Virtual Reality (2025) 29:78
https://doi.org/10.1007/s10055-025-01159-4

ORIGINAL ARTICLE

Medical training in virtual reality: a gamification approach

Luca Turchet[1] · Federico Gentilini[2] · Sarah Malandra[3] · Alessandro Veccia[3] · Alessandro Antonelli[3] · Cesare Scoffone[4]

Received: 28 August 2024 / Accepted: 24 April 2025

Abstract
Medical education is one of the domains that is currently being widely investigated leveraging the capabilities offered by virtual reality (VR) systems. The appeal of such technology is based on the potential cost-effectiveness, portability, safety of training in simulated environments, and the ability to enable training without the need of supervision. One of the approaches that can be utilized during technology-mediated educational activities is gamification, i.e., the use of game design elements in non-game contexts. This approach has the ability to make learning fun, memorable and more effective, as demonstrated by a substantial body of literature. However, whereas a number of studies have investigated the ability of gamification-based VR systems in enhancing learning and training in various domains, the adoption of gamification approaches in VR medical training, and in particular surgical training, is a topic that has been largely overlooked. To bridge this gap, we first co-designed with a pool of urology surgeons a gamification-based VR system for the laser enucleation of the prostate. Subsequently, we conducted a user study with seventeen urology residents to assess the usability and user experience. Our results provide evidence that gamification in VR medical training systems is a valuable strategy to enhance surgical trainees outcomes and motivations. However, our findings also revealed that the lack of realism of the physical aspects involved in real operations, such as force and tactile feedback and visual deformations of the simulated tissues, can drastically hamper the experience that surgeons desire from a VR simulator.

Keywords Virtual reality · Medical training · Gamification · Surgery · Urology · Procedural skills

1 Introduction

Immersive virtual reality (VR) is the three-dimensional digital representation of a real or imagined space with interactive capabilities, which provides the perception of being physically present in such a non-physical space. This technology is rapidly evolving at both the hardware and software level, with devices and applications becoming more user friendly and economically accessible. In particular, VR is an increasingly used medium in educational contexts, leading to innovative forms of training for a large variety of tasks (Freina and Ott 2015; Fowler 2015; Hu et al. 2016; Howard and Gutworth 2020).

Various studies have assessed the benefits of VR training compared to traditional forms of training, encouraging the widespread use of VR technologies in learning contexts (Kern et al. 2019; Abich et al. 2021; Howard et al. 2021). Instead of being passive observers, learners engage in virtual learning environments as active participants, which enables the development of exploration-based learning paradigms. A useful application of VR technologies is that of supporting the development of skills that cannot be easily or safely trained otherwise (e.g., flying, surgery). Indeed, VR offers the possibility of immediate feedback, which promotes more accurate training to self-correct mistakes in environments that are otherwise risky or provide unsafe conditions. In addition, VR-based simulation programs may be more cost-effective than the traditional learning counterparts, they can support ubiquitous learning rather than forcing learners into a particular environment, as well as have the potential to eliminate the need for teaching materials and/or human trainers.

✉ Luca Turchet
luca.turchet@unitn.it

[1] Department of Information Engineering and Computer Science, University of Trento, Trento, Italy

[2] R&D Department, Digital Mosaik, Trento, Italy

[3] Urology Unit, University Hospital of Verona, Verona, Italy

[4] Department of Urology, Cottolengo Hospital, Torino, Italy

Medical education is one of the domains that is currently being widely investigated leveraging the capabilities offered by VR systems (Zhao et al. 2021; Mehrfard et al. 2021). Such a domain requires practitioners to develop clinical skills before dealing with real patients. In part, the acquisition of these skills is traditionally achieved by practicing on artificial models as well as animals' or humans' cadavers. This minimizes to a great extent the incidence of human error during training with the real patient, and relieves the trainee's anxieties of dealing with real patients by acquiring a good level of skill prior to that stage. VR represents a possibility to modernize current teaching methods by offering realistic simulations of real-world training scenarios. The appeal of such technology is based on the potential cost-effectiveness, portability, safety of training in virtual environments, and the ability to allow training without the need for supervision.

Much of the literature about medical training in VR has focused on surgical training (Mao et al. 2021; Li et al. 2021; Ganni et al. 2020; Pérez-Escamirosa et al. 2023; Frederiksen et al. 2020). For instance, research has focused on suturing (Yu et al. 2020), laparoscopy (Jin et al. 2021) and ophthalmology (Thomsen et al. 2017). VR was applied to measure the operative skills among surgeons (Ahlberg et al. 2002), to warm up before the surgeries in expert surgeons (EA Araujo et al. 2014), as well as to decrease mental and physical workload in novice surgeons (Barré et al. 2019). However, to the authors' best knowledge, the use of VR applications for urology surgical training has been scarcely investigated. Only a handful of studies have been conducted on this particular area (Hamacher et al. 2016, 2018).

One of the approaches that can be utilized during technology-mediated educational activities is gamification (Seaborn and Fels 2015; Krath et al. 2021). Gamification has been defined as "the use of game design elements in non-game contexts" (Deterding et al. 2011). Game elements are, for example, points, badges, levels, avatars, quests, social graphs, leader boards, or certificates (Zainuddin et al. 2020). Serious games are the result of the application of the gamification paradigm to a context different from entertainment. They are games designed for a specific purpose related to training (Dörner et al. 2016). Differently from traditional teaching environments based on teacher-centered approach, in which the teacher controls the learning, serious games are focused on a learner-centered approach to education. In this way, the trainee feels in control of an interactive learning process, which facilitates active and critical learning. It is well known that active learning modalities, including games, are known to increase knowledge retention (Schuller et al. 2015; Felszeghy et al. 2019). The implementation of game design elements in real-world contexts for non-gaming purposes has been applied in a variety of educational settings with the aim of fostering students' motivation and performance in relation to a given learning activity. As a matter of fact, gamification has the ability to make learning fun, memorable and more effective, as demonstrated by various scholars (Caponetto et al. 2014; Nah et al. 2014; Dicheva et al. 2015; Kiryakova et al. 2014).

Gamification approaches have been adopted also in medical education (McCoy et al. 2016; Gorbanev et al. 2018) and various studies have demonstrated that gamification can carry several benefits to medical students (for a recent review see (Krishnamurthy et al. 2022)). On the other hand, a number of studies have investigated the ability of immersive virtual reality serious games in enhancing learning and training in various domains (Checa and Bustillo 2020). However, the adoption of gamification approaches in VR medical training is a topic that has been largely overlooked. The literature provides only a handful of studies on such a topic (Wilson et al. 2017; Lenz et al. 2021; Chávez et al. 2020; Yang and Oh 2022), and to the authors' best knowledge no study has investigated the use of gamification for surgery training in VR.

To bridge these gaps we first co-designed with a pool of urology surgeons UROVR (acronym for Urology-VR), a gamification-based VR system for the simulated laser enucleation of the prostate, a common operation in urology (Das et al. 2019; Scoffone and Cracco 2016; Tuccio et al. 2020). This operation requires the surgeon to follow anatomical marks to safely enucleate the prostate adenoma. Subsequently, we conducted a user study with urology residents to assess the system validity and benefits. Specifically, our research questions were:

- How well does a gamification approach support urology residents in their process of learning a surgery procedure?
- How does the developed system compare with traditional learning approaches?
- Does a system like UROVR provide a high user experience, adequate to support surgeons' learning of a surgery procedure?
- How can a system like UROVR integrate into and extend conventional practices to support the learning of surgery practices?

Based on the extensive previous findings on the usefulness of gamification techniques as well as of co-design processes conducted with the end users (Bannon et al. 2018), our hypothesis was that UROVR would have provided urology surgeons trainees with an effective learning environment. Nevertheless, we were also interested in assessing the limitations of the developed system.

2 Co-design

The phase of co-design lasted 8 months and was conducted in parallel with the implementation of the solution and its evaluation in an iterative fashion. It involved three engineers (one expert in human-computer interaction and two VR developers) and five surgeons (two teachers and three trainees). The developed system resulted from a tight interaction between these experts with complementary backgrounds. The requirements gathering and co-design activities included four focus groups and dozens of interviews and online sessions. The engineers were also provided with several videos of the operation, including videos with framings on the hands of the operating surgeons and from the perspective of the laser. They also attended in person twice the operation to further understand the needs of surgeons and the details of the procedures.

The involved surgeons reported to be unsatisfied about traditional training methods and the need to modernize them. Typical training procedures to teach the enucleation of the prostate via a laser involve the presence of a surgeon trainer, the training on cadavers of pigs and humans, and the use of video recordings. These methods were deemed time consuming and not cost-effective. In particular, they required a lot of effort (in terms of bureaucracy and arrangements) to setup the training with cadavers, and the need to move in specific places at determined times. The issue of conducting training in physical presence arose particularly during the social restrictions due to the recent COVID-19 pandemic (Gómez Rivas et al. 2023). VR was seen as a valuable alternative because it could offer ubiquity of the training at an affordable cost.

Therefore, the set goal was to utilize the unique features of the VR medium to provide an enhancement to the traditional methods of delivering urological knowledge to trainees via a tool that augments the teaching process. In particular, it was deemed crucial to provide real-time feedback about correct and incorrect actions performed by trainees at any step of the operation. This is an aspect that is difficult to achieve during traditional teaching methods, which are less accurate than methods based on computational approaches able to identify timely and with high precision the correct and wrong surgical cuts. Gamification was selected as a method to motivate trainees and foster the reuse of the application. In addition, an important requirement emerged from the co-design sessions was to split the training in different parts to allow learners to gradually achieve a sufficient level of autonomy in performing the steps of the operation in the right order.

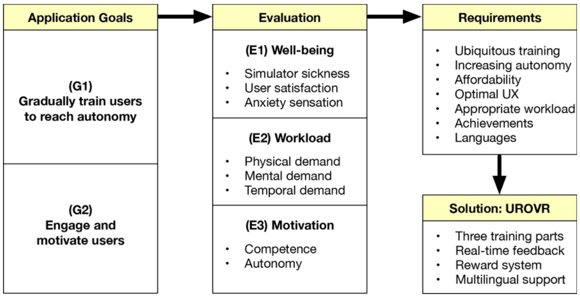


Fig. 1 Overarching methodological framework. Firstly, we derived the evaluation categories from our application goals. Subsequently, we defined the requirements and proposed our solution, named UROVR

However, rather than recreating a highly realistic simulation environment providing the experience of acting in an operating room, it was decided that trainees would have benefited from a learning system focused on explaining the sequential steps of the investigated operation. Therefore, we opted for a simulation environment centered on visual aspects which also utilized a narrating voice for the explanation, rather than rendering the typical soundscape of an operating room during the operation and the haptic sensations resulting from the handling of the surgical tools. Moreover, it was decided that it was not necessary to simulate every single aspect of the operation and that for the aimed training purposes it was sufficient to include in the simulation only the most important components of the overall procedure. The close interaction between engineers and surgeons also helped determine the correct terminology to use in the application. Furthermore, a crucial aspect was that of defining all gamification aspects, including the criteria to use to assign the scores and how to visualize them without distracting too much the trainee from the activity.

Several iterations of design-implementation-evaluation phases were accomplished before achieving the final design. Such interaction design cycles adhered to an overarching methodology, which was defined after the identification of the user needs and pain points described above. Figure 1 provides an overview of the overall methodological framework adopted, which was in part inspired by the study reported in (Kern et al. 2019).

In the first step, we derived the evaluation categories from our application goals. We then defined the requirements and proposed our solution. We defined two primary goals for our VR medical training approach. The first goal was to gradually train users to reach autonomy in the training (G1). We aimed at creating an application that could be used by trainees in full autonomy for self-learning practices. However, the creation of just one application where users could directly practice the surgical actions was deemed to be insufficient to achieve a proper training given the complexity of the operation and its multiple steps. The second goal was to motivate users to learn while providing an engaging virtual environment in which to practice (G2). We determined three subcategories (Well-being, workload, and motivation) for the evaluation of our two primary goals. The first subcategory evaluates the overall well-being (E1) by simulator sickness, user satisfaction, and anxiety sensation. The second subcategory dealt with the evaluation of experienced workload (E2) by physical, mental, and temporal demand. The third subcategory comprised the evaluation of the perceived motivation (E3) after the VR training. This part includes the factors competence and autonomy.

We defined seven requirements for our VR medical training system based on these three evaluation categories.

- *Increasing level of autonomy:* the system should allow the user to achieve a satisfactory self-training level without the need to rely on an instructor either in physical presence or remotely connected;
- *Ubiquitous training:* the system should be portable and bound to a specific place so to allow training ubiquitously;
- *Affordability:* the system should be cost-effective and leverage technologies widely used and easy to find on the market;
- *Optimal user experience:* the system should provide an optimal user experience for trainees, enhancing the sense of presence in the virtual environment and minimizing cyber sickness;
- *Appropriate workload:* the system should not impose an excessive workload on the user in terms of physical, mental, and temporal demands;
- *Achievements:* the system should reward the user for completing certain training goals;
- *Languages:* due to the different language skills of the potential users (i.e., medical students), the application should be implemented in different languages.

3 Implementation

At the software level, the system was developed using the Unity 3D framework and the C# programming language. Both an Italian and English version were created. At the hardware level, it consisted of an Oculus Quest 2 with the two standard accompanying controllers for the hands. Such hardware was selected for its affordable cost and the standalone and wireless capabilities, which enabled meeting the requirements for ubiquitous use and cost-effectiveness. The controller on the dominant hand was utilized to activate the cut, the other controller was used to navigate the application. An accurate system to track the points hit by the simulated laser on the visualized tissue parts was implemented, along with the computations of the regions correctly or wrongly hit. This system was the basis for the control of the gamification method.

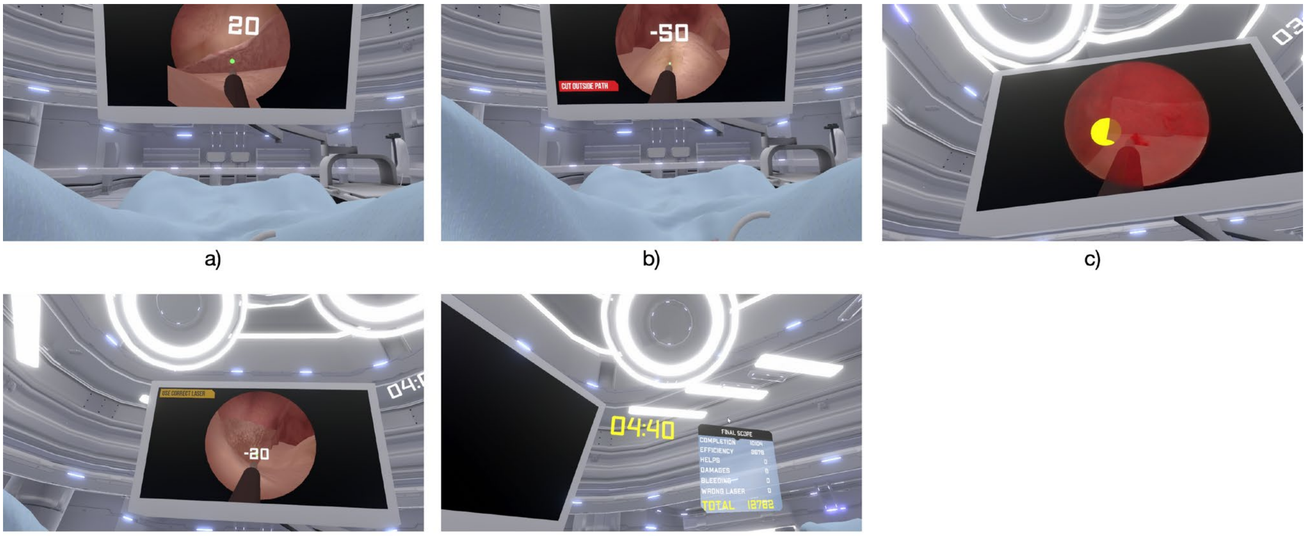
The final design consisted of an application for self-training structured into three parts to be used in sequential order during the training process. The virtual environment represented an operating room with the display of the basic tools used during the prostate enucleation. In particular, a monitor was included, as in the real-world scenario is used to display the video feed from the camera mounted on the laser.

Part 1: Explanation. In this part the user is provided with a guided 3D simulation of the operation, which details the various steps to be performed. A male narrating voice accompanies the various visualizations that the user selects by interacting with a menu. The narrating voice and the visualizations instruct the users on how to perform the actions required by the operation and then ask the user to repeat such actions using the VR hand controllers. The average duration for this phase was designed to last about 20 min.

Part 2: Training with support. This part was devised to assess the learning of the trainee. A gamification approach was included such that the user is immediately informed about the errors made and actions correctly performed. For both correct and wrong actions a score is assigned in real-time. The goal of the serious game is to achieve the highest score. Specifically, an action is considered an error when (i) a wrong laser type is used, (ii) the trainee uses the laser to cut a part of the tissue that is not supposed to be touched, (iii) when the laser is applied for more than 1 s (even in the correct tissue region).

The error notification about the use of the wrong laser type is displayed as a writing ("Use correct laser") and with a negative score (see Fig. 2d). The error notification about the wrong tissue region consists of an audio-visual feedback: at visual level a negative score as well as a writing ("Cut outside path") informs the user that s/he is hitting a wrong tissue region (see Fig. 2b); at auditory level a short alarm sound is provided. A haptic-visual feedback is also produced if no error is made: at visual level it is displayed correct visualization of the tissue opening and the value of the increased score (see Fig. 2a); at haptic level a short continuous vibration is triggered. If the user activates the laser for more than 2 s a visual feedback of bleeding is displayed to inform him/her that s/he is damaging the tissue (see Fig. 2c).

At any time, if users find themselves in difficulty or are blocked on a certain part, they can request an help support, which will show the points where the operation needs to be performed. This will however entail a score decrement. The score is also determined by other aspects, such as the time to complete the operation, the use of the correct laser during different parts of the operation, and the amount of usage of the laser (the less the use, the lower the probability to damage the tissues). Moreover, a penalty is assigned if a bleeding time is too long (a bleeding wound needs to be



FINAL SCORE	
COMPLETION	10104
EFFICIENCY	2678
HELPS	0
DAMAGES	0
BLEEDING	0
WRONG LASER	0
TOTAL	12782

a) b) c) d) e)

Fig. 2 Screenshots of the application displaying feedback to users: **a** correct use; **b** laser application on a wrong tissue region; **c** bleeding as a result of the prolonged use of the laser; **d** use of the wrong laser type; **e** summary of score performances at the conclusion of a training session

closed in reasonable time). In more detail, the total score for a training session is computed using the following formula:

$$\begin{aligned}\text{Total score} = {} & \text{Surgery Execution} \\ & + \text{Efficiency} \\ & - \text{Help} \\ & - \text{Damages} \\ & - \text{Bleedings} \\ & - \text{Wrong Laser Type}\end{aligned} \tag{1}$$

Where

- Surgery Execution $= \alpha * cuts\ done\ on\ colored\ points * combo\ multiplier$, where $combo\ multiplier$ is a variable that has a default value of 1, and that gets constantly increased at each cut performed correctly, and is reset every 5 s (in this way the user is rewarded on the basis of the speed with which performs the cuts);
- Efficiency $= \beta * (1 - (laser\ utilization\ time/game\ time))$, i.e., this score relates to the amount of time in which the laser is used, where the less the utilization the better;
- Help $= \gamma * number\ of\ requested\ help$;
- Damages $= \delta * number\ of\ damages$, i.e., the number of cuts out of target;
- Bleedings $= \epsilon * amount\ of\ bleed\ times$;
- Wrong Laser Type $= \eta * number\ of\ uses\ of\ wrong\ laser\ type$;
- $\alpha = 6,\ \beta = 20,\ \gamma = 80,\ \delta = 50,\ \epsilon = 150,\ \eta = 20$ are constants defined to weight the different contributions of the items above to the final score.

Equation 1, including the exact value for the constants, was defined after an extensive process of trial and error, which encompassed the directions of the surgeons and the feedback of participants involved in the pilot study (see Sect. 4.1).

After a session a user could see his/her performances (see Fig. 2e). We also set in place a reward mechanism such that users can compare their performances in the various training sessions and monitor their improvements. The average duration for this phase was designed to last about 30 min.

Part 3: Training without support. This part was similar to the previous one, with the sole exception that no help support whatsoever is provided. This represents the closest situation to a real operation. The average duration for this phase was designed to last about 30 min. This time is about half of the duration of a traditional training session or the actual operation. Notably for Part 3, Equation 1 did not include the number of requested help, as these were not present.

Notably, in designing the system we carefully considered the twelve tips to harness the power of gamification in medical education reported in (Singhal et al. 2019), as well as the recommendations for designing VR training systems reported in (Abich et al. 2021).

4 User study

4.1 Pilot test

The experiment was preceded by a pilot test. This involved six trainees (2 females, 4 males) aged between 29 and 32 (mean = 31, standard deviation = 1.81) recruited from the Cottolengo Hospital in Turin, Italy. They were all last year residency program students. None of them were involved in the design process of the system. The pilot test allowed us to fine tune the experimental protocol, the scoring system, and data collection procedures, as well as further improve some aspects of the user experience. Nevertheless, taken together the achieved results were generally in line with those reported in the main study described hereinafter.

4.2 Participants

Seventeen trainees (6 females, 11 males) aged between 27 and 31 (mean = 28.7, standard deviation = 1.52) were recruited from the Hospital of Verona, Italy. They were all last year residency program. They were all Italian. All participants reported normal or corrected-to-normal vision and the absence of motor impairments. Participants were blind to the hypothesis of the experiment. None of them was involved in the design process of the system nor in the pilot testing phase. They gave informed consent prior to the start of the study. Fourteen participants reported to have had no previous experience with using VR headsets, while three reported to have had a rather limited experience with VR tools. The experimental procedure, approved by the local ethics committee, was in accordance with the ethical standards of the 1964 Declaration of Helsinki. The experiments were conducted at the premises of the Hospital of Verona.

4.3 Procedure

Participants were given the Oculus Quest 2 VR headset and were asked to use the application running on it for 2 weeks. They were instructed to start from part 1, and proceed with the subsequent part 2 only when they felt they had reached a sufficient level of confidence with part 1. Participants were asked to conduct 5 sessions in both part 2 and part 3. For each session in part 2 and part 3 we recorded the scores related to surgery execution, efficiency, help (only for part 2), damages, bleedings, wrong laser type, total score, as defined in Sect. 3.

After the whole test period was concluded participants were asked to fill in a questionnaire composed by (i) a demographic questionnaire; (ii) a set of questions which in part were based on the technology acceptance model (TAM) (Davis 1989) and on the questionnaire reported in (Chávez et al. 2020), to be evaluated on a 7-point Likert scale [1 = not at all, 7 = very much], which investigated the following dimensions: perceived usefulness, perceived ease of use, satisfaction with the system, and teaching approach (see Figs. 5 and 6); *iii)* the system-usability-scale (SUS) (Brooke 1996); *iv)* an ad-hoc questionnaire of open-ended questions:

- How was your experience in interacting with the system?
- How does the system compare with the traditional teaching method (e.g., operating on cadavers of humans or pigs)?
- What is the added value of the system?
- How would you improve the system?
- Do you have any comment about the system?

4.4 Quantitative results

4.4.1 Scores

Figure 3 illustrates the mean and standard error of the collected scoring metrics for each session in part 2. For each metric, an ANOVA was performed on a linear mixed effect model. These models had the subject as a random factor, and the metric (Surgery Execution, Efficiency, Help, Damages, Bleedings, Wrong Laser Type, Total Score) and session (from 1 to 5) as fixed factors. Post hoc tests were performed on each fitted model using pairwise comparisons adjusted with the Tukey correction. The assumption of normally distributed residuals was visually verified.

Regarding the analysis on surgery execution, a significant main effect was found for factor session (F(4.64) = 8.22, $p < 0.001$). Post hoc tests showed that participants' performances in cutting correctly and quickly were significantly better for session 5 compared to session 1 ($p < 0.001$), 2 ($p < 0.001$) and 3 ($p < 0.05$), and that were significantly better for session 4 compared to session 1 ($p < 0.05$). Concerning Help, a significant main effect was found for factor session (F(4.64) = 3.48, $p < 0.05$). Post hoc tests showed that the number of requested help was significantly lower for session 5 compared to session 1 ($p < 0.05$). As for Bleedings, a significant main effect was found for factor session (F(4.64) = 4, $p < 0.01$). Post hoc tests showed that the amount of bleed times was significantly great for session 1 compared to session 4 and 5 (both $p < 0.05$). Regarding total score, a significant main effect was found for factor session (F(4.64) = 27.48, $p < 0.001$). Post hoc tests showed that participants' overall performances were significantly better for session 5 compared to session 1 ($p < 0.001$), 2 ($p < 0.001$) and 3 ($p < 0.05$), were significantly better for session 4 compared to session 1 and 2 (both $p < 0.001$), as well as were significantly better for session 3 compared to session 2 ($p < 0.05$).

Figure 4 illustrates the mean and standard error of the the collected scoring metrics for each session in part 3. The same analysis conducted for part 2 was performed. Regarding the analysis on Surgery Execution, a significant main effect was found for factor session (F(4.64) = 62.8, $p < 0.001$). Post hoc tests showed that participants' performances in cutting correctly and quickly were significantly better for session 5 compared to session 1 ($p < 0.05$) and 2

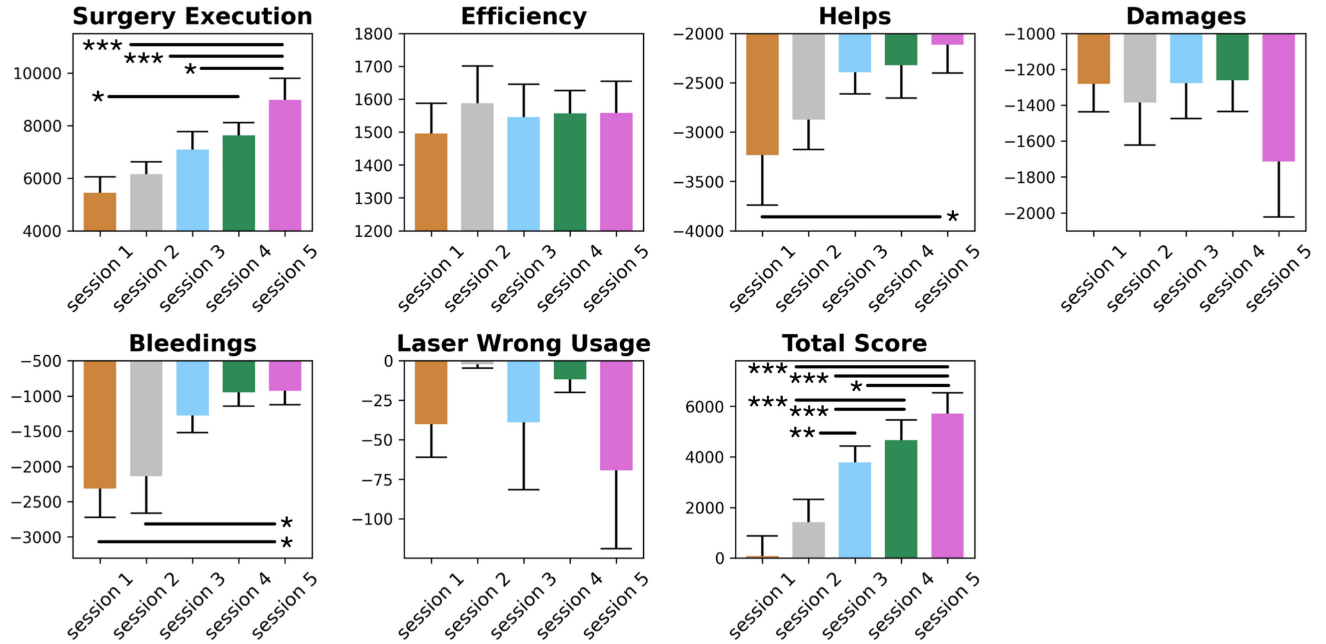
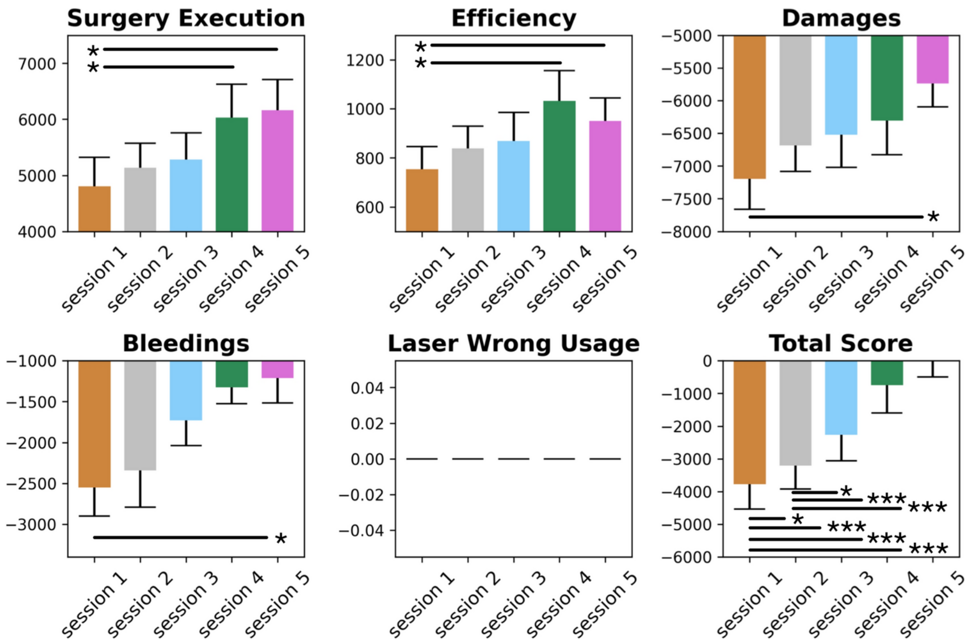


Fig. 3 Mean and standard error of the scoring metrics used in part 2 for each session. * = $p < 0.05$, *** = $p < 0.001$

Fig. 4 Mean and standard error of the scoring metrics used in part 3 for each session. * = $p < 0.05$, *** = $p < 0.001$



($p < 0.05$), and that were significantly better for session 4 compared to session 1 ($p < 0.05$). As for Efficiency, a significant main effect was found for factor session (F(4.64) = 62.82, $p < 0.001$). Post hoc tests showed that participants used the laser significantly more in session 1 compared than in session 4 and 5 (both $p < 0.05$). Concerning Damages, a significant main effect was found for factor session (F(4.64) = 62.81, $p < 0.001$). Post hoc tests showed that the number of cuts out of target was significantly lower in session 5 compared than in session 1 ($p < 0.05$). Concerning, Bleedings, a significant main effect was found for factor session (F(4.64) = 3.02, $p < 0.05$). Post hoc tests showed that the amount of bleed times was significantly great for session 1 compared to session 5 ($p < 0.05$). Regarding Total Score, a significant main effect was found for factor session (F(4.64) = 30.08, $p < 0.001$). Post hoc tests showed that participants' overall performances were significantly better for session 5 compared to session 1 ($p < 0.001$), 2 ($p < 0.001$), 3 ($p < 0.001$) and 4 ($p < 0.05$), and were significantly better for session 4 compared to session 1 ($p < 0.001$), 2 ($p < 0.001$) and 3 (both $p < 0.05$).

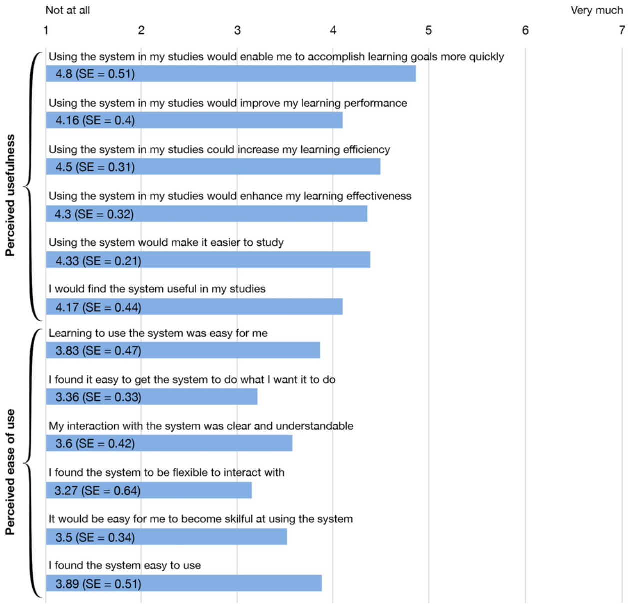


Fig. 5 Mean and standard error of the questionnaire items related to the perceived usefulness and perceived ease of use

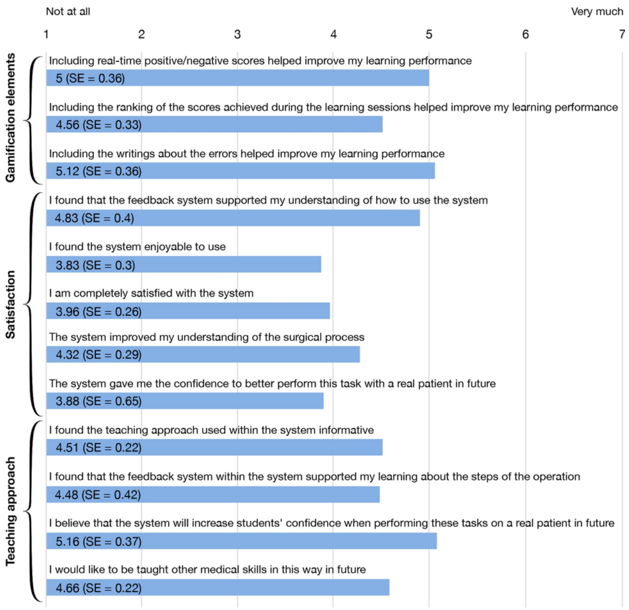


Fig. 6 Mean and standard error of the questionnaire items related to the gamification elements, satisfaction with the system, and teaching approach

4.4.2 Perceived usefulness and ease of use

Figure 5 shows the mean and standard error of the questionnaire items related to the perceived usefulness and perceived ease of use.

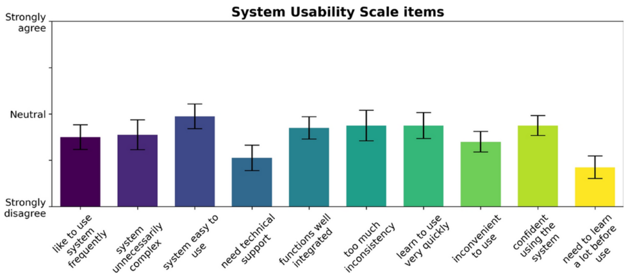


Fig. 7 Mean and standard error of the SUS questionnaire items

4.4.3 Gamification, satisfaction, and teaching

Figure 6 shows the mean and standard error of the questionnaire items related to the gamification elements, satisfaction with the system, and teaching approach.

4.4.4 System usability scale

The SUS metric assesses the usability of a system on a scale from 0 to 100. As a point of comparison, an average SUS score of about 68 was obtained from over 500 studies. The system obtained a mean SUS score of 55.12 (95% confidence interval: [46.29; 63.95]), which is below average. Figure 7 shows the breakdown of the result across the SUS topics. The results indicate that on average, participants did not found the system easy to use without technical support and deemed that it requires significant effort to learn. All other dimensions of the SUS were not above the neutral score.

4.5 Qualitative results

The open-ended questions were analyzed with a reflexive thematic analysis (Braun and Clarke 2019). Through this analysis, conducted by three of the authors, we generated codes that were further organized into the following themes that reflected shared patterns.

Usefulness. Twelve participants commented positively about the usefulness of the system for trainees who are approaching laser enucleation. In particular, the system was deemed to successfully support the learning process of the steps to be conducted, allowing users to repeat them as many times as needed, which is not possible in real-world scenarios (e.g., *"The system felt very useful for understanding the steps of the procedure"; "It is not as realistic as ex vivo models but allows to repeat the task as many times as the user wants, which is very important"; "It still needs improvements to really simulate real-life surgery but provides a good step-by-step learning of the procedure"*). Three participants also commented on the benefit of having a portable system (e.g., *"Thanks to this app I can practice* *at home or wherever and whenever I need"*). Moreover, for six participants an aspect that was found very useful was the scoring system, which allowed for monitoring the learning progresses *"The system is useful because it offers the possibility to provide a score to evaluate the learning process in an objective manner"*). Another aspect that conferred the system with usefulness was the ability of the system to provide immediate feedback to the trainees (e.g., *"The added value of the system is the possibility to have a feedback about the execution of the procedure"; "The simulation of unpredictable bleedings is very useful"*).

Novelty and potential. Four participants reported to have appreciated the novelty and the concept behind the system and to see its potential (e.g., *"The system needs to improve, but I think it could become a great instrument to learn surgery"; "It is a novel concept, but very immature tech. With many major modification it can become a game changer in the long run"*).

Lack of realism. Fifteen participants commented that despite the system being effective and correct in providing the procedure to be learnt, it lacked realism. The main issue concerned the absence of tactile feedback (e.g., *"The steps of the procedure are complete but it needs some improvements in the response to inputs to be more realistic"; "You go through entire structures instead of deforming them as it occurs in a real operation"; "Not very realistic in terms of sensitivity to commands, but allows to learn all the steps of the procedure"*). For nine participants this lack of realism led to a bad experience (e.g., *"Overall my experience was negative because the system was quite far from reproducing reality"; "The inability to physically interact with the fabrics altered my experience with the system"; "Nice idea but for me it is better to work with real tissues"*).

Ethics. Three participants commented that one of the main benefits of the system was its ability to avoid some ethical issues (e.g., *"It is not as realistic as on ex vivo models but definitively overcomes some ethical problems"; "The added value of the system is the possibility to repeat the same procedure in a standardized way without any ethical concern"; "Ethically it is much more acceptable"*).

Improvements requests. Eight participants consistently reported that the system should better support the interactions in the virtual environment, in particular for what concerns the addition of realistic force-feedback and tactile feedback (e.g., *"I suggest to introduce physical interaction between tools and fabrics and eliminate inter-penetrability"*). Relatedly, other five participants requested the ability of the system to support realistic tissue deformations (e.g., *"The most important improvement should be that the instrument can navigate through the tissue, now it cannot and that's not what happens in the reality"*). Four participants suggest improving the visual quality of the scene (e.g., *"The graphics need to be improved so that one really sees a scene like it occurs in a real operation"*). Three participants requested reducing the latency between gestures and their corresponding visual feedback (e.g., *"The system needs some improvements in the speed and precision of response to commands"; "simulator's physics need some improvements, especially in the sensitivity to commands"*).

Cybersickness. Four participants reported to have experienced light symptoms of cybersickness such as nausea and dizziness, although this did not compromise their ability to complete the task. This was in part due to the fact that these participants used for the first time a VR headset (e.g., *"It gave me a bit of headache but nonetheless I kind of enjoyed the experience"*).

5 Discussion

The present study investigated the application of gamification principles to the learning process of surgeons using an immersive VR system, a challenge scarcely addressed thus far. The reason for using gamification was to explore teaching methods alternative to the traditional ones, which could make learning more engaging and motivating. On the other hand, the reasons for using VR were to avoid the typical issues encountered by surgeon trainees in dealing with cadavers of humans or animals, as well as to provide a novel training method that allows for repeating a given operation at will, making errors in a risk-free environment and receiving immediate feedback on the performed actions.

Behavioural data collected during part 2 (which included help) and part 3 (without help) showed the presence of some significant statistical differences between the initial and the final sessions. Specifically, in part 2, a clear and constant improvement trend was observed between the first and last sessions for all metrics except Efficiency, Damages and Laser Wrong Type, while in part 3 all metrics except Laser Wrong Type (which however was zero in all sessions). This is an indication that a learning effect occurred within each part, such that participants could improve their performances based on the gamification techniques set in place. The results of both parts indicate that at least five sessions are needed to show a consistent improvement in the usage of the system and in the participants' performances. However, our results suggest that participants could have conducted more training sessions in part 2 in order to properly transfer to part 3 the skills. Indeed from a comparison between Figs. 3 and 4 it is evident that the usage of help led to better performances compared to those achieved in their absence. Nevertheless, some skills acquired during part 2 transferred to part 3. An indication for this is the absence of any error in part 3 concerning the wrong usage of the laser type, whereas in part 2 participants made some mistakes.

While results on the objective data related to the behavioural performances of participants revealed the usefulness of the adopted gamification approach within the training, the subjective data resulting from the questionnaire responses provided a less positive picture about the actual usability and user experience of users. The responses to the other quantitative items showed that the perceived usefulness was not very high, nor the perceived ease of use (see Fig. 5). In the same vein, the questions related to the gamification elements, satisfaction and teaching approach did not receive on average high scores (see Fig. 6). On average participants judged that the usability of the system was sub-optimal as evident from the responses of the SUS questionnaire (see Fig. 7).

The answers reported to the open-ended questions provide the exact reasons for the relatively low rankings of the quantitative items of the questionnaires. While most participants reported to have appreciated the novelty, concept, usefulness and potential of the tested technology, they also identified major technical and non-technical barriers that hampered a satisfactory experience with the system. First, the realism of the experience was deemed by most participants unsatisfactory. In particular, the absence of force-feedback and tactile feedback properly reflecting real-life situations was found detrimental, along with the lack of visual deformations of the tissue that would occur in the physical world. Moreover, participants commented on the lack of fidelity of the visual rendering in terms of resolution, texture, and colors. Second, a non-technical barrier that could have led to low rankings was the scarce familiarity of participants with VR technology. This is in agreement with other studies that investigated the acceptability of a new technology in relation to technical skills (see e.g., (Venkatesh et al. 2012)). In particular, the study reported in (Bracq et al. 2019) about the acceptability of VR simulators for surgeons training showed that participants who regularly used controllers and/or virtual environments did not have the same attitude towards the VR training simulator as those who were unfamiliar with them. Third, a few participants experienced cybersickness. Although symptoms were not strong, this impacted negatively their experience, which was then reflected in the questionnaire rankings. Whereas technical developments will lead to improvements in the quality of VR headsets in the near future, simulator sickness remains an issue, especially for first-time users.

In their comments, most participants considered the system as useful as a training tool, which can be successfully used to extend conventional learning practices. Nevertheless, an open question is whether participants are successful in transferring the knowledge acquired in VR to real life. This aspect necessitates further investigations, involving longitudinal studies. Notably, the purpose of the proposed VR simulator was not to replace other ways of training students, but rather to assess the validity of the proposed approach in such a way that such training system could be integrated into the curriculum. In general, participants considered VR a promising technology to support learning in a safe and controlled environment, in particular to avoid bureaucratic and ethical issues involved in training with cadavers of humans and animals. Nevertheless, while VR was deemed to have the concrete potential to be an effective enhancement of traditional teaching methods, it was judged to be incapable of substituting the real-world experience of surgical practicing because of the weaknesses related to scarce realism.

The participants' comments about the lack of realism indicate that the expectations from this class of users were very high concerning the quality of experience that a VR training application for surgeons should provide. However, the goal of the system was to support procedural skills training rather than replicating all aspects of real-life operations, in line with the aims set during the co-design phase. Satisfying the requirements of participants would entail a significant engineering effort in terms of exact physics-based rendering which goes well beyond not only the scope of this study but also the capabilities of current technologies with affordable costs (which was one of the target requirements we had established).

Notably, our study has some limitations. First, the sample size (n = 17) was relatively low. Secondly, all participants were Italian. Third, they belonged to one specific specialization school. Fourth, the majority of them were males. Involving a larger pool of participants, from different nationalities, diverse backgrounds from other specialization schools, as well as with greater gender balance, would increase the generalizability of our results. Furthermore, our system did not include any techniques for personalizing the gamification approach, whereas tailored gamification (Klock et al. 2020) is known to lead to performance improvements.

6 Conclusion

This paper described the design, implementation, and evaluation processes of UROVR, a VR application conceived for supporting the procedural skills training of urology surgeons for the enucleation of the adenoma prostate via a laser. The application was co-designed by a team of engineers in collaboration with a set of urology surgeons. This co-designed application adopted the gamification paradigm, which has been scarcely investigated in VR-based medical training thus far. The behavioural results of the user study, conducted with seventeen urology trainees not involved in the co-design sessions, revealed that the system was effective in supporting the learning of the step-by-step procedures involved in the operation. Such findings provide evidence that gamification in VR medical training systems is a valuable strategy to enhance surgical trainees' training outcomes and motivations. Moreover, they suggest that VR can be an effective medium to support training of surgeons. In particular, the power of VR lies in its ability to provide ubiquitous training, even in the absence of a trainer.

On the other hand, the subjective results revealed the system's limitations in delivering a compelling experience in terms of realism. In particular, the lack of effective tactile and force feedback, along with the non-optimal quality of the visual rendering, were deemed the major obstacles to the experience desired by surgeons. Nevertheless, our results also need to be contextualized within the participants' limited familiarity and confidence with VR tools.

An immediate future research direction is to explore via a longitudinal study how learning efficiency and effectiveness evolve with system usage. In future work we also plan to enhance the level of realism of the interaction by investigating novel methods to deliver haptic feedback as well as by providing environmental sounds that typically occur during real operations. Furthermore, we plan to investigate the differences in participants' stress levels across the session types with questionnaires such as NASA-TLX (Hart 2006) or SURG-TLX (Wilson et al. 2011). Finally, we plan to introduce personalization mechanisms where each user could customize the gamification experience.

Acknowledgements This work was supported by Provincia Autonoma di Trento (Legge 6, grant n. PAT/RFS067). The authors would like to thank the developers of company Digital Mosaik, the doctors Cecilia Maria Cracco, Dario Peretti and Carlo Casablanca of Cottolengo Hospital of Turin, as well as the residents at University Hospital of Verona, who contributed to the development of the app and its evaluation.

Author contributions L.T. wrote the main manuscript text and the design of the application. F.G. contributed with the implementation of the application. S.M., A.V. and A.A. contributed to the evaluation of the application. C.S. contributed to the design of the application. All authors reviewed the manuscript.

Funding Open access funding provided by Università degli Studi di Trento within the CRUI-CARE Agreement.

Data availability No datasets were generated or analysed during the current study.

Declarations

Conflict of interest The authors declare no competing interests.

References

Abich J, Parker J, Murphy JS, Eudy M (2021) A review of the evidence for training effectiveness with virtual reality technology. Virtual Real 25(4):919–933

Ahlberg G, Heikkinen T, Iselius L, Leijonmarck C-E, Rutqvist J, Arvidsson D (2002) Does training in a virtual reality simulator improve surgical performance? Surg Endosc Other Interv Tech 16(1):126–129

Bannon L, Bardzell J, Bødker S (2018) Reimagining participatory design. Interactions 26(1):26–32

Barré J, Michelet D, Truchot J, Jolivet E, Recanzone T, Stiti S, Tesnière A, Pourcher G (2019) Virtual reality single-port sleeve gastrectomy training decreases physical and mental workload in novice surgeons: an exploratory study. Obes Surg 29(4):1309–1316

Bracq M-S, Michinov E, Arnaldi B, Caillaud B, Gibaud B, Gouranton V, Jannin P (2019) Learning procedural skills with a virtual reality simulator: an acceptability study. Nurse Educ Today 79:153–160

Braun V, Clarke V (2019) Reflecting on reflexive thematic analysis. Qual Res Sport Exerc Health 11(4):589–597

Brooke J (1996) SUS-A quick and dirty usability scale. Usability Eval Ind 189(194):4–7

Caponetto I, Earp J, Ott M (2014) Gamification and education: a literature review. In: European conference on games based learning, vol 1. Academic Conferences International Limited, p 50

Chávez OL, Rodríguez L-F, Gutierrez-Garcia JO (2020) A comparative case study of 2D, 3D and immersive-virtual-reality applications for healthcare education. Int J Med Inform 141:104226

Checa D, Bustillo A (2020) A review of immersive virtual reality serious games to enhance learning and training. Multimed Tools Appl 79(9):5501–5527

Das AK, Teplitsky S, Humphreys MR et al (2019) Holmium laser enucleation of the prostate (HoLEP): a review and update. Can J Urol 26(4 Suppl 1):13–19

Davis FD (1989) Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Q 13:319–340

Deterding S, Dixon D, Khaled R, Nacke L (2011) From game design elements to gamefulness: defining" gamification". In: Proceedings of the 15th international academic MindTrek conference: envisioning future media environments, pp 9–15

Dicheva D, Dichev C, Agre G, Angelova G (2015) Gamification in education: a systematic mapping study. J Educ Technol Soc 18(3):75–88

Dörner R, Göbel S, Effelsberg W, Wiemeyer J (2016) Serious Games. Springer

EA Araujo S, Delaney CP, Seid VE, Imperiale AR, Bertoncini AB, Nahas SC, Cecconello I (2014) Short-duration virtual reality simulation training positively impacts performance during laparoscopic colectomy in animal model: results of a single-blinded randomized trial. Surg Endosc 28(9):2547–2554

Felszeghy S, Pasonen-Seppänen S, Koskela A, Nieminen P, Härkönen K, Paldanius K, Gabbouj S, Ketola K, Hiltunen M, Lundin M et al (2019) Using online game-based platforms to improve student performance and engagement in histology teaching. BMC Med Educ 19(1):1–11

Fowler C (2015) Virtual reality and learning: Where is the pedagogy? Br J Edu Technol 46(2):412–422

Frederiksen JG, Sørensen SMD, Konge L, Svendsen MBS, Nobel-Jørgensen M, Bjerrum F, Andersen SAW (2020) Cognitive load and performance in immersive virtual reality versus conventional virtual reality simulation training of laparoscopic surgery: a randomized trial. Surg Endosc 34:1244–1252

Freina L, Ott M (2015) A literature review on immersive virtual reality in education: state of the art and perspectives. In: The international scientific conference E-learning and software for education, vol 1, pp 10–1007

Ganni S, Li M, Botden SM, Nayak SR, Ganni BR, Rutkowski A-F, Goossens RH, Jakimowicz J (2020) Virtual operating room simulation setup (VORSS) for procedural training in minimally invasive surgery—a pilot study. Indian J Surg 82:810–816

Gómez Rivas J, Rivero Belenchón I, Li N, Somani B, Andres Esteban E, Cacciamani G, Checcucci E, Puliatti S, Taratkin M, Kowalewski K-F et al (2023) A new era and future of education: the impact of pandemic on online learning-a study from the European school of urology. Cent Eur J Urol 76(3):256–262

Gorbanev I, Agudelo-Londoño S, González RA, Cortes A, Pomares A, Delgadillo V, Yepes FJ, Muñoz Ó (2018) A systematic review of serious games in medical education: quality of evidence and pedagogical strategy. Med Educ Online 23(1):1438718

Hamacher A, Kim SJ, Cho ST, Pardeshi S, Lee SH, Eun S-J, Whangbo TK (2016) Application of virtual, augmented, and mixed reality to urology. Int Neurourol J 20(3):172

Hamacher A, Whangbo TK, Kim SJ, Chung KJ (2018) Virtual reality and simulation for progressive treatments in urology. Int Neurourol J 22(3):151

Hart SG (2006) NASA-task load index (NASA-TLX); 20 years later. In: Proceedings of the human factors and ergonomics society annual meeting, vol 50. Sage publications Sage CA, Los Angeles, pp 904–908

Howard MC, Gutworth MB (2020) A meta-analysis of virtual reality training programs for social skill development. Comput Educ 144:103707

Howard MC, Gutworth MB, Jacobs RR (2021) A meta-analysis of virtual reality training programs. Comput Hum Behav 121:106808

Hu R, Wu Y-Y, Shieh C-J (2016) Effects of virtual reality integrated creative thinking instruction on students' creative thinking abilities. Eurasia J Math Sci Technol Educ 12(3):477–486

Jin C, Dai L, Wang T (2021) The application of virtual reality in the training of laparoscopic surgery: a systematic review and meta-analysis. Int J Surg 87:105859

Kern F, Winter C, Gall D, Käthner I, Pauli P, Latoschik ME (2019) Immersive virtual reality and gamification within procedurally generated environments to increase motivation during gait rehabilitation. In: IEEE conference on virtual reality and 3D user interfaces. IEEE, pp 500–509

Kiryakova G, Angelova N, Yordanova L (2014) Gamification in education. In: Proceedings of 9th international balkan education and science conference, pp 679–684

Klock ACT, Gasparini I, Pimenta MS, Hamari J (2020) Tailored gamification: a review of literature. Int J Hum Comput Stud 144:102495

Krath J, Schürmann L, Von Korflesch HF (2021) Revealing the theoretical basis of gamification: a systematic review and analysis of theory in research on gamification, serious games and game-based learning. Comput Hum Behav 125:106963

Krishnamurthy K, Selvaraj N, Gupta P, Cyriac B, Dhurairaj P, Abdullah A, Krishnapillai A, Lugova H, Haque M, Xie S et al (2022) Benefits of gamification in medical education. Clin Anat 35(6):795–807

Lenz F, Fock M, Kramer T, Petersen Y, Teistler M (2021) A virtual reality game to support visuospatial understanding of medical X-ray imaging. In: IEEE 9th international conference on serious games and applications for health. IEEE, pp 1–6

Li S, Cui J, Hao A, Zhang S, Zhao Q (2021) Design and evaluation of personalized percutaneous coronary intervention surgery simulation system. IEEE Trans Visual Comput Gr 27(11):4150–4160

Mao RQ, Lan L, Kay J, Lohre R, Ayeni OR, Goel DP et al (2021) Immersive virtual reality for surgical training: a systematic review. J Surg Res 268:40–58

McCoy L, Lewis JH, Dalton D (2016) Gamification and multimedia for medical education: a landscape review. J Am Osteopath Assoc 116(1):22–34

Mehrfard A, Fotouhi J, Taylor G, Forster T, Armand M, Navab N, Fuerst B (2021) Virtual reality technologies for clinical education: evaluation metrics and comparative analysis. Comput Methods Biomech Biomed Eng Imag Vis 9(3):233–242

Nah FF-H, Zeng Q, Telaprolu VR, Ayyappa AP, Eschenbrenner B (2014) Gamification of education: a review of literature. In: International conference on Hci in business. Springer, pp 401–409

Pérez-Escamirosa F, García-Cabra DA, Ortiz-Hernández JR, Montoya-Alvarez S, Ruíz-Vereo EA, Ordorica-Flores RM, Minor-Martínez A, Tapia-Jurado J (2023) Face, content, and construct validity of the virtual immersive operating room simulator for training laparoscopic procedures. Surg Endosc 37(4):2885–2896

Schuller MC, DaRosa DA, Crandall ML (2015) Using just-in-time teaching and peer instruction in a residency program's core curriculum: enhancing satisfaction, engagement, and retention. Acad Med 90(3):384–391

Scoffone CM, Cracco CM (2016) The en-bloc no-touch holmium laser enucleation of the prostate (HoLEP) technique. World J Urol 34(8):1175–1181

Seaborn K, Fels DI (2015) Gamification in theory and action: a survey. Int J Hum Comput Stud 74:14–31

Singhal S, Hough J, Cripps D (2019) Twelve tips for incorporating gamification into medical education. MedEdPublish 8(216):216

Thomsen ASS, Bach-Holm D, Kjærbo H, Højgaard-Olsen K, Subhi Y, Saleh GM, Park YS, La Cour M, Konge L (2017) Operating room performance improves after proficiency-based virtual reality cataract surgery training. Ophthalmology 124(4):524–531

Tuccio A, Sessa F, Campi R, Grosso AA, Viola L, Muto G, Scoffone C, Cracco CM, Gómez-Sancha F, Misrai V et al (2020) En-bloc endoscopic enucleation of the prostate: a systematic review of the literature. Minerva Urol Nefrol Ital J Urol Nephrol 72(3):292–312

Venkatesh V, Thong JY, Xu X (2012) Consumer acceptance and use of information technology: extending the unified theory of acceptance and use of technology. MIS Q 36:157–178

Wilson MR, Poolton JM, Malhotra N, Ngo K, Bright E, Masters RS (2011) Development and validation of a surgical workload measure: the surgery task load index (SURG-TLX). World J Surg 35:1961–1969

Wilson AS, O'Connor J, Taylor L, Carruthers D (2017) A case study into the use of virtual reality and gamification in ophthalmology training. In: Joint international conference on serious games. Springer, pp 158–169

Yang S-Y, Oh Y-H (2022) The effects of neonatal resuscitation gamification program using immersive virtual reality: a quasi-experimental study. Nurse Educ Today 117:105464

Yu P, Pan J, Qin H, Hao A, Wang H (2020) Real-time suturing simulation for virtual reality medical training. Comput Anim Virtual Worlds 31(4–5):1940

Zainuddin Z, Chu SKW, Shujahat M, Perera CJ (2020) The impact of gamification on learning and instruction: a systematic review of empirical evidence. Educ Res Rev 30:100326

Zhao G, Fan M, Yuan Y, Zhao F, Huang H (2021) The comparison of teaching efficiency between virtual reality and traditional education in medical education: a systematic review and meta-analysis. Ann Transl Med 9(3):252