Safety Science 165 (2023) 106200

Contents lists available at ScienceDirect

Safety Science

journal homepage: www.elsevier.com/locate/safety

Improving safety outcomes through medical error reduction via virtual reality-based clinical skills training

Grace A.L. Kennedy [a,*], Shiva Pedram [a], Sal Sanzone [b]

[a] SMART Infrastructure Facility, University of Wollongong, Northfields Avenue, Gwynneville, NSW 2522, Australia
[b] Graduate School of Medicine, University of Wollongong, Northfields Avenue, Gwynneville, NSW 2522, Australia

ARTICLE INFO

Keywords:
Medical error
Human error
Training
Virtual reality
Clinical skills
Medical education

ABSTRACT

The reduction of medical error in clinical procedures is a key factor in improving patient safety and health outcomes. This paper describes an empirical study that compared the human error outcomes between two novice groups of medical students performing Arterial Blood Gas collection; both groups of students were given the same traditional training (bookwork, demonstration and simulated practical), however the study group was provided with an interactive Virtual Reality (VR) practical experience developed by Vantari VR prior to the simulated practical. The results of the study showed that students who had undertaken the VR clinical skills training recorded 40% less errors during a simulated practical than the control group. The contributions of this study are threefold: 1) that VR-based clinical skills training is viable and provides improved outcomes for learners, 2) improved insights into the nature of human error in VR training and 3) prospective and retrospective error analyses are both useful in the iterative design of VR procedural training.

1. Introduction

Mistakes in the field of medicine can result in problematic consequences with serious, non-beneficial, long-term outcomes for patients (Higham and Vincent, 2021). Doctors undergo rigorous education and training to gain skills and knowledge that will meet high standards of professional practice. In order to ensure the safety of patients it is therefore essential to reduce or eliminate the risks due to system conditions that factor into human error. During their education, medical students are exposed to clinical procedures with increasing levels of risk (for example, a student doctor might reasonably expect to understand the theory and practical aspects of a specified procedure before undertaking it in a real-life context). Simulation-based education is one mechanism by which student doctors can increase their practical competency, whilst increasing their confidence without putting patients at unnecessary risk. Deliberate, planned scheduling of clinical skills practice is required for students to improve the retention of clinical skills gained through simulation (Offiah et al., 2019), however existing physical simulation experiences can be expensive to repeat due to the cost of consumables and availability of rooms, equipment and resourcing to set up the simulations, in addition to the scheduling of teachers and students to meet in a pre-arranged physical location. This affords an opportunity for computer-mediated simulation for repeated practice in the consolidation and retention of skills. Improving safety through education via VR training is not a new concept, however there are a lack of pedagogies in designing and developing acceptable learning content (Yang and Goh, 2022, Mehrotra and Markus, 2021), evidence of successful learning outcomes (Huber et al., 2017, Vaughan et al., 2016, Pedram et al., 2020), and the long term goal of improved safety in professional practice (Vaughan et al., 2016).

The research described in this paper is part of a project that explores the wider validation of VR training for clinical skills in medical education. An empirical study was designed and undertaken to test two research questions 1) does VR clinical skills training have the potential to reduce the error rate in practical procedures? and 2) can prospective and retrospective error analysis on traditional clinical training inform VR clinical skills training design? The paper is structured as follows: following the introduction, a background summary of the literature around human error analysis and VR medical training is given. The overall methodology for the study is then provided alongside a description of the study. The fourth section presents the task analysis alongside the results from the prospectively performed error identification and risk assessment for the Arterial Blood Gas (ABG) collection procedure. The fifth section outlines the error results and retrospective

* Corresponding author.
E-mail address: gracek@uow.edu.au (G.A.L. Kennedy).

https://doi.org/10.1016/j.ssci.2023.106200
Received 27 October 2022; Received in revised form 18 April 2023; Accepted 11 May 2023
Available online 18 May 2023

error analysis during both the VR ABG training and the subsequent practical ABG training. The following sections discuss the nature of errors in the different forms of training and outlines potential ways that the prospective and retrospective results could be used to feedback into future training design. The paper concludes with brief dialogue on the limitations of the work and implications for beneficiaries of the research.

2. Human error in VR Medical training

High reliability organisations, such as healthcare institutions, are moving to prioritise error reduction and mitigation of risk to improve patient safety, however there are conflicting definitions of medical error (Grober and Bohnen, 2005, Rodziewicz et al., 2022). Amongst the definitions proposed by Rodziewicz et al. (2022), medical error is defined as:

- The failure to complete the intended plan of action or implementing the wrong plan to achieve an aim.
- An unintended act or one that fails to achieve the intended outcome.
- Deviations from the process of care, which may or may not result in harm.
- When planning or executing a procedure, the act of omission or commission that contributes or may contribute to an unintended consequence.

To further specify *errors in operative technique*, we used the definition presented by Seymour et al. (2002) “as specific events that represent significant deviations from optimal performance, without linking these events to adverse outcomes or proximate causes”. The identification and measurement of error (as previously defined) permits the assessment of the effectiveness of VR training in medical error reduction as it decouples the simulated operative procedure from wider system cause and effect. Rapid development of VR environments is supported by the use/reuse of existing generic physical models of the surgical environment, patient and controller avatar/representation whilst shielding the user from the wider variety of scenarios caused by organisational, technological or other variables that cause the system to potentially shift outside its designed for state. Exploring the reaction to different scenarios or case studies that required alternative decision-making pathways and situation awareness were out of scope for this study.

A major goal of VR training is to improve the competency of the user (through attainment of skills, knowledge and abilities) which in turn will decrease the risk to patient safety attributed to human error. Human error analysis and the attribute of human reliability has been of interest to researchers from the earliest days of accident analysis, cognition and engineering of sociotechnical systems be it for retrospective and/or prospective applications (Kirwan, 1994, Reason, 1990). Reduction in human error, improved safety and reduced risk are often cited as an outcome of VR in medical settings. A review undertaken of the literature between January 2016 and September 2021 sought to understand what is currently known about human error in the context of VR for medical training. The following search terms were used [HMD AND “Virtual Reality” AND (Surgery OR Medical) AND (“Education” OR “Training”)] in Science Direct, Scopus, IEEE Explore, PubMed and Emerald Insight databases.

Operating rooms or ’theatres of errors’ have been studied in a number of settings (Barteit et al., 2021, Mirek and Prétot, 2019) and involve the exploration of a user-controlled avatar in observation mode identifying different types of errors demonstrated within the simulated operating theatre. These types of simulations tend to involve a scenario that has a pre-programmed demonstration of different examples of errors where the user’s avatar can navigate around the room observing and assessing for the presence of errors. These types of simulation train the user in awareness of and identification of violations to safety, hygiene or risk policies and procedures.

Gamification of clinical procedures was also common for training technical and non-technical skills (Breitkreuz et al., 2021b). Although many reported the reduction of errors, it is however unclear how or what types of errors were identified and recorded (Breitkreuz et al., 2021b). Bernardo (2017) also report that VR and other simulations can reduce operative time and error by increasing confidence and wasted movements of the surgeon but do not specify how. Corrêa et al. (2019) measured the 3D spatial errors in trajectory and location (depth) of needle insertion. Huber et al. (2017) recorded error rates but these had mixed effects (some exhibited increases and some decreases in mean error rates) compared between different simulated laparoscopic procedures in immersive VR. The variety in the literature found reinforces the view that there are no clear or common definitions or measures for medical error (Rodziewicz et al., 2022).

Bielsa (2021) reported on the application of the Motor Skills Theory of Learning (Fitts and Posner, 1967) for VR in simulated surgical training; 1) the cognitive phase (learning through demonstration or representation), 2) the associative phase (repetitive training and active feedback on errors) and 3) the autonomous phase (repetitive training and motor skill improvement). Identifying and recording errors are most important at the *associative phase*, once these errors are identified it is essential to also determine how to ensure that users are i) aware they have made an error, ii) understand the implications of that error, and iii) identify prevention or recovery actions to mitigate any risks that arise from that error. As VR creates the opportunity of repetitive training, errors made in subsequent sessions can be identified and feedback provided from the trainer or system (Bielsa, 2021). Feedback information from the training is important to both the trainer and the student so that accurate and timely feedback can be received, this information could be based on motivation, level of interest, errors made, or challenges they faced (Fairen et al., 2020) and can be explicit direction from the trainer, pre-programmed in-simulation guidance, performance measures or simulation of the effects of error (Barteit et al., 2021). Appropriate timescales for feedback were inconsistent, students generally preferred to receive feedback during the training, whilst trainers preferred to provide performance feedback right after the training sessions. It is reported that consistent feedback during the training can reduce the cognitive load and frustration for the learner and allow them to focus on the task (Breitkreuz et al., 2021a) in addition, if the simulated tasks are too complex or mix realism and fantasy this can lead to confusion (Mehrotra and Markus, 2021), conversely, higher levels of complexity tend to improve the retention of skills (Arthur and Day, 2018). While performance feedback is based upon direct observation or data harvested during simulation which allows trainers and trainees to debrief the specifics of the tasks or its sub tasks, there is no clear indication of when the right time is to provide feedback. The form of the feedback is more important, Moreno and Mayer (2002) categorize feedback as: “corrective feedback”, where the trainee will be informed on the correctness of actions or decisions made, and “explanatory feedback”, where the explanation is given at to the strengths and weakness in their performance and why. The result of the latter authors’ analysis indicates that trainees who received explanatory feedback performed better in solving complex problems compared to the group who only received corrective feedback.

Although a key outcome of reduction in medical error is the improvement of patient safety, healthcare providers are also beneficiaries of these initiatives because they are at risk of poor outcomes due to medical error incidents, with adverse outcomes of blame, guilt, feelings of anger and inadequacy, and general lowering of self-confidence and wellbeing (Robertson and Long, 2018, Rodziewicz et al., 2022) in addition to anxiety over potential claims of malpractice or litigation and how that will affect the professional’s long term career.

Whilst this section has described a number of examples and case studies where VR has been used successfully to replicate aspects of VR medical training, it is also important to note that human systems integration (HSI) is essential, meaning that adequate consideration of the human is taken within the design and development of VR training systems throughout their whole system lifecycle. Humans should not be an "after thought" brought into the development once the technological system has been mostly created. Stone (2018) describes the necessity of adopting human factors up front in the development of VR, AR and MxR systems right from the early stages of task analysis and explains where human factors has been used to identify the appropriateness of VR solutions for learning. Abstraction of the system boundary is also a necessary consideration in the wider education and healthcare systems of which the training is part of. Booher (2003) describes 'training' as a key domain of HSI, in particular the integration of training into the design and operation of the system. In the case of VR clinical skills simulation, it can be used early on in the design of the VR system for human-in-the-loop simulation with a view to measuring performance and iterating on the design (improving for the learning outcomes required), and then, once sufficiently mature, abstracted for use at the educational system level as a training mechanism.

3. Material and methods

For the purposes of this study, a simple operative task was chosen that emphasizes technical skill acquisition. The Arterial Blood Gas (ABG) collection procedure was mutually selected by the lecturer (Medical Education and Clinical Skills Academic Leader) and the VR developer as a suitable procedure that would be self-contained, appropriate for novice VR users and synchronised with the scheduling of existing medical school ABG collection training. The lecturer was then asked to provide a task analysis based on the written procedural training materials and to provide a high-level prospective error identification based on their experience in teaching this procedure and observations on typical errors seen (knowledge based on many years of teaching ABG collection). The VR developer was given a checklist of tasks based on the task analysis to inform the design of the VR procedure; however they were not provided with the prospective error identification (so as to maintain independence between the prospective and retrospective error comparisons), the contextual information required to build the simulated training for the ABG scenario was provided by experienced medical practitioners within the VR developer's advisory team.

An empirical study was designed that used mixed data collection via user survey and expert observations of a study group (n=25) and control group (n=19) performing ABG collection via VR practical training and traditional practical training. The study was planned as an intervention-style study rather than comparative as it is not intended that VR training be used as an alternative to the traditional practical training, but as an enhancement. This study was approved in the UOW Ethics Committee Protocol #2021/258. The cohort of participants were Phase 2 (2nd year postgraduate) medical students, where existing traditional ABG training consists of formal written course materials (bookwork), a theory lecture, and an ABG practical undertaken using a physical dummy arm with a simulated hydraulic circulatory system. 100% of the cohort participated in the study, with 25 students volunteering to trial the VR training, and the remaining students consenting to participate in the study as the control group. Both control and study groups participated in the existing training, with the study group also having a VR training intervention between the theory lecture and the ABG practical. During both VR training and practical ABG training the students were observed for deviations from tasks, task duration and completion, and measured on how well they performed, how much support they required from the lecturer/trainer and how confidently they approached the task. In addition to observation, the students completed a survey to ascertain their perceptions on the user experience. Following the practicals, the data gathered was analysed and a retrospective error analysis was performed to test the hypothesis that VR training reduces practical error rate and to stimulate discussion around the nature of human error when comparing VR to physical simulation. Finally, consideration of the usefulness of prospective and retrospective error analyses provided a reflection on the implications of both forms to inform future training design improvements. Figure 1 shows the graphical flow of activities.

The proprietary VR training software was developed by Vantari VR (a commercial software provider specialising in immersive simulation for healthcare) using Unreal Engine and was delivered via Oculus/Meta Quest 2 VR Head Mounted Displays (HMDs) tethered via link cable to gaming laptops. The simulation environment features a synthetic operating room environment with the virtual patient laying in a fixed prone position. A virtual trolley is provided with the required equipment for the procedure that the student can interact with. The student is provided with haptic feedback via strength of pulsing vibrations in the hand controller when identifying the correct location of the virtual patient's pulse. The focus of the training is on both procedural learning, but also the psychomotor skill required to collect an ABG sample. A specifically challenging part of this procedure is ensuring that the needle is inserted in the correct area of the arm (in relation to the pulse location), at the correct angle of incidence, and at the correct 3D depth, therefore these constraints have been replicated within the VR simulation and the task will not complete unless all three constraints are met. Prior VR experience was not required, and students were provided with a short five-minute tutorial within a similar VR operating room environment before undertaking the VR ABG training in order to familiarise them with the hand controller buttons, interaction mechanisms and general appearance of the environment. The training involved teaching the student how to manipulate a selection of medical equipment requiring fine motor skills such as forceps, syringes, valves, swabs and scalpels, some of which would be used in the ABG procedure.

4. Task analysis & error identification (prospective)

An error paradigm exists between training/development of skills and putting those skills into action; making errors during training is a key part of learning as students discover the consequences of the error and make cognitive adjustments for the future, as the student gains in expertise it is reasonable to expect the learning curve to result in reduced errors. Paradoxically, VR training must therefore have enough freedom programmed into the scenario to enable errors to be made and learnt from, however these add to the development costs so it is essential to ascertain which types of errors are the ones which most students will make and have the potential for learning from.

A tabular task analysis (TTA) was undertaken, and from this, prospective error identification and risk levels were ascertained from discussions with subject matter experts (UOW clinical skills lecturers and trainers) in order to uncover and prioritise the errors that could lead to procedural failure. Table 1 shows the combined task and error analyses for the ABG task. In this representation, the TTA is shown (all subtask plans being sequential), the most commonly anticipated error identified, the consequences of this error (and potential recovery routes), alongside the estimated levels of severity and likelihood of these errors occurring.

Using the risk matrix (Table 2), the following tasks in Table 3 were identified as potentially vulnerable to risk (for *risks* ≥ 3):

In practical terms, this quantification helps the trainer understand where to focus efforts in reducing the risk due in part to human error vulnerability. For the purposes of this study, the prospective analysis was kept independent from the VR practical development in order to test the hypothesis on whether a prospective error analysis for a practical training session is appropriate for informing VR training design. The VR practical training was designed by the VR developers based on their existing medical procedure simulations and with advice from their in-house medical advisors (medical practitioners with extensive experience in the ABG collection procedure).

5. Results of retrospective error analysis

During the VR training, two of the project research team acted as observers of the participants. A data collection template based on the TTA was used to record whether the participants completed a task and if

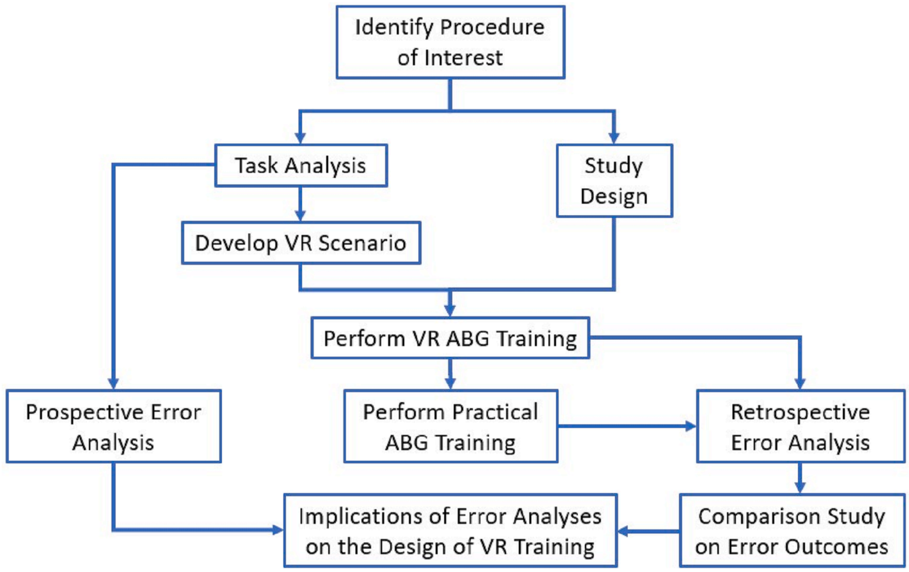

Figure 1. Research methodology.

they made any errors or had difficulties with parts of the procedure. A video of the procedure within the simulated environment was recorded for each participant and reviewed as part of the error identification process.

Four observers were independently recruited from the pool of medical school tutors. Each observer had expertise in the procedure having previously been through the same ABG training during their own studies, and as qualified medical doctors had experience of performing the procedure in hospital settings. Prior to the practical training session the observers were briefed on the task, potential error types they were being asked to perceive and familiarised with the data collection template and overall plan for the training session. The observers were asked to monitor the participant in each task and sub-task noting whether they saw any errors and if so, what happened. Observers were not permitted to interact with the students (i.e. no trainer feedback) during the procedures and were blind to which students had or had not participated in the VR training.

5.1. Error presence percentage

The error presence percentage (represented as % participants who made errors in each subtask) is shown in Table 4, this counts any error (regardless of type or frequency) as a flagged task for each participant and calculates the percentage of students who made any type of error for each task. For example, in Task 4, 24% of participants made some form of error when finding the pulse to identify the location for needle insertion in the VR training, whereas for the same task in the practical training there were very few participants who made some form of error (4.35% in the study group and none in the control group). Error presence in this format is not linked to sub-task completion as all observed errors (deviations from the expected behaviour) were recorded regardless of whether the student met the sub-goal or recovered from the error (i.e. the student made a mistake, but tried again and was successful).

On the most part, the study group's error presence rates were lower across the board than the control group's error presence rates, with the exception of Tasks 4 and 5, however the difference was not significant (e.g. in Task 4, only one student was observed making an error from the study group and no students were observed making an error in the control group). From the remaining tasks, although most showed lower presence of errors in the study group only one met significance level ($p \leq .05$), task '6. Clean skin around puncture site' which shows $M_{study} = 4.35$, $M_{control} = 26.32$, $p = .044$.

As the VR training simulation was in effect a no-fail exercise (and some tasks, 1-3 were autocompleted for the student), there weren't the same opportunities to make errors as in the practical training, therefore it is not appropriate to statistically compare the two different types of training for error presence rate. Further analysis and discussion around the nature of the errors and comparison is given in a later section of the paper.

5.2. Detailed error identification & classification

Each observer made textual notes describing any potential errors that were witnessed as they monitored each subtask being performed by a participant. The error presence percentage gives a high-level estimation of the error rates, however this is limited as the error presence only considers a 1:1 for Task to Error (e.g. if a participant makes errors in more than one subtask then it will only count as one error). Therefore, textual notes and observation forms were revisited to develop a detailed error identification at the sub-task level.

The error rates from observed training may be calculated using the following equations:

$$P(error) = n_e / n_{op}$$

$$Where\ n_{op} = n_T N$$

$$n_e = Number\ of\ errors\ observed,\ n_{op} = Number\ of\ opportunities\ for\ error,$$

$$n_T = Number\ of\ subtasks\ in\ the\ procedure,\ and\ N = Number\ of\ participants$$

Table 5 shows a summary of the calculated error rates for the different forms of training. The discrepancy in n_{op} between the VR, study and control group training is because there were differences in number of participants, and one participant ran out of time during the practical training and any subsequent subtasks from that student were excluded from the study.

Based on the set of participant observations the error rate (shown as a probability) was found to be lower for the study group than the control group ($P(error)_{study} = 0.148 < P(error)_{control} = 0.247$). Therefore, the effect of interventional VR training in this study showed a reduction in overall error rate of 40.02% in the practical training. Similar to the error presence section, it was not reasonable to contrast the VR training error rate to the practicals training due to the nature of the types of errors

Table 1
Tabular task analysis with error identification.

Task	Subtask	Error	Consequences (recovery)	Severity(H/M/L)	Likelihood(H/M/L)
1. Prepares equipment	1.1 Kidney dish				
	1.2 Skin cleaning solution (chlorhex/alcohol)				
	1.3 ABG syringe				
	1.4 Pulls plunger back 1-2mL (may do this later)				
	1.5 23 or 25G needle				
	1.6 Bluey				
	1.7 Gauze	Forget to open gauze packet.	Cannot withdraw needle. (Yes: need help from assistant.)	L	H
	1.8 Small towel				
2. Position patient	2.1 Patient positioned in supine position in bed				
3. Position patient's wrist	3.1 Positions rolled towel under wrist				
	3.2 Ensures wrist is dorsiflexed to 45^o				
4. Identify site of insertion	4.1 Palpates radial pulse at the appropriate anatomical location				
5. Hygiene /safety precautions	5.1 Cleans hands with alcohol handrub	Forget to do.	Risk of infection to patient. (No.)	H	H
	5.2 Dons clean gloves				
6. Clean skin around puncture site	6.1 Cleans skin with chlorhex/ alcohol	Use alcohol-only swab instead.	Risk of infection. (No.)	M	M
	6.2 Allows it to dry				
7. Prepare syringe	7.1 Attach needle to syringe (if not done at 1.5)				
8. Puncture skin	8.1 Locates pulse with finger(s) just proximally to entry				
	8.2 Does <u>not</u> touch skin entry site	Touches previously cleaned puncture site.	Infection. (Yes: clean site again.)	H	M
	8.3 Holds syringe like a dart or pen	Hold it in a different manner.	Less likely to find artery. (Yes: tutor corrects student).	H	H
	8.4 Enters skin at approx. 45^o (30-90^o also acceptable)				
9. Collects blood sample	9.1 Slowly advances until flashback appears	Advances too quickly, or at incorrect angle.	No flashback. (Yes: slowly withdraw while staying beneath skin and change angle.)	M	H
	9.2 Allows syringe to self-fill to 1-2mL				
	9.3 If no success, withdraw and alter angle				
10. Withdraws needle	10.1 Gently withdraws needle				
	10.2 Applies pressure for 3-5 min (may ask assistant)				
11. Finalise sample	11.1 Safely removes needle				
	11.2 Applies black cap to syringe				
	11.3 Expels air from syringe	Cap too loose.	Spray blood everywhere. (No)	H	L
	11.4 Gently roll syringe to mix blood with heparin				

The risk is then quantified against the three levels; high (3), medium (2), low (1) using: $Risk = LikelihoodxSeverity$.

Table 2
Risk matrix

	LOW	MEDIUM	HIGH
LOW	1	2	3
MEDIUM	2	4	6
HIGH	3	6	9

being incomparable, the values are provided for interest, and for potential future benchmarking against other VR tasks or repeated VR training to identify learning curves.

Fig. 2 shows the error rates broken down for each of the subtasks for the observed training modes. This provides a visual indication as to where the errors are most commonly occurring in the procedure. When combined with the severity rating for each subtask, a more complete risk profile can be created and candidate mitigations can be planned (such as more explicit training or feedback).

A visual inspection of the error rate distribution by subtask shows that in general:

$$P(error)_{study} < P(error)_{control}$$

A sign test was also performed to assess whether there were consistent differences between the sets of error rates. The results of the test are shown in Table 6 where:

'positive' is the number of subtasks where $P(error)_{study} < P(error)_{control}$,
'neutral' is the number of subtasks where $P(error)_{study} = P(error)_{control}$,
and 'negative' is the number of subtasks where $P(error)_{study} > P(error)_{control}$.

The test showed that study group's error rate by subtask was significantly lower than that of the control group ($p<.001$).

Table 3
Risk quantification/prioritisation.

Sub-Task	Error	Severity (H/M/L)	Likelihood (H/M/L)	Relative Risk
5.1 Cleans hands with alcohol handrub	Forget to do.	H (3)	H (3)	9
8.3 Holds syringe like a dart or pen	Hold it in a different manner.	H (3)	H (3)	9
8.2 Does not touch skin entry site	Touches previously cleaned puncture site.	H (3)	M (2)	6
9.1 Slowly advances until flashback appears	Advances too quickly, or at incorrect angle.	M (2)	H (3)	6
6.1 Cleans skin with chlorhex/ alcohol	Use alcohol-only swab instead.	M (2)	M (2)	4
1.7 Gauze	Forget to open gauze packet.	L (1)	H (3)	3
11.3 Expels air from syringe	Cap too loose.	H (3)	L (1)	3

Table 4
Error presence percentage (% participants).

Tasks	**VR**	**Study**	**Control**	**Study v Control**
	Error %	**Error %**	**Error %**	***p* (two-tail)**
1. Prepare equipment	-	34.78	52.63	.141
2. Position patient	-	43.48	57.89	.213
3. Positions patient wrist	-	34.78	57.89	.141
4. Identify site of insertion	24.00	4.35	0.00	.370
5. Hygiene/safety precautions	56.00	65.22	63.16	.893
6. Clean skin around puncture site	28.00	4.35	26.32	**.044**
7. Prepare syringe	8.00	4.35	5.26	.893
8. Puncture skin	36.00	60.87	63.16	.883
9. Collects blood sample	20.00	34.78	57.89	.141
10. Withdraws needle	0.00	26.09	27.78	.906
11. Finalises sample	12.00	60.87	77.78	.259

Table 5
Error rates for ABG training.

	VR	**Practical Training**	
		Study	**Control**
n_e	53	102	139
n_{op}	750	690	564
P(error)	**0.071**	**0.148**	**0.247**

6. Error landscapes

As explained earlier in the paper, the nature of the types of errors prevalent in the VR training versus the practical training on the physical dummy arm were not comparable. This has implications for the decision as to the extent to which practical training can be substituted for VR training alternatives. The VR environment has limitations based on which errors have been programmed for, and it is both cost and labour intensive to program every possible deviation from the norm. The study showed that there are types of errors within the VR training that are not probable in real life. Careful consideration must therefore be made as to the replication of errors from real life/physical simulation and VR.

In this section, errors were classified according to the Systematic Human Error Reduction & Prediction Approach (SHERPA) error mode taxonomy (Embrey, 2014). The SHERPA error modes are grouped into 5 classifications.

1. Action errors (concerned with physical actions)
2. Checking errors (concerned with types of checking the status of something)
3. Retrieval errors (concerned with the inability to get the right information at the right time where information has been entered)
4. Communication errors (concerned with the correct and timely communications)
5. Selection errors (concerned with selection from a set of options)

Each of these classifications contain a number of potential error modes (each with an identifier code). Table 7 provides the list of SHERPA error modes.

A landscape of the errors can be viewed by plotting the error modes against the number of errors observed within those modes. Figure 3 shows the error landscapes for the three data sets (VR, study and control groups).

The VR observed errors are focussed around the A4-A7 modes. The most frequent error type was misalignment issues (A5: N=15) which was mostly due to VR 3D positional issues. Within these errors, some were caused by participants attempting to pick up virtual objects but being misaligned (i.e. not being in the right position before pressing the grab button), the other issue participants struggled with was the positioning for both finding the pulse, and locating the puncture site. It was noted that students at the outer centiles of height seemed to have greater problems, the taller students suffered from lack of virtual image depth (often they would not be positioning their controller physically low enough) whereas one shorter student found the top view of the arm was at too oblique an angle. There were also issues with picking up incorrect objects (A6: N=12), this was especially pronounced on the instrument tray. Students either had issues with identifying the correct objects, or accidentally picked up adjacent objects (or even knocked the objects around accidentally). Wrong action on the right object (A7: N=12) was entirely attributed to one sub-task, '5.2 Dons clean gloves'. This subtask was a common challenge for the participants, despite being a relatively simple and innocuous task, the students did not seem to intuitively know how they might put them on within the VR environment. Some tried to place their palms on the gloves and expected them to snap onto their hands, others tried to put the wrong glove on the wrong hand and some couldn't co-ordinate the lifting of the cuff with one hand whilst sliding the glove onto the other hand. Revisiting the error landscape, it should be noted the narrowness of the bands of potential errors within this simulation, many of the errors seen in the practical session were simply not possible to make or observe in the VR training. In addition, a visual task checklist was provided within the VR training environment to prompt the student around the procedural steps, and each subtask had to be completed before the next could be commenced.

The errors observed in the practical training are skewed to omitted subtasks (A8: N_{study}=56, $N_{control}$=77). The vast majority of these omitted tasks focussed on non-collection of specific items (bluey, gauze and towel), skipping of the patient interaction (wrist positioning and pulse identification) in favour of launching straight into the ABG collection off the simulated model arm, and the finalising of the sample (over half of the students did not complete the '11.3 Roll the Sample' subtask). Surprisingly similar numbers of the study and control group omitted the roll the sample subtask, as the VR task required a challenging, but anecdotally intuitive controller joystick action to complete the rolling, it might be expected that the study group would have recalled that subtask. Within the VR training, the students were shown a checklist of the procedure that was highlighted as each sub task was sequentially done so students couldn't omit any tasks as the simulation wouldn't continue to the next subtask until the previous one was completed, this was not the case during practical training. Although the students were provided with a demonstration of the practical training steps immediately before attempting the practical procedure, they were

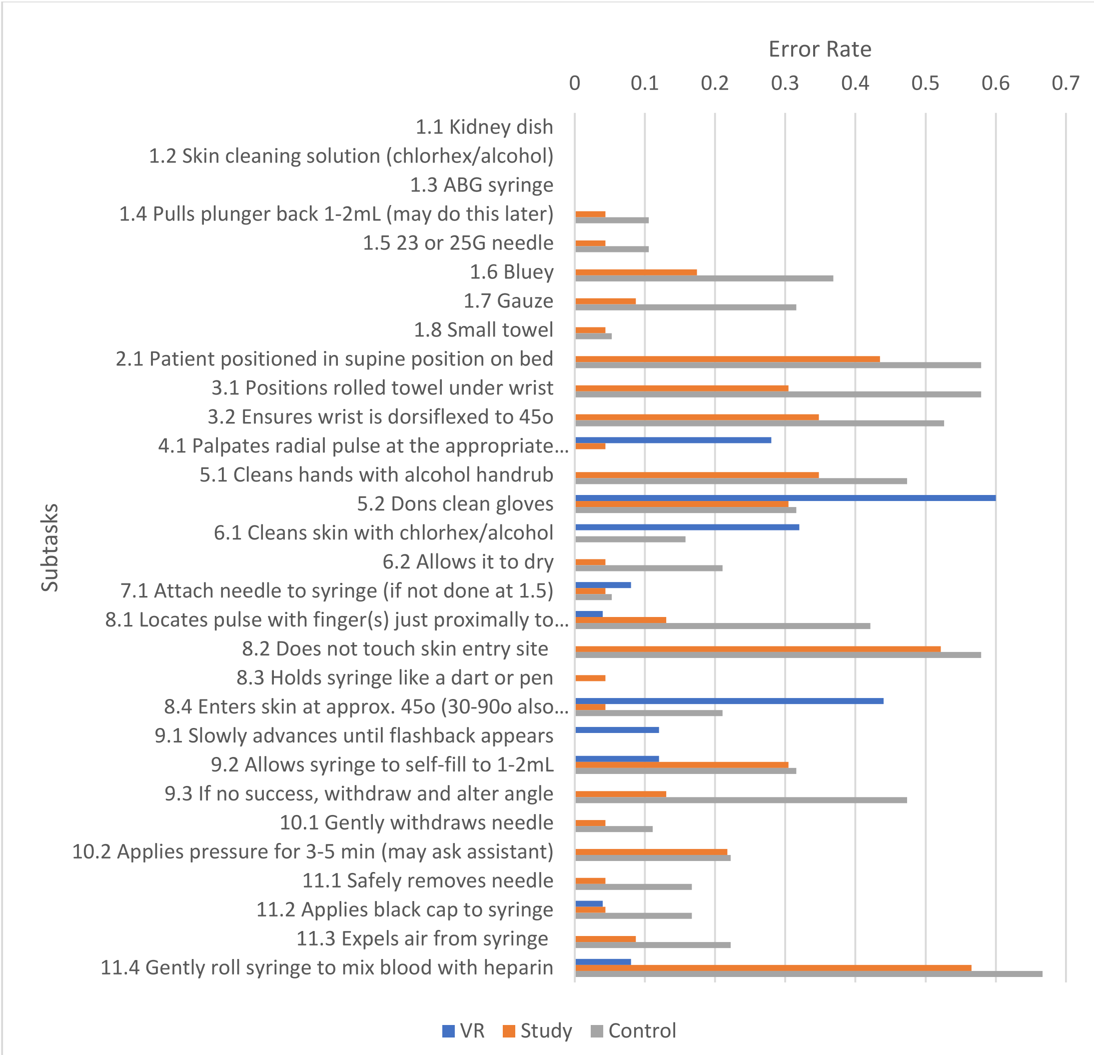

Figure 2. Error rate distribution by subtask.

Table 6
Error rates significance (sign test).

	Positive	Neutral	Negative	p-value
Error Rates per Subtask (difference in error rate)	24	4	2	<.001

not provided with any written task checklists or further prompts unless they specifically requested assistance from the trainer in the room.

Comparing the error landscapes between the study group and the control group, the types of errors and number of errors observed show similar patterns and peaks. Two of the error types showed noticeably reduced numbers of errors. Those students who had participated in VR training made 47.8% fewer 'A4 operation too much/too little' errors and 15.8% fewer 'A8 operation omitted' errors than the control group.

7. Comparison of prospective and retrospective error likelihood

For the purposes of this research, we have translated the error rates into a similar likelihood scale in order to compare the prospective (from lecturer's expert judgement) and retrospective (from calculated observations) as follows: *0-25% Low (L)*, *26-50% Medium (M)*, and *51%-100% High (H)*. Table 8 shows the error likelihood against the subtasks.

Table 7
SHERPA error modes.

Error classification	Code	Error Mode
Action	A1	Operation too long/short
	A2	Operation mistimed
	A3	Operation in wrong direction
	A4	Operation too little/much
	A5	Misalign
	A6	Right operation on wrong object
	A7	Wrong operation on right object
	A8	Operation omitted
	A9	Operation incomplete
	A10	Wrong operation on wrong object
Check	C1	Check omitted
	C2	Check incomplete
	C3	Right check on wrong object
	C4	Wrong check on right object
	C5	Check mistimed
	C6	Wrong check on wrong object
Retrieval	R1	Information not obtained
	R2	Wrong information obtained
	R3	Information retrieval incomplete
Information	I1	Information not communicated
	I2	Wrong information communicated
	I3	Information communication incomplete
Selection	S1	Selection omitted
	S2	Wrong selection made

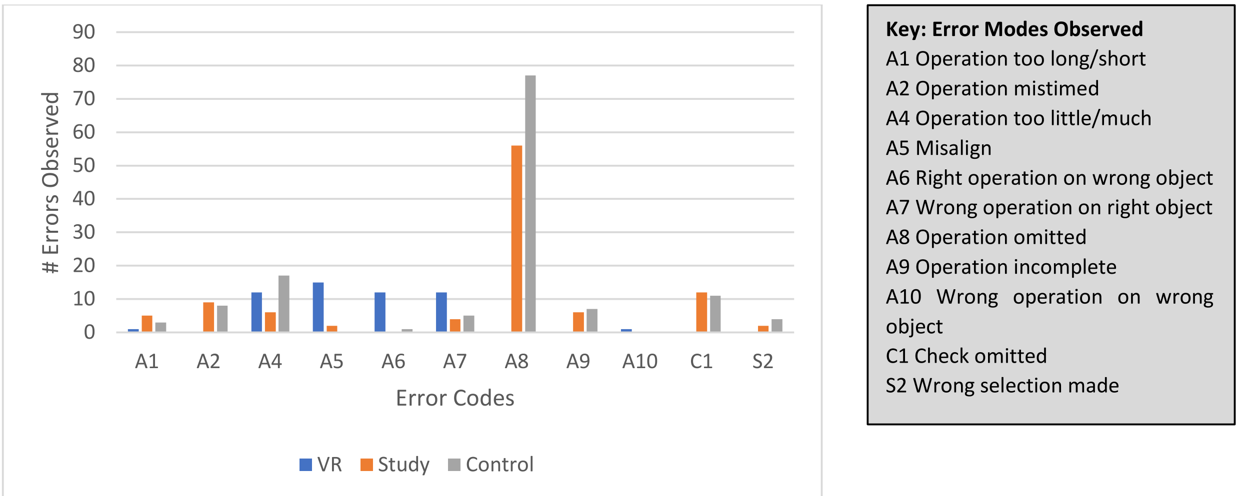

Figure 3. Distribution of observed errors across error modes.

Subtasks 2 and 3 Position patient and position patient's wrist. This subtask was not identified as a common prospective error, however in the practical training errors were seen in relatively high levels. The students seemed to skip the patient interaction tasks, opting to move more quickly to the dummy arm blood collection tasks. Explicit training in this task is required. Although the VR study group showed a lower likelihood of error than the control group, the error likelihood was still Medium level. The VR training did not include any active training for these subtasks (the patient and their wrist were pre-positioned for them in the simulation). There is an opportunity to include these tasks better in the VR training.

Subtask 8.2 Does not touch skin entry site. Reduction of potential infection sources is paramount, infection is attributed in three of the top nine most common medical errors in hospital-acquired conditions according to the Agency for Healthcare Research and Quality (Rodziewicz et al., 2022). This subtask is of interest in the prospective error analysis (Medium) and also likely to be observed in practical training (High in both study and control groups). Interestingly, this was not flagged as an error in the VR training. To improve the VR training, targeted design changes could be used to help the student be aware of where touching of the entry site and potential infection has occurred.

Subtask 8.3 Hold syringe like a dart or pen. This subtask was prospectively identified as a problematic activity (High) however this was not a common observation in either the VR training or practical training.

Subtask 8.4-9.2 Enter skin, advance slowly to allow for flashback and allow syringe to fill. There are some discrepancies between the expected high error probabilities in this series of three sub-tasks. As these actions happen sequentially it is harder to distinguish where the error may have occurred in practice. Our observations showed that the errors were more likely to occur in the angle of attack and not allowing the syringe to fill sufficiently, rather than the lack of pause for flashback. This is confirmed when the prospective error description for 9.1 is revisited "*Advances too quickly, or at incorrect angle*."

Subtask 11.4 Gently roll syringe to mix blood with heparin. This was not identified as a likely source of error in the prospective error identification, however it had a high rate of incidence in the practical training. The error rates for this task were low in the VR training, but unexpectedly high in both the study and control group's practicals. This is indicative that this skill needs better explicit training during the practical.

The comparison between prospective and retrospective error analyses yielded the following reflections:

1. Prospective and retrospective error analyses should be used to inform the replication of types of errors within VR simulated training. This should also include replication of how students identify they have made an error, and plausible routes to error recovery.
2. The use of prospective error analysis of practical training alone to inform VR task design is not appropriate due to the wide discrepancies between prospective and retrospective analyses. For example, the expected high probability of error with task 8.3 'hold syringe like a dart or pen' did not manifest in the practical training (in either study or control groups). Finger posture and grabbing controls are limited by the functionality of VR hand controllers, if the VR development had focused on developing a representation of finger posture and grabbing angle freedom into the scenario it would have been wasted development cost.
3. Whilst prospective error analysis can help to prioritize the higher risk tasks, discussion should still be made over cost trade-offs/value of enacting that type of error in VR versus other training mechanisms within the wider educational provisions that could yield similar learning outcomes.
4. VR training error analysis is a mixture of VR-specific problems and task problems. In further design iterations it will be important to mitigate VR-specific problems. Some of these are due to lack of familiarity with VR controls (pressing incorrect button, or 3D spatial errors) which would improve with use. Some are due to specific idiosyncrasies in the VR design such as the difficulty in intuitively knowing the edge from which to pick up the gloves and how to put them on (in real life the student is unlikely to not know how to don gloves, but were more likely to either forget to don the gloves or put them on too early in the procedure).
5. VR training provides additional procedural support than practical training so retrospective error analysis between VR and practical training is not always comparable. For example, VR training provided a task checklist and students could not omit any tasks, whereas in practical training there was no task list and students were able to omit tasks. In a like-for-like comparison, this is a limitation, however from a wider educational perspective these additional supports for novice learners through VR are beneficial to overall learning outcomes. It is not intended for VR training to entirely replace clinical skills training, but as a means for low-risk practice, increased exposure to procedures, faster acquisition and retention of skill.
6. Prospective and retrospective error rates are not necessarily useful without the added context of the types of task and detailed error analysis around the types of error to be anticipated.
7. Prospective and retrospective error analyses together can be used to inform changes in teaching for improved learning outcomes in terms

Table 8
Error likelihood comparison.

Subtask	Error Likelihood			
	Prospective	Retrospective		
		VR	Study	Control
1.1 Kidney dish		0	0	0
1.2 Skin cleaning solution (chlorhex/alcohol)		0	0	0
1.3 ABG syringe		0	0	0
1.4 Pulls plunger back 1-2mL (may do this later)		0	0.04	0.11
1.5 23 or 25G needle		0	0.04	0.11
1.6 Bluey		0	0.17	0.37 (M)
1.7 Gauze	H	0	0.09	0.32 (M)
1.8 Small towel		0	0.04	0.05
2.1 Patient positioned in supine position on bed		0	0.43 (M)	0.58 (H)
3.1 Positions rolled towel under wrist		0	0.30 (M)	0.58 (H)
3.2 Ensures wrist is dorsiflexed to 45°		0	0.35 (M)	0.53 (H)
4.1 Palpates radial pulse at the appropriate anatomical location		0.28 (M)	0.04	0
5.1 Cleans hands with alcohol handrub	H	0	0.35 (M)	0.47 (M)
5.2 Dons clean gloves		0.60 (H)	0.30 (M)	0.32 (M)
6.1 Cleans skin with chlorhex/ alcohol	M	0.32 (M)	0	0.16
6.2 Allows it to dry		0	0.04	0.21
7.1 Attach needle to syringe (if not done at 1.5)		0.08	0.04	0.05
8.1 Locates pulse with finger(s) just proximally to entry		0.04	0.13	0.42 (M)
8.2 Does <u>not</u> touch skin entry site	M	0	0.52 (H)	0.58 (H)
8.3 Holds syringe like a dart or pen	H	0	0.04	0
8.4 Enters skin at approx. 45° (30-90° also acceptable)		0.44 (M)	0.04	0.21
9.1 Slowly advances until flashback appears	H	0.12	0	0
9.2 Allows syringe to self-fill to 1-2mL		0.12	0.30 (M)	0.32 (M)
9.3 If no success, withdraw and alter angle		0	0.13	0.47 (M)
10.1 Gently withdraws needle		0	0.04	0.11
10.2 Applies pressure for 3-5 min (may ask assistant)		0	0.22	0.22
11.1 Safely removes needle		0	0.04	0.17
11.2 Applies black cap to syringe		0.04	0.04	0.17
11.3 Expels air from syringe	L	0	0.08	0.22
11.4 Gently roll syringe to mix blood with heparin		0.08	0.56 (H)	0.67 (H)

of explicit direction/written materials and both VR training and practical training.

In the training of clinical skills it is just as important to learn the theoretical skills through correct actions as it is to experience failures (and their effects) and learn to either avoid these actions again, or learn how to recover from or make corrective actions. In terms of error probabilities, the novice user would be expected to be making high numbers of errors (error tolerance for absolute beginners can be expected to be as high as 30%). It will be essential to consider which errors can be mitigated through design, or processes, and which errors need to be mitigated with improved training (or a combination of both if the severity warrants such measures).

8. Limitations & further work

The study has five identified limitations. Firstly, the study reported within this paper is a single study with a single cohort of medical students (N=44). Repeated studies with a larger sample are required to validate the results. Secondly, the time horizon covered one training experience, further investigation is required to understand the learning curve under further "doses" of VR training, skills retention and decay, and error reduction over time and into professional practice. Clinical skill includes development of the psychomotor skills as well as the cognitive skills (involving both declarative and procedural knowledge (Cecilio-Fernandes et al., 2018)), which are subject to skill decay over time, and are more likely to decay where procedures have been practiced less (Offiah et al., 2019), from the obverse perspective, skill retention is greater with practice and is needed to gain mastery or expertise of the task (Arthur and Day, 2018). Thirdly, the variety of the population is also a potential concern; the cohort of students were mostly in their 20s to mid-30s. The cohort did not contain any students for which individual's conditions that might impinge on the experience of using VR (e.g. eye conditions that would affect stereoscopic vision or motion-stability issues). The fourth limitation is that of broader applicability, the ABG collection task and teaching procedures were developed from one university's perspective, other institutions may perform and teach the procedure in different ways. Finally, the observations team (two researchers familiar with the VR ABG collection task and four medical tutor assessors familiar with the ABG practical task) were briefed together in the observations procedure, however for the subjective judgement on performance, no benchmarking occurred between observers post-assessment.

Both forms of simulated training (VR and practical) represented a constrained sub-task in an ideal context with limited trainer feedback and considered students at the novice stage of learning. It is anticipated that repetitive practice would happen on the VR training system in the future. Further investigations should consider not just further practice, but also the intervals between practices. Limited studies in surgical skill retention have shown that training in intervals outperform massed practiced, however it is unclear what intervals are appropriate for different horizons of skill retention (Cecilio-Fernandes et al., 2018). A further line of consideration is the development of adaptive VR training that can provide feedback to the student to meet their needs (for example, if they are struggling, or make an error, the system could flag this to the user and provide corrective and explanatory feedback), adaptive scenarios could also be used for the advanced user to practice on unusual patient conditions or more complex environments. Another key improvement would be to consider supporting the students' associative phase of learning by programming more realistic cause and effects during error events, particularly where the risk is higher, such as not having the gauze available to apply pressure after needle withdrawal could result in arterial spray.

9. Conclusions

This paper has focused on the link between VR training interventions and human error with the intended outcome of improved patient safety. The study has three main conclusions:

1) VR-based clinical skills training is viable and provides improved outcomes for learners. The results of this evidence-based study showed that VR training resulted in significant reduction in error rate across the set of subtasks and a reduced incidence of human error across the whole task by 40% in the ABG collection procedure which has an impact on not only the patient's safety, but the proficiency of the learner and ultimately the wider healthcare systems that the future doctors will be part of.

2) Improved insights into the nature of human error in VR training. The authors' intention for the paper is to report not only on the impact of VR training on medical error reduction, but also to further the debate on

the use of error analysis in VR. The paper gives a clear and comprehensive report of a methodology for error identification and analysis that can be used by other researchers and teachers in the field of VR for medical training to better understand the effect of VR training on student performance. The results could also be used to assist in benchmarking medical error and more widely human reliability in VR training for different domains.

3) Prospective and retrospective error analyses are both useful in the iterative design of VR-based clinical skills training. Whilst existing studies have tended to focus on feasibility and user experience, it is equally important to understand the nature of the types of error, how they are translated from real-life to VR and vice versa and finally how those who develop VR learning systems should use such error analyses to derive requirements for the design of the VR training. By doing this, the field will be able to more closely show traceability between errors, feedback, and associative learning that will support the learning outcomes that are desired. Co-development of VR medical procedure training with the stakeholders (learner, teacher and VR developer) are key to the development and improvement of training that will be fit for purpose.

CRediT authorship contribution statement

Grace A.L. Kennedy: Conceptualization, Methodology, Formal analysis, Investigation, Writing – original draft, Writing – review & editing, Visualization. **Shiva Pedram:** Conceptualization, Investigation, Writing – review, Funding acquisition. **Sal Sanzone:** Conceptualization, Resources, Validation.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

This project was funded by Vantari VR (NSW, Australia) and UOW Advantage SME (Wollongong, Australia) Tech Voucher program, as part of NSW Treasury Boosting Business Innovation Program. The authors would like to thank the personnel of Vantari VR and the staff and students of the UOW Graduate School of Medicine for their guidance and support.

References

Arthur, W., and Day, E.A., 2018. Skill decay: The science and practice of mitigating skill loss and enhancing retention. In: Ward, P. (ed.) The Oxford Handbook of Expertise, Oxford Library of Psychology, pp. 1085-1108.

Barteit, S., Lanfermann, L., Bärnighausen, T., Neuhann, F., Beiersmann, C., 2021. Augmented, mixed, and virtual reality-based head-mounted devices for medical education: systematic review. JMIR Serious Games 9.

Bernardo, 2017. Virtual reality and simulation in neurosurgical training. World Neurosurgery 106, 1015–1029.

Bielsa, V.F., 2021. Virtual reality simulation in plastic surgery training. Literature review. J. Plast. Reconstruct. Aesthet. Surg. 74, 2372–2378.

Booher, H.R., 2003. Handbook of human systems integration, John Wiley & Sons.

Breitkreuz, K., Kardong-Edgren, S., Gilbert, G.E., Anderson, P., Deblieck, C., Maske, M., Hallock, C., Lanzara, S., Parrish, K., Rossler, K., Turkelson, C., Ellertson, A., 2021. Nursing faculty perceptions of a virtual reality catheter insertion game: a multisite international study. Clin. Simulat. Nurs. 53, 49–58.

Cecilio-Fernandes, D., Cnossen, F., Jaarsma, D.A., Tio, R.A., 2018. Avoiding surgical skill decay: a systematic review on the spacing of training sessions. J. Surg. Educat. 75, 471–480.

Corrêa, C.G., Nunes, F.L.S., Ranzini, E., Nakamura, R., Tori, R., 2019. Haptic interaction for needle insertion training in medical applications: The state-of-the-art. Med. Eng. Phys. 63, 6–25.

Embrey, D., 2014. Application of SHERPA to predict and prevent use error in medical devices. Proceedings of the International Symposium on Human Factors and Ergonomics in Health Care, 2014. SAGE Publications Sage CA: Los Angeles, CA, pp. 246-253.

Fairen, M., Moyes, J., Insa, E., 2020. VR4Health: Personalized teaching and learning anatomy using VR. J. Med. Syst. 44.

Fitts, P.M., Posner, M.I., 1967. Human performance. brooks. Cole, Belmont, CA, 5, 7-16.

Grober, E.D., Bohnen, J.M., 2005. Defining medical error. Can. J. Surg. 48, 39.

Higham, H., Vincent, C., 2021. Human Error and Patient Safety. Textbook of Patient Safety and Clinical Risk Management, Springer, Cham.

Huber, T., Paschold, M., Hansen, C., Wunderling, T., Lang, H., Kneist, W., 2017. New dimensions in surgical training: immersive virtual reality laparoscopic simulation exhilarates surgical staff. Surg Endosc 31, 4472–4477.

Kirwan, B., 1994. A guide to practical human reliability assessment. CRC press.

Mehrotra, D., Markus, A.F., 2021. Emerging simulation technologies in global craniofacial surgical training. J. Oral Biol. Craniofacial Res. 11, 486–499.

Mirek, S., Prétot, R., 2019. Le "bloc des erreurs" en simulation, apprendre autrement. Inter bloc 6502, 28–31.

Moreno, R., Mayer, R.E., 2002. Learning science in virtual reality multimedia environments: Role of methods and media. J. Educat. Psychol. 94, 598–610.

Offiah, G., Ekpotu, L.P., Murphy, S., Kane, D., Gordon, A., O'Sullivan, M., Sharifuddin, S. F., Hill, A.D., Condron, C.M., 2019. Evaluation of medical student retention of clinical skills following simulation training. BMC Med. Educat. 19, 1–7.

Pedram, S., Palmisano, S., Skarbez, R., Perez, P., Farrelly, M., 2020. Investigating the process of mine rescuers' safety training with immersive virtual reality: a structural equation modelling approach. Comput. Educat. 153, 103891.

Reason, J., 1990. Human error, Cambridge University Press.

Robertson, J.J., Long, B., 2018. Suffering in silence: medical error and its impact on health care providers. J. Emergency Med. 54, 402–409.

Rodziewicz, T.L., Houseman, B., Hipskind, J.E., 2022. Medical error reduction and prevention. StatPearls [Internet].

Seymour, N.E., Gallagher, A.G., Roman, S.A., O'Brien, M.K., Bansal, V.K., Andersen, D.K., Satava, R.M., 2002. Virtual reality training improves operating room performance: results of a randomized, double-blinded study. Ann. Surg. 236, 458.

Stone, R.J., 2018. Blending the Best of the Real with the best of the Virtual: Mixed Reality Case Studies in Healthcare and Defence. Springer, Augmented Reality and Virtual Reality.

Vaughan, N., Dubey, V.N., Wainwright, T.W., Middleton, R.G., 2016. A review of virtual reality based training simulators for orthopaedic surgery. Med. Eng. Phys. 38, 59–71.

Yang, F., Goh, Y.M., 2022. VR and MR technology for safety management education: an authentic learning approach. Saf. Sci. 148, 105645.