



HAL
open science

Training situational awareness for scrub nurses: Error recognition in a virtual operating room

Marie-Stéphanie Bracq, Estelle Michinov, Marie Le Duff, Bruno Arnaldi, Valérie Gouranton, Pierre Jannin

► To cite this version:

Marie-Stéphanie Bracq, Estelle Michinov, Marie Le Duff, Bruno Arnaldi, Valérie Gouranton, et al.. Training situational awareness for scrub nurses: Error recognition in a virtual operating room. *Nurse Education in Practice*, 2021, 53, pp.1-10. 10.1016/j.nepr.2021.103056 . hal-03216504

HAL Id: hal-03216504

<https://hal.science/hal-03216504>

Submitted on 5 May 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

TRAINING SITUATIONAL AWARENESS FOR SCRUB NURSES: ERROR RECOGNITION IN A VIRTUAL OPERATING ROOM

Marie-Stéphanie Bracq^{1,2}, Estelle Michinov¹, Marie Le Duff², Bruno Arnaldi³, Valérie
Gouranton³, Pierre Jannin²

1. Univ Rennes, LP3C (EA 1285), F-35000 Rennes, France. mariestephrem@gmail.com;
estelle.michinov@univ-rennes2.fr
2. Univ Rennes, Inserm, LTSI - UMR 1099, F-35000 Rennes, France. marie.le-
duff@univ-rennes1.fr; pierre.jannin@univ-rennes1.fr
3. Univ Rennes, INSA Rennes, Inria, CNRS, IRISA, F-35000 Rennes, France.
Bruno.Arnaldi@irisa.fr; Valerie.Gouranton@irisa.fr

Manuscript type: Research paper

Word count: 5889 words (including the abstract, text and references)

Funding

This work has received a French government support granted to the CominLabs excellence laboratory and managed by the National Research Agency in the "Investing for the Future" program under reference ANR-10-LABX-07-01.

Conflict of Interest

The authors declare that they have no conflict of interest.

Please address any correspondence regarding this manuscript to: Estelle Michinov, Université Rennes 2, Département de Psychologie, Place du Recteur Henri le Moal, 35043 Rennes Cedex (France). Phone Number: + 33 2 99 14 19 44

E-mail: estelle.michinov@univ-rennes2.fr

Acknowledgments

The authors are particularly grateful to the engineers for their technical assistance and patience (A. Audinot), to P.L. Hénaux and B. Nogues from the neurosurgery department of the CHU Rennes for her helpful comments, and to C. Gautier, head of the Pôle de Formation des Professionnels de Santé – CHU Rennes for their support.

Abstract

Virtual reality simulation provides interesting opportunities to train nurses in a safe environment. While the virtual operating room has proven to be a useful training tool for technical skills, it has been less studied for non-technical skills. This study aimed to assess “Error recognition in a virtual operating room”, using a simulation scenario designed to improve situation awareness. Eighteen scrub-nurse students and 8 expert scrub-nurses took part in the experiment. They were immersed in a virtual operating room and reported any errors they observed. There were nineteen errors with various degrees of severity. Measures were retrieved from logs (number of errors, time for detection, movements) and from questionnaires (situation awareness, subjective workload, anxiety and user experience). The results showed that the participants who detected most errors had a higher level of situation awareness, detected high-risk errors faster and felt more immersed in the virtual operating room than those detecting fewer errors. They also felt the workload was lighter and

experienced more satisfaction. Students explored the operating room more than experts did and detected more errors, especially those with moderate risk. Debriefings confirmed that virtual simulation is acceptable to trainees and motivates them. It also provides useful and original material for debriefings.

Keywords: non-technical skills, situation awareness, virtual reality simulation, scrub nurses, education

Introduction

Most scrub nurses discover the operating room (OR) for the first time during their nursing internships, by watching surgical interventions, under the supervision of experienced scrub nurses, surgeons and surgical residents. This clinical training is stressful for students and very demanding in terms of tutor availability. It also leads to differences in learning outcomes due to the diversity of practices, levels and pedagogical qualities of the tutors. Virtual reality (VR) simulation technology provides interesting opportunities to train scrub nurses in a safe environment and is less demanding in tutor resources (Flurry et al., 2012). Although VR simulation has been shown to be a useful training tool for psychomotor or procedural skills, it has been less widely studied for non-technical skills (NTS) training (Author blinded, 2019a). However, NTS are crucial for patient safety and quality of care and training scrub nurses in these skills has become essential (Kang et al., 2015).

Among NTS, situation awareness (SA) appears crucial in the OR as it has a direct impact on other NTS such as communication, decision making, leadership and teamwork (Flin et al., 2015; Gluyas et al., 2019; Stubbings et al., 2012). SA is defined as the ability to gather information from the environment, recognize and understand it and anticipate its future

state (Endsley, 1995; Flin et al., 2008). SA is a major NTS for scrub nurses who are responsible for hygiene and safety and who are also expected to “think ahead of the surgeon” (Mitchell et al., 2012, 2011). Many factors can have a negative influence on SA (Endsley, 2012; Gluyas et al., 2019): environmental factors (e.g. noise, equipment, time pressure), individual factors (e.g. anxiety, fatigue, stress) and cognitive factors (e.g. attentional capacity, information overload, task interruptions). Several studies have demonstrated that specific training is required to improve SA (Endsley, 2015; Gluyas et al., 2019; Salas et al., 2008). Simulation can improve SA and facilitate clinician workflows (Brady et al., 2013; Stubbings et al., 2012). The WHO developed a set of tools to help health professionals use simulation technology to mimic professional situations for training purposes (WHO, 2017).

VR simulation technology enables the development of low-cost, realistic, easy-to-use, easily configurable simulators that reduce safety, ethical and health problems. VR simulation can be used to assist trainees acquire skills in a safe environment with low levels of anxiety (e.g., Kardong-Edgren et al., 2019). Setting up a simulation of the operating room or using a real operating theatre is expensive and unrealistic, so VR simulation technology is attractive. Since the development of head-mounted displays (HMDs) and other accessible VR systems, it has become easier to generate virtual environments. In medical education, the 3D Virtual Operating Room has been used as a safe setting to train healthcare professionals (e.g., Christensen et al., 2018; Gallagher et al., 2005; Pons Lelardeux et al., 2017). The usability and acceptability of virtual OR in nursing education has been tested and validated with expert and student nurses (Author blinded, 2019b). Some recent results suggest that VR simulation technology that promotes a high “sense of presence” improves the user’s subsequent performance (Grassini et al., 2020). VR simulation offers advantages compared with traditional training methods and can improve performance (Alhalabi, 2016; Webster, 2016),

although these findings have recently been challenged (Makransky et al., 2019, 2020). There is thus a need to examine the effectiveness of VR simulation technology in nursing education (Author blinded, 2019a; Chen et al., 2020; Kyaw et al., 2019).

In the present study, the objective was to develop a simulation in a VR operating room we called “Error recognition in a virtual OR” and examine its value as a SA training tool for scrub nurses, as it involves effective information gathering and anticipation during the surgical intervention. Based on previous studies related to factors influencing SA (e.g. Endsley, 2012, 2015; Flin et al., 2008; Gluyas et al., 2019), we expected that participants with higher SA, lower perceived workload, lower anxiety and higher immersion levels in the VR simulator would detect more errors in the OR and more rapidly. We also expected that participants who detected more errors would explore the virtual OR more thoroughly and have a better user experience.

Methods

Participants

The simulation scenario was pretested by three second-year students in a French school for scrub nurses. This gave us an overview of the reactions to the scenario and an estimate of the number of reported errors. It also allowed us to adjust the duration of the simulation.

For the virtual OR simulation session, 18 scrub nurse students (15 females, 3 males, $M_{age} = 35$ years) in the same class participated voluntarily. Eight expert scrub nurses (all females, $M_{age} = 45$ years) working in a University hospital also participated. The experiment took place in the local nursing school. The sessions were led by a doctoral student in psycho-

ergonomics and two computer engineers. The study took place over eight half-days between May and August 2019.

Ethical approval

The study protocol was approved by the Ethics Committee of the University Hospital of Rennes, France (No. 16.137). The student nurses were invited to participate in the VR simulation at the beginning of their semester and the expert nurses were informed of the project during a training session in the nursing school. The project and the information concerning a VR simulation was presented to students and nurses by a doctoral student in Psychology and Ergonomics and all the participants provided their written consent. They were informed that their responses would be processed anonymously and used for research purposes only. No personal information was collected.

Materials and procedure

The simulation session comprised the three classic steps of healthcare simulation: 1) briefing, 2) VR simulation and 3) collective debriefing, in accordance with the International Nursing Association for Clinical Simulation and Learning (INACSL).

Briefing. Before entering the VR simulator, participants watched a short VR scenario showing how to move around in the virtual OR. They were then briefed about the aim of the simulation session and told that their task was to check the instrument table and ensure that quality standards, safety and hygiene rules were respected. They were asked to report any surgical error they observed by indicating the non-conforming object or situation. Before entering the simulator, they read a paper version of the patient's file, which was also

accessible in the virtual environment. The case was a craniotomy for Mr Jean Dupond, aged 54, with a left frontal meningioma.

VR simulation equipment. Each participant was equipped with a virtual reality headset. The equipment used for the experiment was an HTC Vive system, composed of a Head Mounted Display and two hand controllers. A large screen enabled the experimenters to monitor the progress of the scenarios. The virtual OR was developed by computer science researchers (author blind, 2019b), based on #FIVE (Bouville et al., 2015) and #SEVEN (Claude et al., 2016) tools that fit into UNITY (<http://unity.com>). The scenario is obtained from real observations processed by a calculation technique on Test-and-Flip networks (Caillaud, 2013). The integration in Unity allows rapid development (single interface) and multi-platform deployment (screen, headsets, tablets) of the applications.

In the “Virtual Operating Room of errors”, participants are immersed in a virtual OR with no guidance. They can move and interact freely in the environment between five categories of location: equipment, colleagues, instrument tables, patient and tactile screen (see Figure 1). The virtual OR contains 121 medically related objects and three avatars representing colleagues (surgeon and scrub nurse) and the patient.

Task in the VR simulation. In the “Virtual Operating Room of Errors”, scrub nurses had to make sure that quality standards, safety and hygiene rules were respected and report any surgical error they observed. The errors were selected after discussion with expert scrub nurses and a review of the literature on existing tools. “Error recognition in an OR” is a well-known simulation scenario in France, involving real-life situations in a mock-up OR aimed at raising healthcare professionals’ awareness of quality and safety standards and hygiene rules (Guilloton, 2016; Mirek & Prétot, 2019). We also took into account the representability of

errors in a virtual OR environment after discussion with VR engineers and we asked scrub nurse teachers to validate their pedagogical interest. Finally, 19 errors were introduced, related to hygiene and risk of infection ($n = 12$), disruptions to the surgical procedure ($n = 4$), identity monitoring ($n = 1$), trophic lesions ($n = 1$) and medical risk ($n = 1$). For each one, the level of risk was evaluated by two teachers and two scrub nurses and we differentiated between moderate- ($n = 12$) and high-risk errors ($n = 7$). The VR simulation session lasted a total of 14 minutes.

Debriefing. Each session was followed by a short individual debriefing (5 to 10 minutes) to obtain participants' immediate reactions and their perception of self-efficacy (Bandura, 1977). Next, a 90-minute collective debriefing was organized with two teachers and experimenters, following the 3D model: Defusing, Deepening and Discovering (Zigmont et al., 2011). First, participants were asked to express their feelings and reactions after the simulation (Defusing phase). Next, all the errors were reviewed (Discovering phase). Finally, participants were asked about their strategies and what they had learnt for their future clinical practice (Deepening phase). At the request of the nursing students during the collective debriefing, a second session was organized six months later with the same protocol and scenario.

Measures

Detected errors, time for detection and trajectory patterns in the virtual OR. The simulator recorded objective data for each participant: number and nature of detected errors, time needed for detection, number and nature of places visited. Retrieved logs from each session provided graphical movement representations enabling us to compare trajectory patterns (Gallagher & Satava, 2002). The OR was divided into four zones: entrance with

screen and tables; centre with patient, equipment and colleagues; corner 1 with other equipment; and corner 2 with microscope (see Figure 3).

Situation awareness. The Situation Awareness Rating System (SART; Taylor, 1990) was used to assess participants' SA during the simulation session. It has 10 items rated on a seven-point Likert scale, grouped in three dimensions: attentional demand, attentional supply and understanding. Scores vary from -15 to +39.

Subjective workload. The NASA Task Load Index (NASA-TLX; Hart and Staveland, 1988; simplified version of Byers et al., 1989) was used to assess participants' subjective workload during the simulation session. It has six items: mental demand, physical demand, temporal demand, performance demand, effort and frustration, rated on a 0–100 scale.

Anxiety. The short form of the State-Trait Anxiety Inventory (STAI-Y; Spielberger, 1983; STAI: Y-6 item; Marteau & Bekker, 1992) was used to assess participants' anxiety. It has six items rated on a four-point Likert scale, from 1= not at all to 4= very much. Total score is calculated using Marteau and Bekker's formula. Scores range between 34 and 36 under normal conditions and between 50 and 61 under stressful conditions (Bekker et al., 2003).

Debriefing analysis. Debriefings were recorded and transcribed. Collective debriefing was analysed using a five-minute timeline pattern, summarizing the number of oral interactions by students, silences, interruptions such as laughter and their arguments and affect.

Attitudes and satisfaction post-simulation. Participants were asked to assess their perception of the simulation with 13 items rated on a seven-point Likert scale and grouped in five dimensions: ease of use, immersion, efficiency, overall satisfaction and behavioural intention.

Data analysis

Statistical analysis was performed with JASP (JASP Team, 2018). Descriptive statistics were computed for the level of performance and for the participants' status (nursing students vs expert nurses). For level of performance, participants were median-split to create two groups: participants in the Low Detection Group (LDG) detected fewer than 7.5 errors, while those in the High Detection Group (HDG) detected more than 7.5. We decided to use a median split (variable-oriented) approach rather than a cluster analysis (person-oriented) as our objective was not to examine individual differences, but rather to split the sample into two groups based on performance (low versus high detection of errors). A dichotomized measure represents groups more appropriately and enables analyses to be conducted in terms of group rather than individual differences (MacCallum et al., 2002). Means were compared using the Wilcoxon test for paired samples and Mann-Whitney test for independent samples. A qualitative analysis was conducted to identify the pattern of movements in the virtual OR and the collective debriefing sessions were examined through thematic content analysis.

Results

Effects of the level of performance

The results for the two groups are presented in Table 1. HDG detected more errors than LDG in both categories: moderate risk ($U = 7.50, p < .001$) and high risk ($U = 27.50, p = .001$). High-risk errors were detected faster by HDG than LDG, $U = 46.00, p = .043$. HDG had a higher SA score than LDG, $U = 48.50, p = .034$ and a higher level of immersion in virtual OR, $U = 43.00, p = .017$. Marginally significant results show that LDG had a higher level of subjective workload than HDG, $U = 115, p = .062$ and that HDG felt more satisfied than LDG $U = 54, p = .057$.

Next, we examined the movements in the virtual OR (see Table 2). Analysis revealed four patterns: (1) Pattern A: participants moved all around the OR (See Figure 4); (2) Pattern B: one corner (1 or 2) was ignored; (3) Pattern C: both corners were ignored; (4) Pattern D, centre and corners were ignored (see Figure 5).

Participants in HDG explored the virtual OR more than those in LDG. In HDG, seven participants followed pattern A (i.e. moved all around the OR), four followed pattern B (i.e. one corner was ignored) and two followed pattern C (i.e. both corners were ignored). In LDG, six participants followed pattern A, six followed pattern B and 1 followed pattern D. Regarding the first place visited, eight participants in each group started with the screen. In HDG, the five remaining participants chose the instrumentation table first. In LDG, one participant began with the two colleagues, one with the equipment in the centre and one with the other equipment in a corner.

Effects of participants' status

The results revealed some differences between student and expert nurses (see Table 3). Students detected more errors than the experts, $U = 113.5$, $p = .011$. Students detected more moderate risk errors, $U = 120.00$, $p = .004$ and were quicker than experts, $U = 38.00$, $p = .031$. Students visited more places than the experts, $U = 115$, $p = .009$. However, experts were more satisfied than students, $U = 41$, $p = .041$.

Analysis of movements revealed that 12/18 students and 4/8 experts started with the screen, while 5/18 students and 2/8 experts started with the instrumentation table. No participant started with the patient or the single colleague. Regarding movement patterns, the main difference between experts and students was for pattern A: 11/18 students explored the whole virtual OR compared with only 2/8 experts (see Table 2).

Analyses of debriefings after the 1st simulation session

During individual debriefings, students were asked how they evaluated their performance (self-efficacy): 3/18 assessed it as poor, 10/18 as average and 5/18 as good.

The details of the collective debriefing are presented in Table 4. The defusing phase lasted for 55 minutes. Defusing gave participants the opportunity to express themselves spontaneously and there were many verbal interactions, with several moments of interruption and one moment of silence. Participants discussed the simulator and the scenario: realism, immersion, need for time to get used to it, difficulty with the controllers or to read information in small print, problems identifying some elements, choice of the surgical specialty and questions about interpreting the errors. They also stressed the difference between their behaviour in the simulator and in real life, especially regarding the patient, as most of them realized they had not checked him first. They also felt divided between playing a game and being assessed and most of them asked to be given a mark. Despite positive feedback about the environment and acknowledging that they had learnt something, they mostly expressed negative affect: confusion, frustration, exhaustion, loneliness, lack of communication, fear (reaction, failure, sickness) and ambivalence.

The discovering phase lasted 10 minutes. There were fewer exchanges, no silence and fewer interruptions. However, participants commented on the errors they had or had not detected and mentioned again some difficulty in interpreting information or details. They also asked about some elements that they had identified as errors. They mainly expressed surprise when they discovered errors they had not found.

The deepening phase lasted 20 minutes. There were fewer exchanges than in the two previous phases, fewer interruptions but more silences. Students expressed their awareness

that the last OR they worked in was their reference point and that the simulation had alerted them to points for their next training period. They also asked for a second session, which was organized six months later.

Results of the second simulation session

Results are presented in Table 5. The mean number of detected errors was higher in session 2 than session 1, $U=7.50, p < .001$. The mean number of reported non-errors was higher in session 2 than session 1, $U=13.50, p = .004$. Moderate risk detection rate was higher in session 2 than session 1, $U=1, p < .001$. High risk detection rate was higher in session 2 than session 1, $U=19.5, p = .020$. SA was higher in session 1 than session 2, $U=162, p < .001$. Subjective workload was higher in session 1 than session 2, $U=142, p = .006$. Ease of use was higher in session 2 than session 1, $U=19, p = .019$.

Main results regarding movements in the virtual OR are presented in Table 2. Students generally started with the tactile screen, but unlike the first session, some of them started with the colleagues or the equipment. They explored the OR more fully in the second session, though the patterns of movements were reduced: moving through the whole OR (pattern A) or ignoring the empty corner (pattern B).

After this second session, a collective debriefing was organized (see Table 5). Students found it easier and felt they had responded faster in the virtual OR in the second than first session. They stressed the importance of immersion and the need to become familiar with the environment first. Overall, students expressed more satisfaction, more enjoyment and less pressure. Participants acknowledged the impact of their recent training periods and the simulation session on their ability to adapt and cope with novelty. Again, they expressed difficulty identifying small details and asked for a score or feedback of their performance to

check their progress between the two sessions. However, they said that they enjoyed immersion and were eager to use the virtual OR again.

Discussion

The objective of this study was to assess “Error recognition in a virtual OR”, a VR-based simulation scenario designed to improve the SA of scrub nurses.

Concerning the effects of the level of performance, the results partly confirm our prediction, showing that participants who detected most errors had higher SA scores and a higher level of immersion than participants who detected fewer errors. Participants who had higher SA scores perceived their workload to be lighter, felt more satisfied with the design of the scenario and also considered the simulator to be a better learning tool than those with a lower level of SA. Participants who perceived a heavier subjective workload were also more anxious and rated the scenario as a less efficient learning tool. These results suggest that this VR scenario could be a useful training tool to improve SA. It also stresses the importance of immersion in the process, not only for hedonic purposes but also for learning outcomes, by projecting oneself into the situation and engaging in it (e.g., Grassini et al. 2020; Kardong-Edgren et al., 2019).

We also observed some unexpected differences between students and experts. The results revealed that nursing students detected more errors, especially those involving moderate risk. They also explored the virtual OR more than expert nurses. However, expert nurses expressed more satisfaction with the scenario than students. Several elements can explain these differences. First, the sample of experts was small and the results should be interpreted with caution. Another explanation concerns the notion of expertise, with the idea that experts are more efficient, focusing on essential issues and wasting less time on moderate problems

(Benner, 1982, 2004). By contrast, other studies stress that intuition or experience do not necessarily have an impact on SA and decision-making, which could explain why students detected more errors (Stubbings et al., 2012). However, our results could also simply be explained by the fact that the students had just completed a course on safety and hygiene rules. This is congruent with previous research on compliance with standard precautions by OR professionals that tends to decrease with time, stressing the need for regular reminders after a few years of practice (Erasmus et al., 2010; Gammon et al., 2008). This suggests that this scenario could be useful for both initial and in-service training, especially for expert nurses who expressed more satisfaction than students with this new training modality.

Analysis of the collective debriefing highlights the need to give participants time to discover the environment and get used to HMD material. They enjoyed the realism of the scenario and said that they would like to have another session, indicating the acceptability of the scenario, which is in line with our previous study of another VR simulator we developed (Author blinded, 2019b). Students also wondered about the differences between their behaviour in the virtual OR and in a real-life situation, as they realized that they had not given their main attention to the patient. It also made them aware of their perception of errors, which is an important issue for reflective analysis and clinical reasoning (Lavoie et al., 2013).

Finally, the results of the second session with students six months later demonstrated that learning occurred in a VR simulation. The students detected more errors in the second session and perceived the workload to be lighter. They also found the simulator easier to use and expressed more satisfaction with their performance in the debriefing. Although they still said they had some difficulty identifying certain details and wanted to be given a score or feedback, they found the session had made them more aware of their ability to cope with

novelty. They enjoyed it and were eager to use the simulator again. This second session showed that the simulator had a learning effect and that students used strategies to detect a wider range of errors.

Limitations and future studies

The average number of errors detected by our participants was lower than the mean value of 9.5. More sessions are needed to see if this value is stable and to determine whether it is due to our selection or the representation of errors, or to the HMD modality that may have unsettled some participants. Participants were not asked about their previous experience of video games or HMD, as the acceptability study of the environment demonstrated that it had no impact (Author blinded, 2019b), but it could have been interesting to check that this was still valid.

One experimenter was present during each simulation, to start the session and make sure that there were no problems. This social presence may have influenced participants' behaviour and performance, especially for the expert nurses who may have felt they were being assessed. It would be interesting to control for this effect with sessions where participants are left alone.

The sample size of the present study was limited and concerned only students and professionals from one hospital. Future studies should test the external validity of the "Error Recognition in a Virtual OR" scenario in other samples and see whether its use can be extended to other professions. Future development of the scenario is also expected, introducing more errors as well as an authoring interface to enable teachers to focus on a category of errors relevant to the profile of their students or to their teaching objectives. The

use of aerial views of movements for debriefing could also be studied (Levett-Jones and Lapkin, 2014).

Conclusion

In this study, we studied the perceived workload, anxiety and SA of participants immersed in a virtual surgical simulation scenario called “Error Recognition in a Virtual OR”. The results confirm our expectations, namely that participants who look for more information about the situation have better SA and identify more errors in the virtual OR. However, contrary to our expectations, the students performed better than the experts, which indicates the importance of providing in-service training for healthcare professionals. Moreover, as our scenario dealt with safety and hygiene, our results confirm the need for regular reminders about this issue.

The scenario enabled SA to be assessed and thus seems to be a suitable tool to improve this NTS, in both initial and in-service training. The VR technology used for the simulator was generally well accepted and motivated trainees. It provides data that can be used constructively in debriefing. Further development will involve using the scenario with other healthcare professionals working in the OR.

References

Alhalabi, W. (2016). Virtual reality systems enhance students’ achievements in engineering education. *Behaviour & Information Technology*, 35(11), 919–925.

<https://doi.org/10.1080/0144929X.2016.1212931>

Author blinded (2019a)

Author blinded (2019b)

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change.

Psychological Review, 84(2), 191.

Bekker, H. L., Legare, F., Stacey, D., O'Connor, A., & Lemyre, L. (2003). Is anxiety a suitable measure of decision aid effectiveness: A systematic review? *Patient Education and Counseling*, 50(3), 255–262. [https://doi.org/10.1016/S0738-3991\(03\)00045-4](https://doi.org/10.1016/S0738-3991(03)00045-4)

Benner, P. (1982). From novice to expert. *American Journal of Nursing*, 82, 402–407.

Benner, P. (2004). Using the Dreyfus Model of Skill Acquisition to Describe and Interpret Skill Acquisition and Clinical Judgment in Nursing Practice and Education. *Bulletin of Science, Technology & Society*, 24(3), 188–199. <https://doi.org/10.1177/0270467604265061>

Bouville, R., Gouranton, V., Boggini, T., Nouviale, F., & Arnaldi, B. (2015). # FIVE: High-level components for developing collaborative and interactive virtual environments. *Proceedings of Eighth Workshop on Software Engineering and Architectures for Realtime Interactive Systems (SEARIS 2015), conjunction with IEEE Virtual Reality (VR)*, Mar 2015, Arles, France, 33–40.

Brady, P. W., Muething, S., Kotagal, U., Ashby, M., Gallagher, R., Hall, D., Goodfriend, M., White, C., Bracke, T. M., DeCastro, V., & others. (2013). Improving situation awareness to reduce unrecognized clinical deterioration and serious safety events. *Pediatrics*, 131(1), e298–e308.

- Byers, J. C., Bittner, A. C., & Hill, S. G. (1989). Traditional and raw task load index (TLX) correlations: Are paired comparisons necessary? In E. Mital (Ed.), *Advances in industrial ergonomics and safety* (pp. 481–485). Taylor & Francis.
- Caillaud, B. (2013). Surgical Process Mining with Test and Flip Net Synthesis. *Application of Region Theory (ART)*, Jul 2013, Barcelona, Spain. pp.43-54. hal-00872284
- Chen, F. Q., Leng, Y. F., Ge, J. F., Wang, D. W., Li, C., Chen, B., & Sun, Z. L. (2020). Effectiveness of Virtual Reality in Nursing Education: Meta-Analysis. *Journal of Medical Internet Research*, 22(9), e18290. <https://doi.org/10.2196/18290>
- Christensen, N. H., Hjermitsev, O. G., Stjernholm, N. H., Falk, F., Nikolov, A., Kraus, M., Poulsen, J., & Petersson, J. (2018). Feasibility of team training in virtual reality for robot-assisted minimally invasive surgery. *Proceedings of the Virtual Reality International Conference-Laval Virtual*, 1–4.
- Claude, G., Gouranton, V., Caillaud, B., Gibaud, B., Jannin, P., & Arnaldi, B. (2016). From Observations to Collaborative Simulation: Application to Surgical Training. *International Conference on Artificial Reality and Telexistence, Eurographics Symposium on Virtual Environments*. <https://hal.archives-ouvertes.fr/hal-01391776/>
- Endsley, M. (2012). Situation awareness. In G. Salvendy (Ed.), *Handbook of human factors and ergonomics*. John Wiley & Sons.
- Endsley, M. R. (1995). Toward a theory of situation awareness in dynamic systems. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 37(1), 32–64. <https://doi.org/10.1518/001872095779049543>

- Endsley, M. R. (2015). Final reflections: Situation awareness models and measures. *Journal of Cognitive Engineering and Decision Making*, 9(1), 101–111.
<https://doi.org/10.1177/1555343415573911>
- Erasmus, V., Daha, T. J., Brug, H., Richardus, J. H., Behrendt, M. D., Vos, M. C., & van Beeck, E. F. (2010). Systematic Review of Studies on Compliance with Hand Hygiene Guidelines in Hospital Care. *Infection Control & Hospital Epidemiology*, 31(3), 283–294.
<https://doi.org/10.1086/650451>
- Flin, R., O'Connor, P., & Crichton, M. (2008). *Safety at the sharp end, a guide to non-technical skills*. Ashgate Publishing Limited.
- Flin, R., Youngson, G. G., & Yule, S. (2015). *Enhancing Surgical Performance: A Primer in Non-technical Skills*. CRC Press.
- Flurry, M., Brooke, S., Micholetti, B., Natoli, N., Moyer, K., Mnich, S., & Potochny, J. (2012). Nurse training with simulation: An innovative approach to teach complex microsurgery patient care. *Annals of Plastic Surgery*, 69(4), 459–461.
<https://doi.org/10.1097/SAP.0b013e31824b3db4>
- Gallagher, A.G., & Satava, R. M. (2002). Virtual reality as a metric for the assessment of laparoscopic psychomotor skills. *Surgical Endoscopy*, 16(12), 1746–1752.
<https://doi.org/10.1007/s00464-001-8215-6>
- Gallagher, Anthony G., Ritter, E. M., Champion, H., Higgins, G., Fried, M. P., Moses, G., Smith, C. D., & Satava, R. M. (2005). Virtual reality simulation for the operating room. *Annals of Surgery*, 241(2), 364–372. <https://doi.org/10.1097/01.sla.0000151982.85062.80>

- Gammon, J., Morgan-Samuel, H., & Gould, D. (2008). A review of the evidence for suboptimal compliance of healthcare practitioners to standard/universal infection control precautions. *Journal of clinical nursing*, *17*(2), 157–167. <https://doi.org/10.1111/j.1365-2702.2006.01852.x>
- Gluyas, H., Stomski, N., Andrus, P., Walters, J., Morrison, P., Williams, A., Hopkins, M., & Sandy, M. (2019). Performance based situation awareness observations in a simulated clinical scenario pre and post an educational intervention. *Nurse Education in Practice*, *36*, 20–27. <https://doi.org/10.1016/j.nepr.2019.02.011>
- Grassini, S., Laumann, K., & Rasmussen Skogstad, M. (2020). The Use of Virtual Reality Alone Does Not Promote Training Performance (but Sense of Presence Does). *Frontiers in psychology*, *11*, 1743. <https://doi.org/10.3389/fpsyg.2020.01743>
- Guilloton, D. (2016). De la chambre des erreurs au bloc opératoire des erreurs : Une expérience de simulation. *Interbloc 35*, 182-185. <https://doi.org/10.1016/j.bloc.2016.07.013>
- Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In *Advances in Psychology* (Vol. 52, pp. 139–183). Elsevier. [https://doi.org/10.1016/S0166-4115\(08\)62386-9](https://doi.org/10.1016/S0166-4115(08)62386-9)
- JASP Team. (2018). *JASP (Version 0.9)[Computer software]*. <https://jasp-stats.org/>
- Kang, E., Massey, D., & Gillespie, B. M. (2015). Factors that influence the non-technical skills performance of scrub nurses: A prospective study. *Journal of Advanced Nursing*, *71*(12), 2846–2857. <https://doi.org/10.1111/jan.12743>

- Kardong-Edgren, S. S., Farra, S. L., Alinier, G., & Young, H. M. (2019). A call to unify definitions of virtual reality. *Clinical Simulation in Nursing*, *31*, 28–34.
<https://doi.org/10.1016/j.ecns.2019.02.006>
- Kyaw, B. M., Saxena, M., Posadzki, P., Vseteckova, J., Nikolaou, C. K., George, P. P., Divakar, U., Masiello, I., Kononowicz, A. A., Zary, N., & Car, L. T. (2019). Virtual reality for health professions education: Systematic review and meta-analysis by the digital health education collaboration. *Journal of Medical Internet Research*, *21*(1), e12959.
<https://doi.org/10.2196/12959>
- Lavoie, P., Pepin, I., & Boyer, L. (2013). Reflective debriefing to promote novice nurses' clinical judgment after high-fidelity clinical. *Dynamics*, *24*(4), 36–41.
- Levett-Jones, T., & Lapkin, S. (2014). A systematic review of the effectiveness of simulation debriefing in health professional education. *Nurse Education Today*, *34*(6), e58–e63.
<https://doi.org/10.1016/j.nedt.2013.09.020>
- MacCallum, R. C., Zhang, S., Preacher, K. J., & Rucker, D. D. (2002). On the practice of dichotomization of quantitative variables. *Psychological Methods*, *7*(1), 19–40.
<https://doi.org/10.1037/1082-989x.7.1.19>
- Makransky, G., Terkildsen, T. S., & Mayer, R. E. (2019). Adding immersive virtual reality to a science lab simulation causes more presence but less learning. *Learning and Instruction*, *60*, 225–236. <https://doi.org/10.1016/j.learninstruc.2017.12.007>
- Markansky, G. andreasen, N. K., Baceviciute, S., & Mayer, R. E. (2020). Immersive virtual reality increases liking but not learning with a science simulation and generative learning

- strategies promote learning in immersive virtual reality. *Journal of Educational Psychology*. Advance online publication. <https://doi.org/10.1037/edu0000473>
- Marteau, T. M., & Bekker, H. (1992). The development of a six-item short-form of the state scale of the Spielberger State-Trait Anxiety Inventory (STAI). *British Journal of Clinical Psychology*, *31*(3), 301–306. <https://doi.org/10.1111/j.2044-8260.1992.tb00997.x>
- Mirek, S., & Prétot, R. (2019). Le “bloc des erreurs” en simulation, apprendre autrement. *Inter Bloc*, *6502*(1), 28–31. <http://dx.doi.org/10.1016/j.bloc.2019.01.004>
- Mitchell, L., Flin, R., Yule, S., Mitchell, J., Coutts, K., & Youngson, G. (2011). Thinking ahead of the surgeon. An interview study to identify scrub nurses’ non-technical skills. *International Journal of Nursing Studies*, *48*(7), 818–828. <https://doi.org/10.1016/j.ijnurstu.2010.11.005>
- Mitchell, L., Flin, R., Yule, S., Mitchell, J., Coutts, K., & Youngson, G. (2012). Evaluation of the Scrub Practitioners’ List of Intraoperative Non-Technical Skills (SPLINTS) system. *International Journal of Nursing Studies*, *49*(2), 201–211. <https://doi.org/10.1016/j.ijnurstu.2011.08.012>
- Pons Lelardeux, C., Panzoli, D., Lubrano, V., Minville, V., Lagarrigue, P., & Jessel, J.-P. (2017). Communication system and team situation awareness in a multiplayer real-time learning environment: Application to a virtual operating room. *The Visual Computer*, *33*(4), 489–515. <https://doi.org/10.1007/s00371-016-1280-6>
- Salas, E., Cooke, N. J., & Rosen, M. A. (2008). On teams, teamwork and team performance: discoveries and developments. *Human factors*, *50*(3), 540–547. <https://doi.org/10.1518/001872008X288457>

Spielberger, C. D. (1983). *Manual for the State-Trait Anxiety Inventory STAI (Form Y)*.

Consulting Psychologists Press.

Stubblings, L., Chaboyer, W., & McMurray, A. (2012). Nurses' use of situation awareness in decision-making: An integrative review: Nurses' use of situation awareness in decision-making: an integrative review. *Journal of Advanced Nursing*, *68*(7), 1443–1453.

<https://doi.org/10.1111/j.1365-2648.2012.05989.x>

Taylor, R. M. (1990). Situation awareness rating technique (SART): The development of a tool for aircrew systems design. *Situational Awareness in Aerospace Operations (Chapter 3)*.

Webster, R. (2016). Declarative knowledge acquisition in immersive virtual learning environments. *Interactive Learning Environments*, *24*(6), 1319–1333.

<https://doi.org/10.1080/10494820.2014.994533>

World Health Organization. (2017). *WHO simulation exercise manual: A practical guide and tool for planning, conducting and evaluating simulation exercises for outbreaks and public health emergency preparedness and response*. World Health Organization.

<https://apps.who.int/iris/handle/10665/254741>

Zigmont, J. J., Kappus, L. J., & Sudikoff, S. N. (2011). The 3D Model of Debriefing: Defusing, Discovering and Deepening. *Seminars in Perinatology*, *35*(2), 52–58.

<https://doi.org/10.1053/j.semperi.2011.01.003>

Table 1. Effects of the levels of performance on the different measures

Measures	Low Detection Group			High Detection Group			Diff.	
	(n=13)			(n=13)			U	p
	M (SD)	Min	Max	M (SD)	Min	Max		
Number of detected errors	4.69 (1.75)	2.00	7.00	10.38 (2.14)	8.00	15.00	0.00	<.001
% of detected errors	24.68	10.52	36.84	54.63	42.10	78.95	0.00	<.001
Mean detection rate	0.25 (0.09)	0.11	0.37	0.55 (0.11)	0.42	0.79	0.00	<.001
For moderate risk	0.27 (0.13)	0.08	0.50	0.60 (0.14)	0.33	0.83	7.50	<.001
For high risk	0.21 (0.14)	0.00	0.43	0.45 (0.19)	0.14	0.71	27.50	.001
Mean detection time	5.86 (1.71)	2.40	9.22	6.34 (1.95)	3.45	10.80	71.00	.505
For moderate risk	6.26 (2.25)	2.35	10.26	5.92 (2.21)	3.35	9.93	95.00	.614
For high risk	5.33 (2.31)	1.46	8.32	7.50 (2.62)	3.41	12.26	46.00	.043
Number of visited places	13.46 (3.60)	8.00	19.00	15.08 (4.73)	4.00	24.00	65.50	.339
Situation awareness	15 (11.75)	-8.00	28.00	23.31 (7.42)	10.00	34.00	48.50	.034
Subjective workload	54.61 (19.90)	11.50	82.00	45.55 (19.04)	14.00	76.50	115.00	.062
Anxiety	40.26 (14.11)	20.00	70.00	41.28 (14.31)	23.33	63.33	80.00	.602
Ease of use	3.60 (1.19)	1.00	5.00	3.59 (0.97)	1.33	5.00	93.50	.660

Immersion	3.58 (1.11)	1.25	5.00	4.46 (0.68)	3.00	5.00	43.00	.017
Efficiency	3.85 (1.21)	1.50	5.00	4.19 (0.60)	3.00	5.00	79.50	.811
Satisfaction	3.61 (1.32)	1.00	5.00	4.46 (0.56)	3.50	5.00	54.00	.057
Behavioural intention	4.15 (1.34)	1.00	5.00	5.00 (0.00)	5.00	5.00	NaN	

Note: We used one-tailed test (Mann-Whitney) with hypothesis of Low Detection Group > High Detection Group for subjective workload and anxiety measures and Low Detection Group < High Detection Group for moderate and high risk detection rates, mean, moderate and high risk detection time, SA, ease of use, immersion, efficiency, satisfaction and behavioural intention measures.

Table 2. First places visited and patterns of movements in the virtual OR according to the levels of performance and status of participants

	Low Detection Group	High Detection Group	Nursing Students		Expert Nurses
			Session 1	Session 2	
Screen first	8	8	12	10	4
Instrumentation table first	2	5	5	2	2
2 colleagues first	1	0	0	2	1
Patient first	0	0	0	0	0
Equipment first	1	0	0	4	1
Other equipment first	1	0	1	0	0
Pattern A	6	7	11	13	2
Pattern B	6	4	5	5	5
Pattern C	0	2	1	0	1

Pattern D 1 0 1 0 0

Table 3. Effects of participants' status on the different measures

Measures	Nursing students (n= 18)			Expert nurses (n= 8)			Diff.	
	M (SD)	Min	Max	M (SD)	Min	Max	U	P
Number of detected errors	8.56 (3.57)	3.00	15.00	5.25 (1.91)	2.00	8.00	113.5	.011
% of detected errors	45.05	15.79	78.95	27.63	10.53	42.10	113.5	.011
Mean detection rate	0.45 (0.19)	0.16	0.79	0.28 (0.10)	0.11	0.42	113.5	.011
For moderate risk	0.51 (0.21)	0.08	0.83	0.26 (0.09)	0.08	0.33	120.00	.004
For high risk	0.34 (0.22)	0.00	0.71	0.30 (0.18)	0.00	0.57	75.00	.887
Mean detection time	5.88 (1.26)	3.45	8.51	6.59 (2.62)	2.40	10.80	59.50	.505
For moderate risk	5.54 (1.80)	3.35	9.56	7.32 (2.62)	2.35	10.26	38.00	.031
For high risk	6.80 (2.20)	3.20	12.26	5.56 (3.66)	1.46	11.68	79.00	.357
Number of visited places	15.67 (3.55)	9.00	24.00	11.12 (4.02)	4.00	16.00	115.00	.009
Situation awareness	18.33 (10.47)	-8.00	33.00	21.00 (11.10)	4.00	34.00	59.50	.504
Subjective workload	51.39 (20.07)	14.00	82.00	47.12 (19.6)	11.50	67.50	83.00	.559

Anxiety	41.85 (13.82)	23.30	70.00	38.33 (14.80)	20.00	63.30	83.50	.539
Ease of use	3.66 (0.96)	1.33	5.00	3.46 (1.33)	1.00	5.00	76.00	.844
Immersion	3.96 (1.08)	1.25	5.00	4.15 (0.89)	2.50	5.00	64.50	.694
Efficiency	3.94 (0.89)	1.50	5.00	4.19 (1.13)	2.00	5.00	53.00	.288
Satisfaction	3.81 (1.16)	1.00	5.00	4.56 (0.68)	3.50	5.00	41.00	.041
Behavioural intention	4.50 (1.15)	1.00	5.00	4.75 (0.71)	3.00	5.00	65.00	.600

Note: We used one-tailed test (Mann-Whitney) with hypothesis of Nursing students > Expert nurses for mean and moderate risk detection rates and number of visited places measures and Nursing students < Expert nurses for moderate risk detection time and satisfaction measures.

Table 4. Collective debriefings with nursing students

	Session 1			Session 2		
	Defusing	Discovering	Deepening	Defusing	Discovering	Deepening
Duration	55 min (56.25%)	10 min (12.50%)	20 min (31.25%)	5 min (25%)	10 min (50%)	5 min (25%)
Oral interactions	161	52	48	14	14	11
Silences	1	0	4	0	2	0
Laughter	23	16	8	1	2	0
Students' assertions	Realism, immersion,	Comments on the errors they had	Awareness that the last OR they	Easier and quicker	Importance of their experience	Ideas for future use

	time to adapt.	found or not	worked is their reference point	Importance of immersion and time for appropriation	in the last training periods
	Difficulties with controllers, reading, identification for some elements, surgical specialty (neurosurgery), interpretation of errors	Difficulties for reading information	Having a reference OR gives a feeling of security but need to detach from it before the next training period	More perspective on their activity	Awareness of their ability to cope with novelty
	Differences with ordinary behaviours, contrast between game and assessment	Interpretation of errors with question on elements that were not considered as errors in the scenario	Warning points for the next training period		Difficulty in reading or memorize information
	Need for a score		Ask for a second session		Need for a score or feedback on the number of detected errors
Students' affects	Confusion, frustration, exhaustion, loneliness, lack of communication, fear (reaction, failure, sickness), ambivalence	Surprise	Fun aspect of simulation Positive conclusion	Satisfaction, fun, less pressure	Enjoyed immersion, eager to go back in the virtual OR
	Admiration for the work, reassured to see the knowledge acquired				

Table 5. Comparisons of the two simulation sessions

Measures	Session 1			Session 2			Diff.	
	(n=18)			(n=18)			U	<i>p</i>
	M (SD)	Min	Max	M (SD)	Min	Max		
Number of detected errors	8.56 (3.57)	3	15	11.67 (2.14)	7	16	7.50	<.001
% of detected errors	45.05	15.79	78.95	61.42	36.84	84.21	7.50	<.001
Nb of reported non errors	1.78 (1.66)	0	6	3.72 (2.89)	0	12	13.50	.004
Mean detection rate	0.45 (0.19)	0.16	0.79	0.61 (0.11)	0.37	0.84	9.00	<.001
For moderate risk	0.51 (0.21)	0.08	0.83	0.70 (0.15)	0.42	1	1	<.001
For high risk	0.34 (0.22)	0.00	0.71	0.45 (0.12)	0.29	0.71	19.50	.020
Mean detection time	5.88 (1.36)	3.45	8.51	5.16 (1.70)	2.47	7.99	118.00	.163
For moderate risk	5.54 (1.80)	3.35	9.56	4.62 (1.81)	1.99	7.89	116.00	.196
For high risk	6.80 (2.20)	3.20	12.26	6.54 (2.11)	2.78	9.83	95.00	.702
Number of visited places	15.67 (3.55)	9.00	24.00	17.83 (3.62)	13.00	26.00	50.50	.132
Situation awareness	18.33 (10.47)	-8.00	33.00	5.33 (1.19)	2	7	162	<.001
Subjective workload	51.39 (20.07)	14.00	82.00	41.37 (13.01)	19.50	72.50	142	.006
Ease of use	3.66 (0.96)	1.33	5.00	4.17 (0.51)	3	4.67	19	.019

Immersion	3.96 (1.08)	1.25	5.00	4.11 (0.57)	3.00	5.00	57.50	.909
Efficiency	3.94 (0.89)	1.50	5.00	4.08 (0.65)	3.00	5.00	31.00	.549
Satisfaction	3.81 (1.16)	1.00	5.00	3.82 (0.67)	2.50	5.00	66.00	.750
Behavioural intention	4.50 (1.15)	1.00	5.00	4.72 (0.46)	4.00	5.00	4.00	.410

Note: We used one-tailed test (Wilcoxon) with hypothesis of Session 1 > Session 2 for SA and Workload,

and Session 1 < Session 2 for number of detected errors and non-errors, moderate and high risk detection rates,

and ease of use measures.

Figure captions

Figure 1- Photo of teleportation in the virtual OR



Figure 2- Photo of an error



Figure 3- Photo of an aerial view of the virtual OR with the four zones

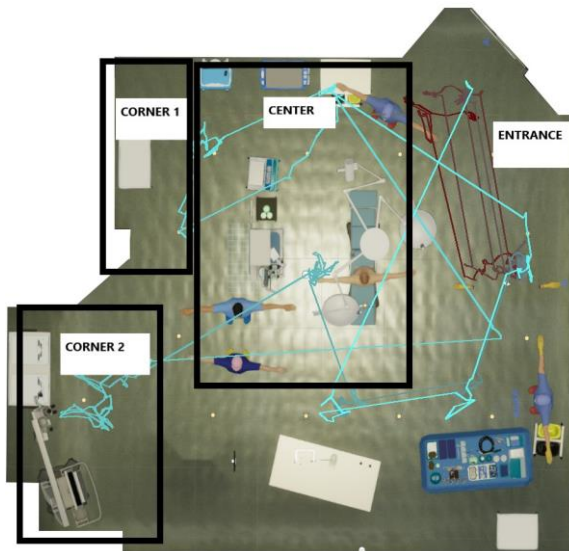


Figure 4- Movements in the OR: pattern A

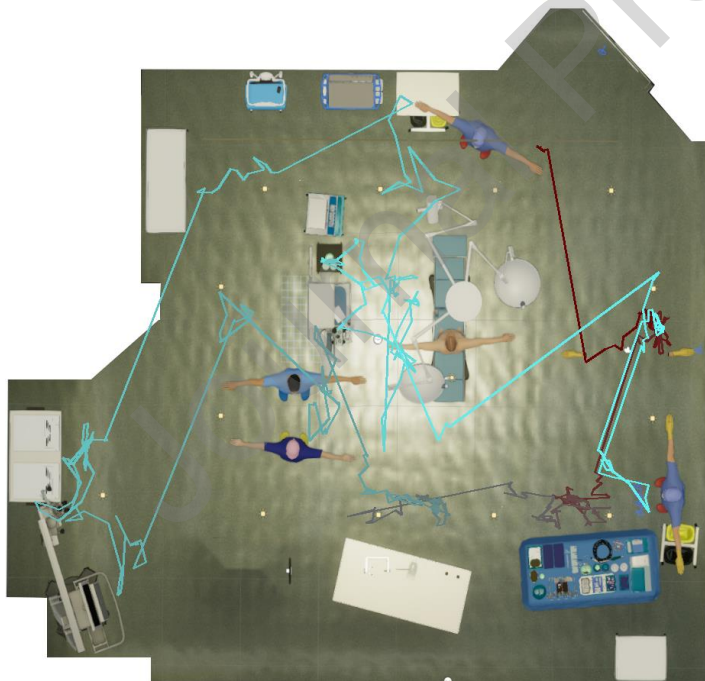
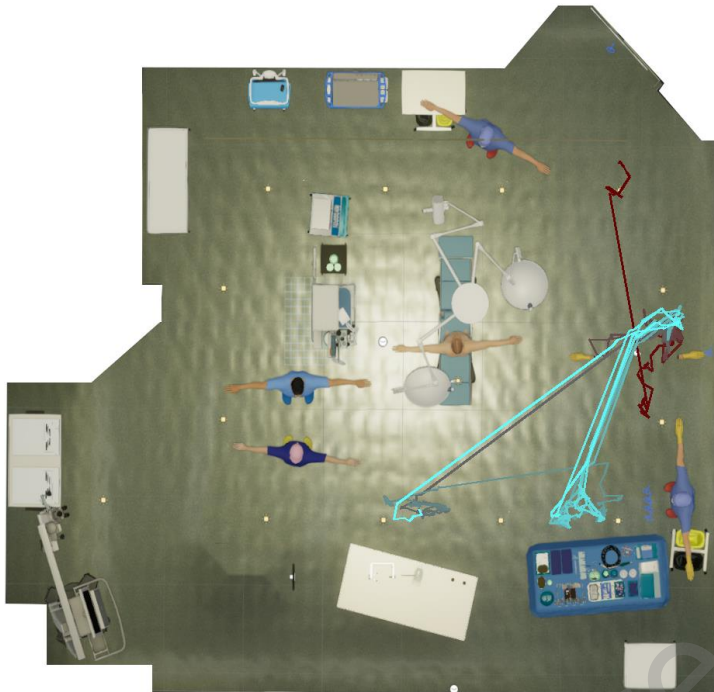


Figure 5- Movements in the OR: pattern D



Author statement

MS Bracq, E Michinov, P Jannin, B Arnaldi, V Gouranton: conceived and designed the experiments. MS Bracq, M Le Duff: performed the experiments. MS Bracq, E Michinov, P Jannin: analyzed and interpreted the data. MS Bracq, E Michinov, P Jannin, B Arnaldi, V Gouranton, M Le Duff: contributed materials and procedure, analysis tools or data. MS Bracq, E Michinov, P Jannin, B Arnaldi, V Gouranton: wrote and revised the paper

Journal Pre-proof

Highlights

- The VR scenario allows an assessment of situation awareness
- The VR scenario is a suitable tool for initial and in-service training
- VR technology is well accepted and motivates trainees
- VR simulator provides data for personalized debriefing (patterns of movements)

Journal Pre-proof