limitations

The participants' lack of familiarity with VR technology may have been a confounding variable that could have affected the learning experience and usability of the application. Language barriers also made interpreting the instructions and feedback difficult, leading to increased cognitive load and usability issues for some participants. In addition, the "novelty effect" of VR may have possibly influenced the participants' engagement and their subjective evaluation of the learning experience.

Another point that deserves attention is the limited time frame of the experiments. This limitation prevented a thorough analysis of the optimal calibration of the parameters used to calculate the Expertise Level and thus an in-depth evaluation of the advantages and disadvantages of AL and AT in this particular context. Furthermore, due to the same limitation, it was not possible to evaluate the Recommender Module.

Finally, the study focused on the comparison of VR and lecture-based training in terms of declarative and procedural knowledge, which are essential cognitive aspects of DAM training. However, psychomotor skills, which are crucial for mastering DAM, require hands-on training with physical simulators such as manikins. Future work should include an assessment of practical skills through hands-on testing after cognitive training to provide a more comprehensive evaluation of how VR training complements traditional methods in preparing healthcare professionals.

5.6 Future work

Regarding usability and UX, while responses were generally positive and indicated a user-friendly interface, the need for more intuitive interactions and improved feedback mechanisms became apparent. Future developments will focus on improving the naturalness of verbal interactions. The ability to offer multilingual features for both scaffolding and verbal interactions would reduce the barriers associated with the participants' language skills. Future versions should also include a comprehensive debriefing component, critical for consolidating learned concepts and reflecting on the learning experience.

The current study focused only on the procedural aspects of DAM. Future studies could expand this focus to include hardware specific to psychomotor skill development, which could significantly impact the effectiveness of our VR approach, especially in learning fine motor skills and physical manipulation.

Future research should also introduce a longitudinal study design to assess the long-term knowledge retention and lasting effects of VR training. In addition, such a study could allow for a fine-tuning of the hyperparameters of the Presenter Module, which could assess as well its robustness and generalization properties, and an evaluation of the effectiveness of the Recommender Module, which was not investigated in this study. Similarly, the incorporation of new techniques, such as the use of biometric data (Ben Abdessalem and Frasson 2017; Blankendaal and Bosse 2018; Dey et al. 2019) and AI algorithms (Chen et al. 2011; Huang et al. 2018), could improve the optimization of

AL systems in our implementation. Finally, a larger number of subjects could provide insights into the effectiveness of the VR application across different demographics and prior experience levels.

6 Conclusions

This work presents a novel immersive Virtual Reality adaptive application for training emergency physicians in the Difficult Airway Management procedure and related clinical interventions. To the best of our knowledge, this is the first immersive VR application for DAM focused on procedural learning and the first to use an adaptive system in this context based on the Expertise Reversal Effect, which dynamically modulates both the instructor's verbosity and the learning path to the learner's level of expertise.

The experimental results show that the Virtual Reality application is more effective in teaching new knowledge than traditional methods and achieves comparable results in a similar time frame and with a similar cognitive load. However, no significant difference was found between the adaptive and non-adaptive systems, which we attribute to limitations in the experimental design. The evaluation of adaptive systems remains a complex task as there are no standardized and validated frameworks. Developers of adaptive Virtual Reality applications in medical education (and beyond) should consider designing long-term studies or selecting experimental samples with different expertise levels to better assess the impact and benefits of these systems.

Users rated the application very positively in terms of immersion and usefulness, but several areas for improvement were identified. These include localizing the application for non-English speakers, improving interactions with virtual agents, providing better feedback when errors occur, and providing better guidance in the initial stages of the learning experience.

Beyond the specific application for DAM training, this study highlights the broader potential of VR-based adaptive systems to address critical challenges in medical education. By providing scalable, immersive, and customized learning experiences, Virtual Reality technologies can complement traditional methods, reduce logistical and financial burdens, and enable healthcare professionals to develop the cognitive and procedural skills essential for high-risk scenarios. As medical education increasingly relies on technology-enhanced approaches, Virtual Reality has the potential to redefine the teaching of complex skills, bridge the gap between theoretical knowledge and practical application, and ultimately contribute to improved patient safety and better care outcomes.

In future developments, efforts will focus on refining the usability and improving the user experience. We will then evaluate the effectiveness of the application in refreshing expert knowledge and additionally train psychomotor skills with a Human Patient Simulator. We will also conduct longitudinal studies to assess the long-term impact of VR training on knowledge retention and skill acquisition, including a thorough evaluation of the effectiveness of adaptive learning and recommender systems over time.

Acknowledgements None.

Author contributions The concept and design of the study were jointly developed and approved by all authors. The application was designed and developed by EB. EB, FS and AB prepared the data analysis and the first draft of the manuscript. All authors commented on previous versions of the manuscript and read and approved the final manuscript.

Funding Open access funding provided by Politecnico di Torino within the CRUI-CARE Agreement.

Data availability Data discussed in the paper is publicly avalable on figshare at this link: https://doi.org/10.6084/m9.figshare.25983403.v2.

Declarations

Conflict of interest The authors have no relevant financial or non-financial interests to disclose.

Ethical approval The authors state that there are no potential Conflict of interest. Research was conducted involving human participants, with approval by University's Institutional Review Board. Participants were given informed consent prior to experimental procedures.

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

Apfelbaum J, Hagberg C, Caplan R, Blitt C, Connis R, Nickinovich D, Hagberg C, Caplan R, Benumof J, Berry F et al (2013) American society of anesthesiologists task force on management of the difficult airway practice guidelines for management of the difficult airway: an updated report by the American society of anesthesiologists task force on management of the difficult airway. Anesthesiology 118(2):251–270

Rourke J (2006) Manual of rural practice. Can J Rural Med 11(4):295–296

Yang D, Wei Y-K, Xue F-S, Deng X-M, Zhi J (2016) Simulation-based airway management training: application and looking forward. J Anesth 30:284–289

Haerling KA (2018) Cost-utility analysis of virtual and mannequin-based simulation. Sim Healthc 13(1):33–40

Mergen M, Meyerheim M, Graf N (2023) Reviewing the current state of virtual reality integration in medical education-a scoping review protocol. Syst Rev 12(1):1–6

Liu K, Zhang W, Li W, Wang T, Zheng Y (2023) Effectiveness of virtual reality in nursing education: a systematic review and meta-analysis. BMC Med Educ 23(1):710

Carruth DW (2017) Virtual reality for education and workforce training. In: 2017 15th international conference on emerging elearning technologies and applications (ICETA), pp. 1–6. IEEE

Xie B, Liu H, Alghofaili R, Zhang Y, Jiang Y, Lobo FD, Li C, Li W, Huang H, Akdere M et al (2021) A review on virtual reality skill training applications. Front Virtual Reality 2:645153

Zahabi M, Abdul Razak AM (2020) Adaptive virtual reality-based training: a systematic literature review and framework. Virtual Reality 24:725–752

Kelley CR (1969) What is adaptive training? Hum Factors 11(6):547–556

Kaplan A (2021) Higher education at the crossroads of disruption: the university of the 21st century, pp. 121–134. Emerald Group Publishing, Leeds

Radianti J, Majchrzak TA, Fromm J, Wohlgenannt I (2020) A systematic review of immersive virtual reality applications for higher education: design elements, lessons learned, and research agenda. Comput Edu 147

Makransky G, Petersen GB (2021) The cognitive affective model of immersive learning (camil): A theoretical research-based model of learning in immersive virtual reality. Educ Psychol Rev 33(3):937–958

Dabija M, Fedog F, Engström Å, Gustafsson S (2019) Difficult airways: key factors for successful management. J Perianesth Nurs 34(1):151–159

Rosen P, Sloane C, Ban KM, Lanigra M, Wolfe R (2006) Difficult airway management. Intern Emerg Med 1:139–147

Sweller J, Ayres PL, Kalyuga S, Chandler P (2003) The expertise reversal effect. Educ Psychol 38(1):23–31

Kalyuga S (2009) The expertise reversal effect. In: Managing cognitive load in adaptive multimedia learning, pp. 58–80

Sun Y, Pan C, Li T, Gan TJ (2017) Airway management education: simulation based training versus non-simulation based training-a systematic review and meta-analyses. BMC Anesthesiol 17(1):1–7

Morgan PJ, Cleave-Hogg D, McIlroy J, Devitt JH (2002) Simulation technology: a comparison of experiential and visual learning for undergraduate medical students. J Am Soc Anesthesiol 96(1):10–16

Multak N, Euliano T, Gabrielli A, Layon AJ (2002) Human patient simulation: a preliminary report of an innovative training tool for physician assistant education. J Physician Assist Educ 13(2):103–105

Hall RE, Plant JR, Bands CJ, Wall AR, Kang J, Hall CA (2005) Human patient simulation is effective for teaching paramedic students endotracheal intubation. Acad Emerg Med 12(9):850–855

Kory PD, Eisen LA, Adachi M, Ribaudo VA, Rosenthal ME, Mayo PH (2007) Initial airway management skills of senior residents: simulation training compared with traditional training. Chest 132(6):1927–1931

Wenk M, Waurick R, Schotes D, Wenk M, Gerdes C, Van Aken HK, Pöpping DM (2009) Simulation-based medical education is no better than problem-based discussions and induces misjudgment in self-assessment. Adv Health Sci Educ 14:159–171

Hallikainen J, Väisänen O, Randell T, Tarkkila P, Rosenberg PH, Niemi-Murola L (2009) Teaching Anaesthesia induction to medical students: comparison between full-scale simulation and supervised teaching in the operating theatre. Eur J Anaesthesiol| EJA 26(2):101–104

Modell JH, Cantwell S, Hardcastle J, Robertson S, Pablo L (2002) Using the human patient simulator to educate students of veterinary medicine. J Vet Med Educ 29(2):111–116

Demirel D, Butler KL, Halic T, Sankaranarayanan G, Spindler D, Cao C, Petrusa E, Molina M, Jones DB, De S et al (2016) A hierarchical task analysis of cricothyroidotomy procedure for a virtual airway skills trainer simulator. Am J Surg 212(3):475–484

Demirel D, Alexander Y, Halic T, Sankaranarayanan G, Ryason A, Spindler D, Butler KL, Caroline C, Petrusa E, Molina M et al (2016) Virtual airway skills trainer (vast) simulator. Stud Health Technol Inform 220:91

Rajeswaran P, Varghese J, Kumar P, Vozenilek J, Kesavadas T (2019) Airwayvr: Virtual reality trainer for endotracheal intubation. In: 2019 IEEE conference on virtual reality and 3D user interfaces (VR), pp. 1345–1346. IEEE

Rajeswaran P, Kesavadas T, Jani P, Kumar P (2019) Airwayvr: virtual reality trainer for endotracheal intubation-design considerations and challenges. In: 2019 IEEE conference on virtual reality and 3d user interfaces (VR), pp. 1130–1131

Samosorn AB, Gilbert GE, Bauman EB, Khine J, McGonigle D (2020) Teaching airway insertion skills to nursing faculty and students using virtual reality: a pilot study. Clin Simul Nurs 39:18–26

Teasdale G, Jennett B (1974) Assessment of coma and impaired consciousness: a practical scale. The Lancet 304(7872):81–84

Rajeswaran P, Hung N-T, Kesavadas T, Vozenilek J, Kumar P (2018) Airwayvr: learning endotracheal intubation in virtual reality. In: 2018 IEEE conference on virtual reality and 3D user interfaces (VR), pp. 669–670. IEEE

Vaughan N, Gabrys B, Dubey VN (2016) An overview of self-adaptive technologies within virtual reality training. Comput Sci Rev 22:65–87

Ben Abdessalem H, Frasson C (2017) Real-time brain assessment for adaptive virtual reality game: A neurofeedback approach. In: Brain function assessment in learning: first international conference, BFAL 2017, Patras, Greece, Sept 24–25, 2017, Proceedings 1, pp. 133–143. Springer

Dey A, Chatburn A, Billinghurst M (2019) Exploration of an eeg-based cognitively adaptive training system in virtual reality. In: 2019 IEEE Conference on Virtual Reality and 3D User Interfaces (vr), pp. 220–226. IEEE

Blankendaal RA, Bosse T (2018) Using run-time biofeedback during virtual agent-based aggression de-escalation training. In: Advances in practical applications of agents, multi-agent systems, and complexity: the PAAMS collection: 16th international conference, PAAMS 2018, Toledo, Spain, June 20–22, 2018, Proceedings 16, pp. 97–109. Springer

Lang Y, Wei L, Xu F, Zhao Y, Yu L-F (2018) Synthesizing personalized training programs for improving driving habits via virtual reality. In: 2018 IEEE conference on virtual reality and 3D user interfaces (VR), pp. 297–304. IEEE

Jeelani I, Han K, Albert A (2017) Development of immersive personalized training environment for construction workers. In: Computing in civil engineering 2017, pp. 407–415

Lin Y, Wang S (2019) The study and application of adaptive learning method based on virtual reality for engineering education. In: Image and graphics: 10th international conference, ICIG 2019, Beijing, China, Aug 23–25, 2019, Proceedings, Part III 10, pp. 372–383. Springer

Pham T, Roland L, Benson KA, Webster RW, Gallagher AG, Haluck RS (2005) Smart tutor: a pilot study of a novel adaptive simulation environment. Stud Health Technol Inform 111:385–389

Mariani A, Pellegrini E, Enayati N, Kazanzides P, Vidotto M, De Momi E (2018) Design and evaluation of a performance-based adaptive curriculum for robotic surgical training: a pilot study. In: 2018 40th annual international conference of the IEEE engineering in medicine and biology society (EMBC), pp. 2162–2165. IEEE

Yovanoff MA, Chen H-E, Pepley DF, Mirkin KA, Han DC, Moore JZ, Miller SR (2018) Investigating the effect of simulator functional fidelity and personalized feedback on central venous catheterization training. J Surg Educ 75(5):1410–1421

Chan J, Chan C, Chia P, Goy R, Sng B (2024) Novice learners' perspectives on obstetric airway crisis decision-making training using virtual reality simulation. Int J Obstet Anesth 57:103926

Billings D (2012) Efficacy of adaptive feedback strategies in simulation-based training. Mil Psychol 24(2):114–133

Serge SR, Priest HA, Durlach PJ, Johnson CI (2013) The effects of static and adaptive performance feedback in game-based training. Comput Hum Behav 29(3):1150–1158

Dalgarno B, Lee MJ (2010) What are the learning affordances of 3-D virtual environments? Br J Edu Technol 41(1):10–32

Mallampati SR, Gatt SP, Gugino LD, Desai SP, Waraksa B, Freiberger D, Liu PL (1985) A clinical sign to predict difficult tracheal intubation: a prospective study. Can Anaesth Soc J 32:429–434

Johnson CI, Whitmer DE, Entinger J, Peterson EK, Sobel BM (2022) Interacting with virtual reality with a controller instead of the body benefits performance and perceptions. In: Proceedings of the human factors and ergonomics society annual meeting, vol. 66, pp. 1294–1298. SAGE Publications Sage CA: Los Angeles, CA

Monteiro P, Gonçalves G, Coelho H, Melo M, Bessa M (2021) Hands-free interaction in immersive virtual reality: a systematic review. IEEE Trans Visual Comput Graphics 27(5):2702–2713

Strada F, Bottino A, Lamberti F, Mormando G, Ingrassia PL (2019) Holo-blsd-a holographic tool for self-training and self-evaluation of emergency response skills. IEEE Trans Emerg Top Comput 9(3):1581–1595

Bloom CP, Loftin RB (2020) Facilitating the development and use of interactive learning environments. CRC Press, Boca Raton

Nour M, Abed E, Hegazi N (1995) A proposed student model algorithm for an intelligent tutoring system. In: SICE'95. Proceedings of the 34th SICE annual conference. international session papers, pp. 1327–1333. IEEE

Yaghmaie M, Bahreininejad A (2011) A context-aware adaptive learning system using agents. Expert Syst Appl 38(4):3280–3286

Brusilovskiy PL (1994) The construction and application of student models in intelligent tutoring systems. J Comput Syst Sci Int 32(1):70–89

Sakles JC, Laurin EG, Rantapaa AA, Panacek EA (1998) Airway management in the emergency department: a one-year study of 610 tracheal intubations. Ann Emerg Med 31(3):325–332

Brooke J et al (1996) SUS-a quick and dirty usability scale. Usability Eval Indus 189(194):4–7

Kalawsky RS (1999) VRUSE-a computerised diagnostic tool: for usability evaluation of virtual/synthetic environment systems. Appl Ergon 30(1):11–25

Leppink J, Paas F, Vleuten CP, Van Gog T, Van Merriënboer JJ (2013) Development of an instrument for measuring different types of cognitive load. Behav Res Methods 45:1058–1072

Leppink J, Paas F, Van Gog T, Der Vleuten CP, Van Merrienboer JJ (2014) Effects of pairs of problems and examples on task performance and different types of cognitive load. Learn Instr 30:32–42

Biocca F, Harms C, Gregg J (2001) The networked minds measure of social presence: Pilot test of the factor structure and concurrent validity. In: 4th annual international workshop on presence, Philadelphia, PA, pp. 1–9

Bangor A, Kortum PT, Miller JT (2008) An empirical evaluation of the system usability scale. Int J Hum Comput Interact 24(6):574–594

Bangor A, Kortum P, Miller J (2009) Determining what individual SUS scores mean: Adding an adjective rating scale. J Usability Stud 4(3):114–123

Boet S, Borges B, Naik V, Siu L, Riem N, Chandra D, Bould M, Joo H (2011) Complex procedural skills are retained for a minimum of 1 year after a single high-fidelity simulation training session. Br J Anaesth 107(4):533–539

Chen Y, Baran M, Sundaram H, Rikakis T (2011) A low cost, adaptive mixed reality system for home-based stroke rehabilitation. In: 2011 Annual international conference of the IEEE engineering in medicine and biology society, pp. 1827–1830. IEEE

Huang X, Naghdy F, Naghdy G, Du H, Todd C (2018) The combined effects of adaptive control and virtual reality on robot-assisted fine hand motion rehabilitation in chronic stroke patients: a case study. J Stroke Cerebrovasc Dis 27(1):221–228

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