

Proficiency From Immersion

A Human-Centered Design in Cross-Cultural Surgical Training

Li, Meng; Ganni, Sandeep; Albayrak, Armagan; Rutkowski, Anne F.; van Eijk, Daan; Jakimowicz, Jack

DOI

[10.3389/frvir.2021.675334](https://doi.org/10.3389/frvir.2021.675334)

Publication date

2021

Document Version

Final published version

Published in

Frontiers in Virtual Reality

Citation (APA)

Li, M., Ganni, S., Albayrak, A., Rutkowski, A. F., van Eijk, D., & Jakimowicz, J. (2021). Proficiency From Immersion: A Human-Centered Design in Cross-Cultural Surgical Training. *Frontiers in Virtual Reality*, 2, Article 675334. <https://doi.org/10.3389/frvir.2021.675334>

Important note

To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Copyright

Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy

Please contact us and provide details if you believe this document breaches copyrights.
We will remove access to the work immediately and investigate your claim.



Proficiency From Immersion: A Human-Centered Design in Cross-Cultural Surgical Training

Meng Li^{1,2*}, Sandeep Ganni³, Armagan Albayrak¹, Anne F. Rutkowski⁴, Daan van Eijk¹ and Jack Jakimowicz⁵

¹Human-Centered Design Department, Industrial Design Engineering Faculty, Delft Technology of University, Delft, Netherlands, ²Industrial Design Department, Mechanical Engineering School, Xi'an Jiaotong University, Xi'an, China, ³GSL Medical College, Rajahmundry, India, ⁴Department Management, Tilburg School of Economics and Management, Tilburg University, Tilburg, Netherlands, ⁵Skills Lab, Catharine Hospital, Eindhoven, Netherlands

OPEN ACCESS

Edited by:

Lap-Fai Yu,
George Mason University,
United States

Reviewed by:

Frederic Merienne,
ParisTech École Nationale Supérieure
d'Arts et Métiers, France
Mark Dennison,
United States Army Research
Laboratory, United States

*Correspondence:

Meng Li
m.li-4@tudelft.nl

Specialty section:

This article was submitted to
Virtual Reality in Industry,
a section of the journal
Frontiers in Virtual Reality

Received: 02 March 2021

Accepted: 14 May 2021

Published: 09 June 2021

Citation:

Li M, Ganni S, Albayrak A,
Rutkowski AF, van Eijk D and
Jakimowicz J (2021) Proficiency From
Immersion: A Human-Centered Design
in Cross-Cultural Surgical Training.
Front. Virtual Real. 2:675334.
doi: 10.3389/frvir.2021.675334

Ensuring surgeons are well-trained in various skills is of paramount importance to patient safety. Surgical simulators were introduced to laparoscopy training during the last 2 decades for basic skills training. The main drawback of current simulation-based laparoscopy training is their lack of true representation of the intro-operative experience. To create a complete surgical surrounding, the required amount of resources is demanding. Moreover, organizing immersive training with surgical teams burdens daily clinical routines. High-end virtual reality (VR) headsets bring an opportunity to generate an immersive virtual OR with accessible and affordable expenses. Pilot studies reveal that personalization and localization are key needs of the virtual operating room (VOR). They are therefore key in this study. The focus of this study was to explore the effect of different human factors, such as domain knowledge, culture, and familiarity of VR technologies, on the perception of VOR experience. A human-centered design approach was applied to investigate the presence and usability of a VOR. Sixty-four surgical practitioners joined the study in the Netherlands and India. The surgeons were referred to as “experts” and surgical trainees as “novices.” The VOR system we used is composed of a laparoscopic simulator, a graphic virtual OR surrounding, and an Oculus Rift VR headset. Participants conducted the “complete Lapchol” task with the VOR. Afterward, four questionnaires were used to collect subjective ratings on presence and usability. Participant’s qualitative feedback was collected using a semi-structural interview as the final stage. Results showed the surgical knowledge only affected perceived mental demand when using a VOR. The cultural difference would alter the rating on the majority of items in these questionnaires. VR experience mainly affected the judgment on presence including “quality of interface” and “reversible actions.” The interaction effects between surgical knowledge either with culture difference or with VR experience were obvious. This study demonstrated the influences of cultural differences on the perception of immersion and usability. Integrating immersive technologies such as virtual reality and augmented reality to human-centered design opens a brand new horizon for health care and similar professional training.

Keywords: immersive training, virtual reality, cultural difference, intraoperative distractors, human-centered design

INTRODUCTION

Ensuring surgeons are well trained in different kinds of skills is of paramount importance to patient safety. Despite the benefits on saving costs for healthcare systems and improving patient's well-being, mastering laparoscopy (also known as minimally invasive surgery, keyhole surgery, or microsurgery) challenges the limitations of training budget and duration, as well as trainees' mental and physical capabilities (Berguer et al., 2003, 968; Berguer et al., 2001, 1205–1206). It takes 60 months for a resident to become a surgeon (PRISMA health, 2021). Simulation-based training is at the cornerstone of learning demanding tasks such as piloting and driving as it allows immersive visualization and replicates real-world scenarios (Strachan, 2000). Surgical simulators were introduced to the laparoscopy training during the last 2 decades, which effectively helps the acquirement of basic laparoscopic skills, such as eye-hand dexterity and surgical procedures (Seymour et al., 2002, 460–462; Munz et al., 2004, 491–494; Schijven et al., 2005, 1222–1225).

In a real operating room (OR), numerous distractions are occurring during operations, which increase the task demand and stress level of the surgeons (Wiegmann, 2007, 660–662). Surgeons have to demonstrate high dexterity, concurrently appraise the intro-operative situation, and control the surgical flow to avoid adverse events as well (Henrickson Parker, 2010, 356–358). Surgeons are thus required to not only remember the proper sequence of a given procedure but also avoid distractions while conversing with the teams. Patient safety has been proven to be negatively impacted when surgeons are inadequately trained to use complex technology or perform new procedures with long learning curves. Such tasks are particularly taxing on the surgeon's resources. The surgical profession is one of the most stressful occupations. While surgeons are generally in a healthy state, it has been demonstrated that the long hours of work, as well as ongoing disruptions, significantly drained out their physiological and mental resources (Pluyter et al., 2010, 903–905). Training the awareness of impact from these factors is increasing in the field of surgery (Taekman and Shelley, 2010, 111–114).

The main drawback of current simulation-based laparoscopy training is its lack of true representation of the intro-operative experience (Jakimowicz and Buzink, 2015, 28). Most laparoscopic simulators replicate surgical tasks in a 2D display without the environments containing busy and often chaotic operating theatres. To create a complete surgical surrounding, the required amount of spatial, financial, personnel, and technological resources is demanding (Badash et al., 2016, 453; Jakimowicz and Buzink, 2015, 28). Besides, organizing team training burdens the already busy daily clinical routines (Van de Ven et al., 2017, 133–136). Since the upsurge of high-end VR headsets in 2016, it became accessible and affordable to virtually generate an immersive environment of an OR. Different studies highlighted that the presence of the immersive virtual environments, the “being there” effect, brings new opportunities to turn the impossible into possible for learning and training fields (Bowman and McMahan, 2007, 38). Despite

TABLE 1 | Participant information.

	Surgical knowledge		VR experience			Total
	Experts	Novices	Yes	None	NA	
Dutch	8	28	12	17	7	36
India	13	15	15	13	0	28
Total	21	43	27	30	7	64

heightened motivation and engagement of surgical trainees, immersive environments are necessary but not sufficient (Li et al., 2020, 568–570; Ganni et al., 2020, 4–5). Pilot studies revealed that personalization and localization are the key needs of surgical trainees.

Human-centered design is a well-known approach to develop safe, easy-of-use, and affordable products and services for the end-users, especially in the healthcare field (Bowman et al., 2002). However, the majority of current user studies or user cases were done with the WEIRD population (Western, educated, industrialized, rich, and democratic), and then generalized the results applying to human beings in general (Henrich et al., 2010, 61; Nisbett, 2004, 45–48). In product and service development, lacking careful consideration of local culture leads to market failures or even fatal incidents (Hao, 2019, 45–46). Culture alters the way users perceive, understand, and communicate with the surrounding individuals, the local communities, and the world. According to the cultural psychologist Henrich, one has to differentiate people from industrialized western countries, like the Dutch, and the more traditional societies of countries like India regarding how culture shapes their cognition: the western mindset (WEIRD) is more individualist, concerned with universal values, and focused on abstract thinking; in contrast, peoples' mindsets in traditional cultures are more collectivist, concerned with particularistic values, and stressing holistic thinking (Henrich, 2020, 34–38, 40–45).

Several studies showed that cultural differences do exist under several virtual environments (Hornbeck and Barrett, 2013, 23–26; Lin et al., 2020, 10–11). The authors hence focused on the following research questions in this study:

RQ: Are there differences in the perception of the VOR system in the sense of immersion and usability on the following factors:

Sub-Q1: Level of expertise of the surgical knowledge and skills.

Sub-Q2: The adaptability to the trainee's culture.

Sub-Q3: With or without experience on VR technologies.

METHOD

Participants

Sixty-four surgical practitioners enrolled as participants from Catharina Hospital, Eindhoven, the Netherlands, and GSL Medical College, Rajahmundry, India, from June 2018 to February 2019. Among them, twenty-one were surgeons and forty-three were surgical trainees. In this article, the surgeons were referred to as “experts,” while the surgical trainees as “novices.” There were thirty-nine males and twenty-five

TABLE 2 | The heuristic principles and their sub-principles (“consistent and standardized” as an example, and the full version is in **Supplementary Material**).

Heuristics	Sub-principles
The system is consistent and standardized	a. Sequences of actions (skill acquisition) b. Color (for categorization) c. Layout and position (spatial consistency) d. Font and capitalization (levels of the organization) e. Terms (e.g., delete, del, and remove) and language (words and phrases) f. Interaction rules (e.g., for unvisited hyperlinks) g. Touch (e.g., the textures, force, and movement)

females at the average age of 33 years ($SD = 8$). 68.8% of the participants had experience in box surgical trainers, while 71.9% of them had experience with laparoscopic simulators. Twenty-seven participants (42.2%) have used low-cost or high-end VR or AR devices (**Table 1**). Every participant took part in the study voluntarily and was given informed consent. The study was approved by the ethical committees both from the Catharina Hospital and the GSL Medical College.

Materials Measurements

In this study, the factors influencing the perception of presence and immersion were investigated *via* 2 questionnaires. The Presence Questionnaire (PQ), a well-known presence assessment scale, was modified on sound and haptic aspects. Previous studies validated PQ except for “haptic” and “sound” factors (Witmer and Singer, 1998, 235–236; Witmer et al., 2005, 308–310). In this study, we added two items (i.e., accuracy of gestures and realistic resistance of tissue) on “haptic” and one item on “sound” (realistic sound effect) according to the features of the VOR, and applied an extended 7-point scale (1 = not at all and 7 = completely) to survey the level of immersion in fine gradients. A scale was developed based on the fourteen heuristic principles for medical devices (Zhang et al., 2003, 25–26). An example of these heuristics is shown in **Table 2**. Participants used these principles as guidelines to rate their experience with a 5-point scale at the system level, in which one means a low level of realism and five is high realism.

The usability of the VOR was evaluated with a combination of two questionnaires. First, intuitiveness, in other words subconsciously applying prior knowledge, was evaluated *via* the Questionnaire for Intuitive Use (QUESI) (Naumann and Hurtienne, 2010, 401–402). The QUESI was applied across multiple professions, including health care, to quantify the intuitiveness of virtual environments (Saalfeld et al., 2015, 147–149; Li et al., 2018, 304–306). The validated assessment asked if the VOR appears intuitive and satisfying using a 5-point Likert scale (1 = fully disagree and 5 = fully agree). Second, the mental workload of performing the task in the VOR was measured using the NASA-TLX (Hart, 2006, 906–908). This validated tool has been extensively used for assessing the task demand of surgeons when performing laparoscopic surgeries or training (Zhang et al., 2012, 2746–2747; Lee et al., 2014, 458–460). The participants gave a score to the levels of mental, physical, and temporal demands they perceived, as well as their effort, performance, and frustration during the task. The Raw Task

Load Index (RTLX) and subscales were calculated into a score between 0 and 100 (0 = low and 100 = high) (Hart, 2006, 906–908).

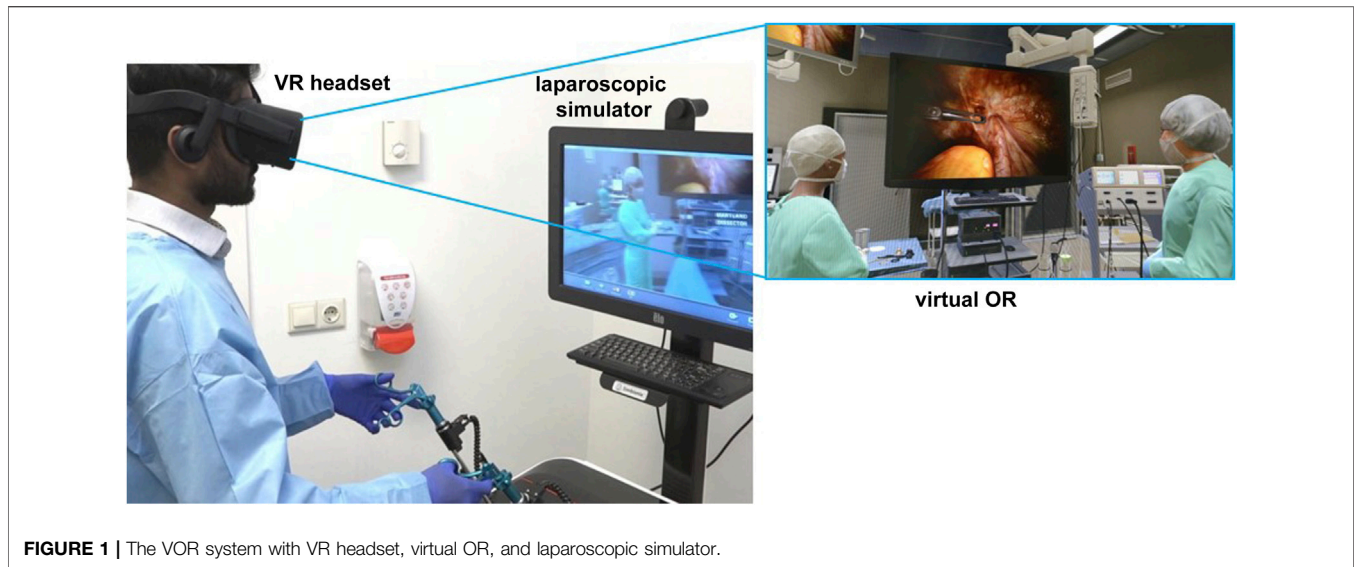
Participants reflected their personal experience of the VOR with two questions: 1) How satisfied are you with the VOR experience? 2) Which factors were not compelling or not realistic in the VOR experience?

Setup

The VOR system we applied comprised three components: a VR headset, graphic virtual OR surrounding, and a VR laparoscopic simulator. This system is the mainstream of commercially available immersive laparoscopy training platforms. Hence, the researchers chose it as the object for this evaluation.

The system contained an Oculus Rift VR headset, providing stereoscopic images (1080 * 1200 per eye and 110° field of view), 3D audio, and 6 DOF head tracking. The virtual OR surrounding was a graphically generated virtual reality application that replicates typical laparoscopic ORs in Western countries, including a full setup of instruments and equipment, a surgical team, a patient, and various distractions. The simulated distraction covered the typical types, such as door openings, phones/pagers/beepers, radio, as well as case-related communication (van Houwelingen et al., 2019, 4527–4531). The simulated auditory distraction and communication distraction are all in English.

LapMentor VR™ (Symbionix, 3D Systems Corporation, United States) with MentorLearn software was applied as the laparoscopic simulator. It contains two integrated modules: 1) the interface module replicated an operation table including a patient’s abdomen, two trocars and handholds, a camera, and a double foot switch. These handholds quipped five DOFs and haptic feedback. The foot switch could activate electrosurgical functions. The camera could be frozen, allowing trainees to finish the training alone. The height of the interface module is adjustable from 62.99” (1.60 m) to 70.86”(1.80 m). 2) The processing module contained a two-unit industrial PC with a 24” touch-screen monitor (1920*1080 dpi): 1) the simulation unit is a 3.1-GHz Intel Core i7-4770S and an Intel™ motherboard; 2) the VOR unit is an NVIDIA GeForce GTX 1060 graphic card and an Intel™ SHARKBAY motherboard. Both units run on Windows 7 Professional (x64) operating system. The MentorLearn software includes a basic skill trainer and a procedural skill trainer. The basic skill modules allow trainees to practice tasks for basic psychomotor abilities. The procedural skill module simulated an entire procedure of laparoscopic



cholecystectomy. The trainee could see a computer-generated cavity from virtual patients through the monitor.

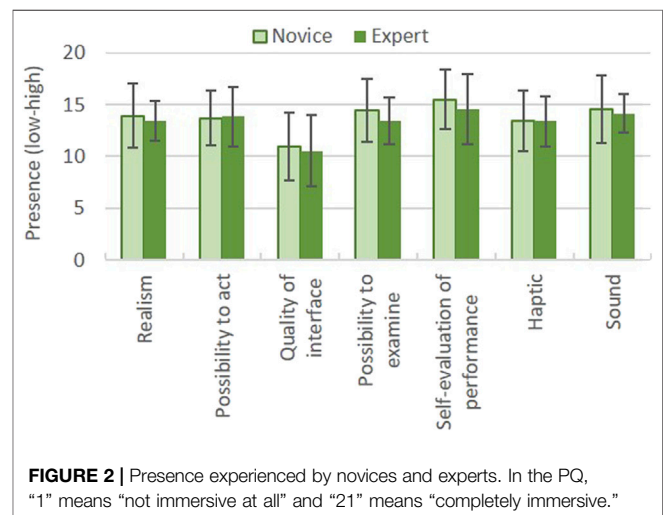
The VOR system displays a graphic virtual OR surrounding the simulator *via* the VR headset (**Figure 1**). If the trainee changes a tool in the VOR, there are several differences from the LapMentor: 1) the tool menu is floating at the eye level; 2) turn a knot at the front of the handle to choose tools instead of pulling the instrument. To simulate the electrosurgical coagulation, a foot switch is displayed underneath the simulated monitor. A video from Simbionix.com demonstrated the interactions with the VOR system (<https://simbionix.com/simulators/lap-mentor/lap-mentor-vr-or/>).

Procedure

A protocol was developed for the experiment, starting with a standard introduction to the objective of the study and on the VOR system. The participants then filled in their informed consent. During the experiment, the participants could explore the virtual OR freely with the LapChol task “LapMentor VR: complete cholecystectomy” for a maximum of 15 min to control the symptoms of simulation sickness. After the hands-on session, participants filled in the questionnaires on the presence and usability of the VOR experience. In the end, the participants were interviewed to collect their qualitative feedback.

Data Analysis

Mean and standard deviation (SD) were calculated with SPSS v.25 for each questionnaire. To compare different groups in novice and expert surgeons, Euro-Asian cultures, and with or without VR experience, a two-step process was applied: 1) tests of normality was conducted using the Kolmogorov–Smirnova test at a significant level of 0.05; 2) when the significant value was less than or equaled to 0.05, indicating a non-nominal distribution of the data, the Mann–Whitney U test (MWW) was utilized to compare the two groups; if the value was larger than 0.05, the classical independent sample *t* test was applied. Then the two-way



ANOVA was used to indicate the main effects and interactive effect among surgical experience, culture, and VR experience on the presence, realism, mental workload, and intuitiveness *via* interaction plot, interaction effect value, and the significance. A *p* value of < 0.05 was considered statistically significant (*), while $p < 0.01$ indicates statistical moderate significance (**), and $p < 0.001$ indicates statistical high significance (***)

RESULTS

Sub-Q1: Level of Expertise of the Surgical Knowledge and Skills Immersion: Presence Questionnaire and Heuristic Scale

In general, both novices and experts experienced moderate presence (PQ mean: total = 13.64, SD = 2.90) with the VOR

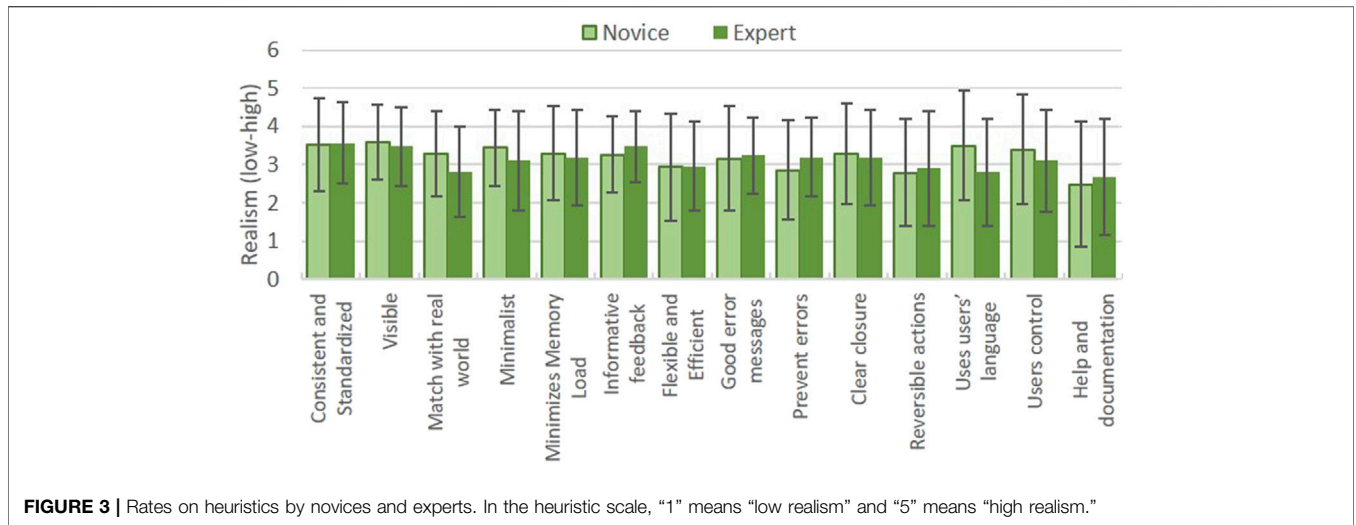


FIGURE 3 | Rates on heuristics by novices and experts. In the heuristic scale, “1” means “low realism” and “5” means “high realism.”

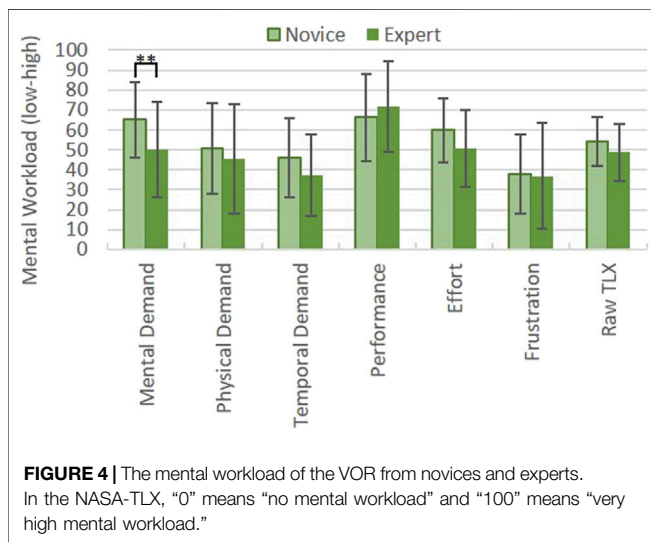


FIGURE 4 | The mental workload of the VOR from novices and experts. In the NASA-TLX, “0” means “no mental workload” and “100” means “very high mental workload.”

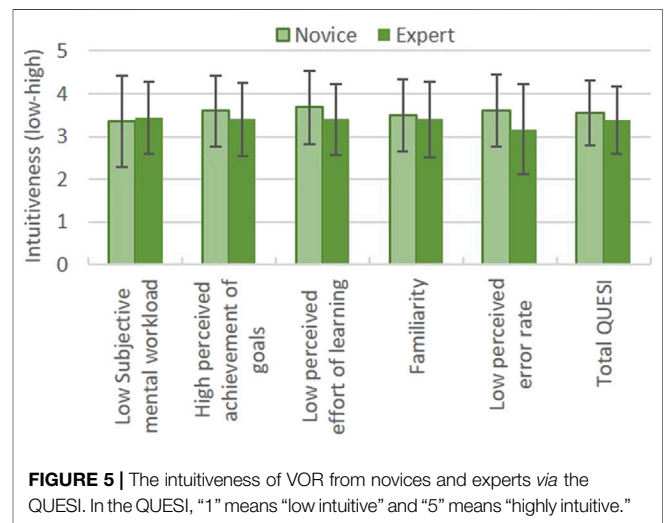


FIGURE 5 | The intuitiveness of VOR from novices and experts via the QUESI. In the QUESI, “1” means “low intuitive” and “5” means “highly intuitive.”

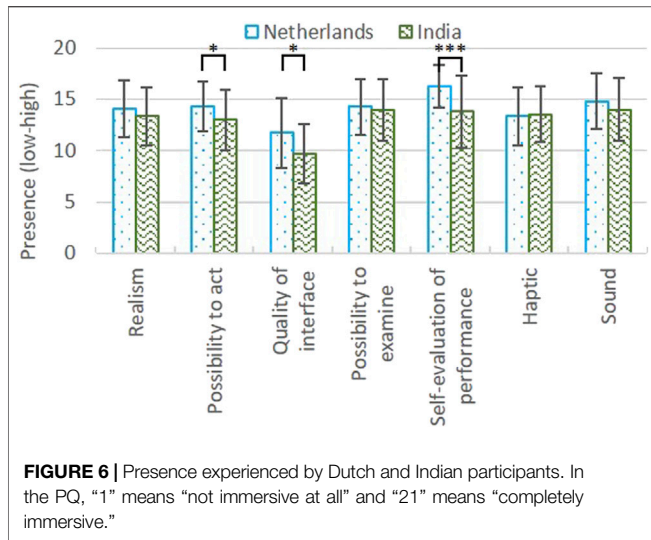
system except for the “quality of interface” (novices = 10.97, SD = 3.28; experts = 10.54, SD = 3.45). The novices were rated between 13.45 (2.90) and 15.50 (2.90), while experts ranged between 13.38 (2.40) and 14.55 (3.38). There were no significant differences between them. The “self-evaluation of performance” (novices = 15.50, SD = 2.90; experts = 14.55, SD = 3.38) and “sound” (novices = 14.56, SD = 3.26; experts = 14.14, SD = 1.89) provided the highest presence (Figure 2).

According to the heuristics shown in Figure 3, the VOR was properly designed (HE means: novices = 3.20, SD = 1.28; experts = 3.12, SD = 1.21), except for “flexible and efficient” (novices = 2.93, SD = 1.40; experts = 2.95, SD = 1.16), “reversible actions” (novices = 2.79, SD = 1.41; experts = 2.90, SD = 1.51), and “help and documentation” (novices = 2.49, SD = 1.64; experts = 2.67, SD = 1.53). There were no significant differences of the novices and experts on the perception of realism.

Usability: NASA-TLX and QUESI

The raw TLX showed that the novice’s mental workload of using the VOR around the mid-point (50) both for the novices and experts (RTLX: novices = 54.28, SD = 12.24; experts = 48.65, SD = 14.40) (Figure 4). The “performance” (total = 67.97, SD = 22.28) associating with nonsuccess tasks was the subscale with the highest mental workload, especially for the experts (experts = 71.67, SD = 22.99). The second high subscale “mental demand” was significantly higher for the novices than the experts (novices = 65.12, SD = 18.95; experts = 50.24, SD = 23.79). On the contrary, the perception of “frustration” (total = 37.58, SD = 22.20) was the lowest.

Both novices and experts thought the VOR was intuitive to use (total QUESI: novices = 3.54, SD = 0.77; experts = 3.37, SD = 0.79) (Figure 5). The “low perceived effort of learning” (Dutch = 3.68, SD = 0.86; Indian = 3.40, SD = 0.83), “low perceived error rate” (novices = 3.60, SD = 0.86; experts = 3.17, SD = 1.06), and “high perceived achievement of goals” (novices = 3.60, SD = 0.84; experts = 3.40, SD = 0.86) were the most intuitive factors for VOR experience.



Sub-Q2: The Adaptability to the Trainee’s Culture

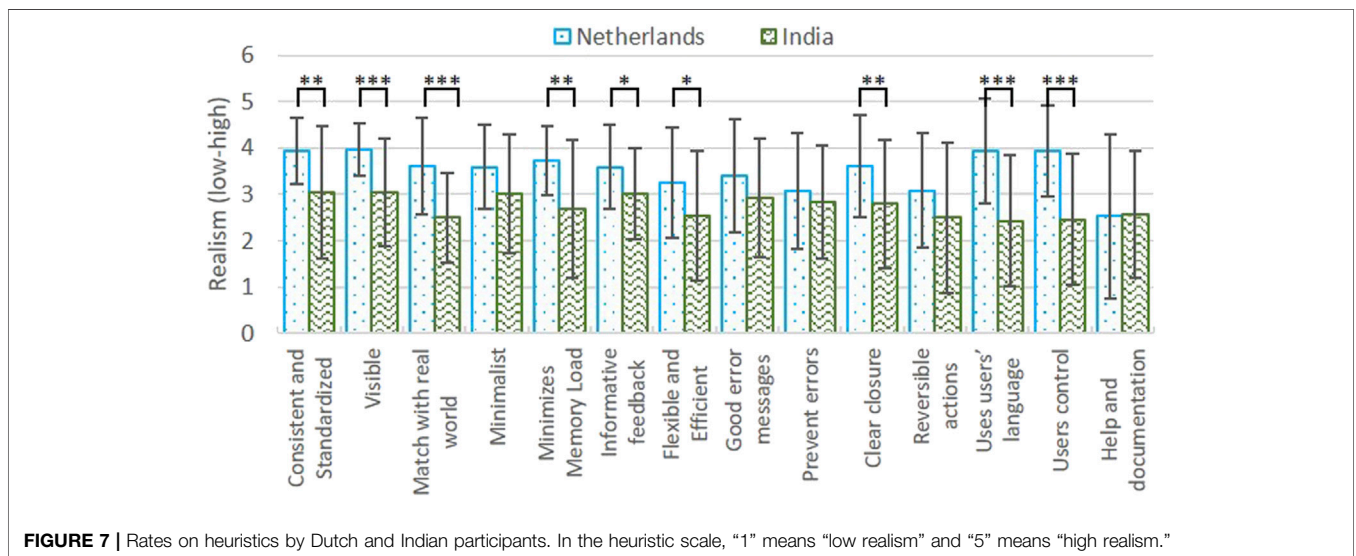
Immersion: Presence Questionnaire and Heuristic Scale

The perception of immersion significantly differentiated across the Indian and the Dutch groups (PQ means: Indian = 13.03, SD = 2.99; Dutch = 14.11, SD = 2.70). For the Indian participants, the degree of immersion was evenly distributed on every aspect of presence except for “quality of interface” (Indians = 9.70, SD = 2.91); in contrast, the Dutch participants attributed a higher degree of immersion to “self-evaluation of performance” (Dutch = 16.28, SD = 2.10) (Figure 6). The Indian participants rated “possibility to act” (Indians = 12.98, SD = 2.97; Dutch = 14.33, SD = 2.40) and “quality of interface” (Indians = 9.70, SD = 2.91; Dutch = 11.70, SD = 3.38) significantly lower than the Dutch. The difference in “self-evaluation of performance” (Indians = 13.79, SD = 3.56; Dutch = 16.28, SD = 2.10) was highly significant.

The majority of the fourteen heuristics were rated lower by the Indian participants than the Dutch (HE means: Indian = 2.74, SD = 1.31; Dutch = 3.51, SD = 1.05), and they significantly differentiated on nine items (Figure 7). Among these heuristics, four items were highly significant, including “visible” (Dutch = 3.97, SD = 0.56; Indian = 3.04, SD = 1.17), “match with the real world” (Dutch = 3.61, SD = 1.05; Indian = 2.50, SD = 0.96), “use user’s language” (Dutch = 3.92, SD = 1.13; Indian = 2.43, SD = 1.40), and “user’s control” (Dutch = 3.94, SD = 0.98; Indian = 2.46, SD = 1.43). The differences were modestly significant between Dutch and Indian in “consistent and standardized” (Dutch = 3.94, SD = 0.72; Indian = 3.04, SD = 1.43), “minimizes memory load” (Dutch = 3.72, SD = 0.74; Indian = 2.68, SD = 1.49), and “clear closure” (Dutch = 3.61, SD = 1.10; Indian = 2.79, SD = 1.37). The items related to system efficiency such as “informative feedback” (Dutch = 3.58, SD = 0.91; Indian = 3.00, SD = 0.98) and “flexible and efficient” (Dutch = 3.25, SD = 1.18; Indian = 2.54, SD = 1.40) were significantly lower for the Indian participants than the Dutch.

Usability: NASA-TLX and QUESI

Figure 8 showed that the Indian participants had a significantly higher mental workload than the Dutch counterparts (RTLX: Dutch = 48.50, SD = 11.62; Indian = 57.50, SD = 13.45). The subscales “performance” (Dutch = 68.19, SD = 21.32; Indian = 67.68, SD = 23.66), “mental demand” (Dutch = 56.39, SD = 21.17; Indian = 65.18, SD = 21.62), and “effort” (Dutch = 56.39, SD = 17.91; Indian = 57.32, SD = 17.61) were the main mental workload resources of the VOR system indifferently for Indian and the Dutch. Dutch participants had significantly lower mental workload than Indian participants on these subscales: “physical demand” (Dutch = 40.14, SD = 20.72; Indian = 60.00, SD = 24.49), as well as on the “temporal demand” (Dutch = 38.06, SD = 21.29; Indian = 49.82, SD = 17.35) and “frustration” (Dutch = 31.81, SD = 20.78; Indian = 45.00, SD = 22.11).



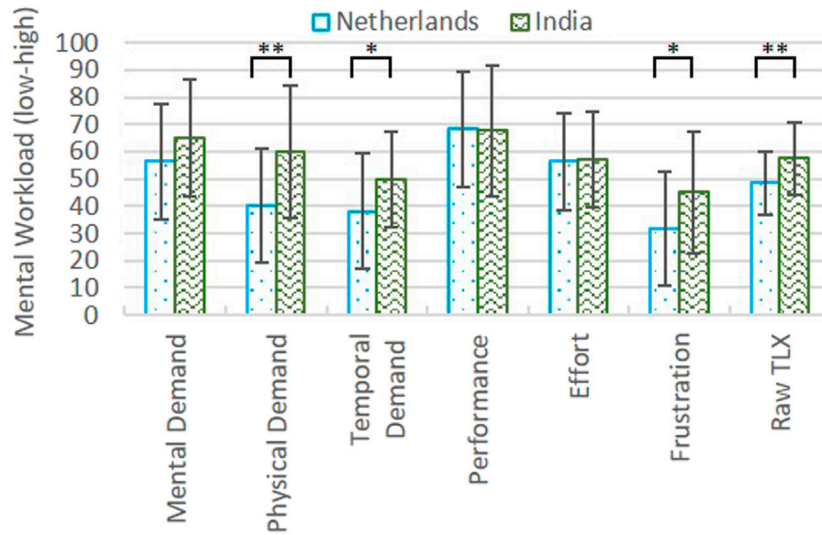


FIGURE 8 | The mental workload of the VOR from the Dutch and Indian participants. In the NASA-TLX, “1” means “very low mental workload” and “100” means “very high mental workload.”

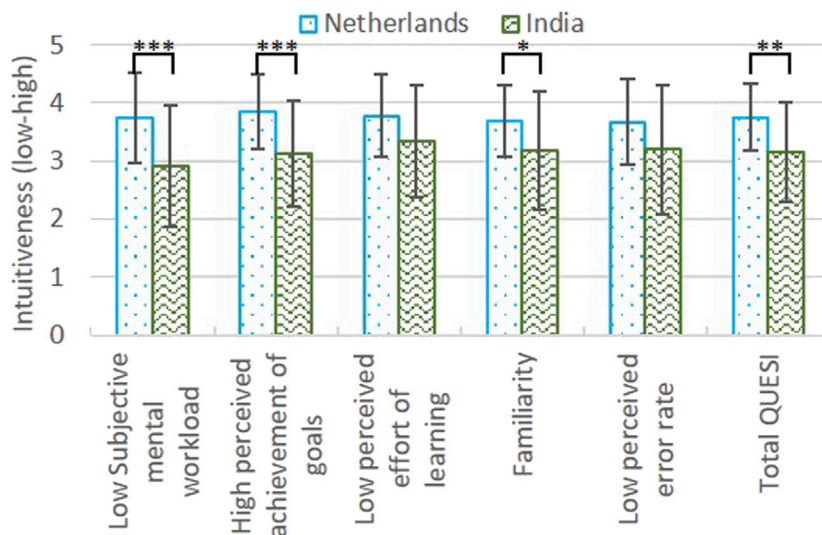


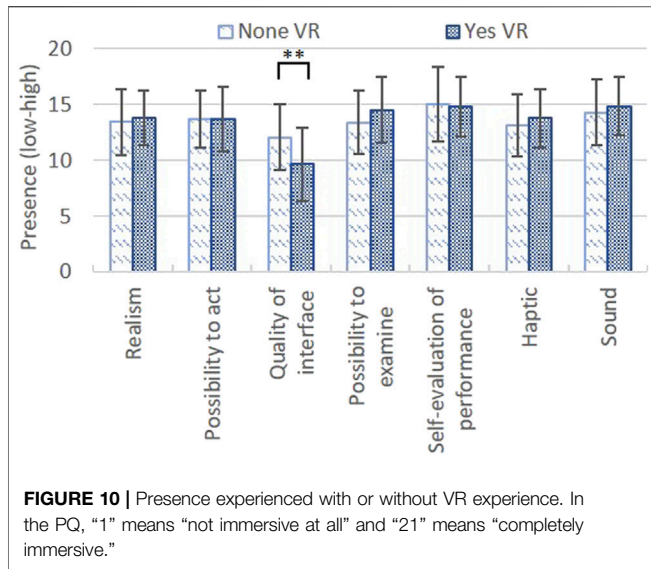
FIGURE 9 | The intuitiveness of VOR from Dutch and Indian participants *via* QUESI. In the QUESI, “1” means “low intuitive” and “5” means “highly intuitive”. Sub-Q3: With or without experience on VR technologies.

The degrees of intuitiveness were also significantly different across the Indian and the Dutch responders. The Dutch rated total QUESI significantly higher than the Indian counterparts (total QUESI: Dutch = 3.75, SD = 0.58; Indian = 3.15, SD = 0.86), and the scores of “low subjective mental workload” (Dutch = 3.75, SD = 0.58; Indian = 2.92, SD = 1.05) and “high perceived achievement of goals” (Dutch = 3.85, SD = 0.63; Indian = 3.12, SD = 0.91) were significantly higher (Figure 9). The Dutch participants significantly felt more familiar with the VOR than the Indian participants (Dutch = 3.69, SD = 0.61; Indian = 3.18, SD = 1.02). Both Dutch and Indian participants perceived “low effort

of learning” (Dutch = 3.78, SD = 0.70; Indian = 3.35, SD = 0.97) and “low error rate” (Dutch = 3.67, SD = 0.74; Indian = 3.20, SD = 1.12).

Sub-Q3: With or Without Experience on VR Technologies Immersion: Presence Questionnaire and Heuristic Scale

In general, non-VR and VR users would have a similar experience on immersion (PQ means: non-VR = 13.57, SD = 2.92; VR = 13.58, SD = 2.80), except for “quality of interface”



(Figure 10). The VR users rated the interface quality of the VOR significantly lower than the non-VR users (“quality of interface”: non-VR = 12.07, SD = 2.91; VR = 9.63, SD = 3.31).

In Figure 11, the VR users demonstrated a significantly higher need for autonomy such as “reversible actions” (HE means: non-VR = 12.07, SD = 2.91; VR = 9.63, SD = 3.31). The requirements associated with customization such as “use user’s language” (non-VR = 3.37, SD = 1.61; VR = 2.96, SD = 1.32) and “user control” also seemed higher for the VR users (non-VR = 3.50, SD = 1.41; VR = 2.93, SD = 1.47).

Usability: NASA-TLX and QUESI

In brief, the workload of using the VOR was proper (RTLX: non-VR = 51.78, SD = 13.78; VR = 52.31, SD = 13.91), and VR users seemed to have slightly higher mental workload. “Performance” (non-VR = 68.50, SD = 20.39; VR = 72.22 SD = 22.84), “mental demand” (non-VR = 57.83, SD = 21.32; VR = 60.74, SD = 24.21), and “effort” (non-

VR = 55.50, SD = 15.61; VR = 56.11, SD = 20.58) were the main resources of mental workload for both groups, while “frustration” was the lowest mental workload (total = 37.02, SD = 23.01). The VR and non-VR users showed no significant differences on the subscales of mental workload (Figure 12).

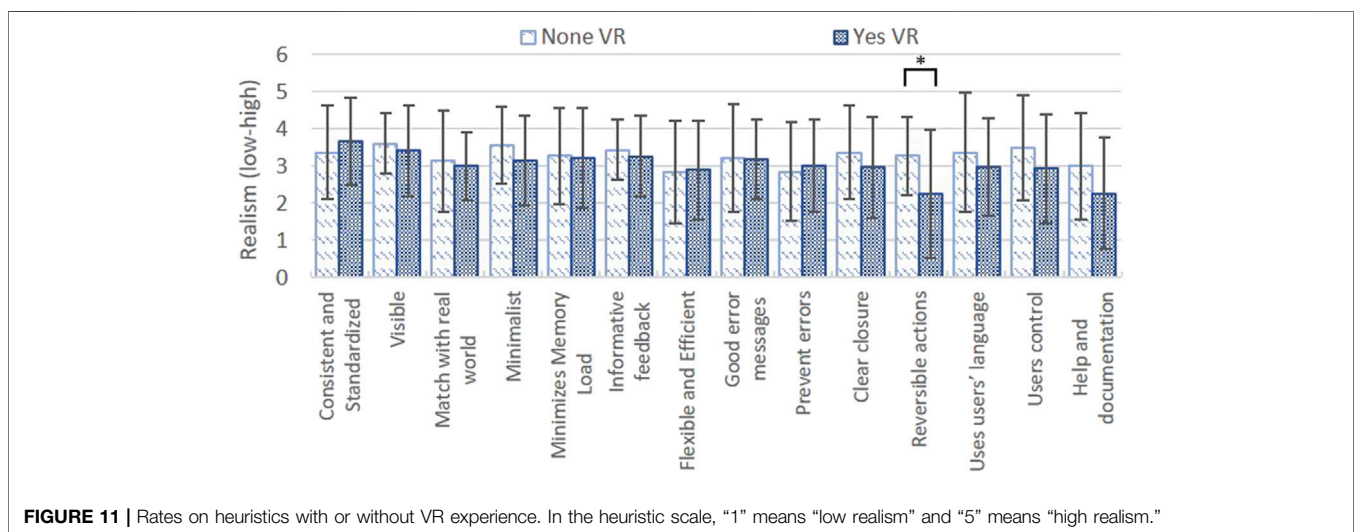
Using the VOR system was modestly intuitive (total QUESI: non-VR = 3.38, SD = 0.82; VR = 3.52, SD = 0.76) with or without VR experience. The VR users seemed to rate the intuitiveness slightly higher than the non-VR users, especially on “low subjective mental workload” (non-VR = 3.22, SD = 0.98; VR = 3.40, SD = 1.06), “high perceived achievement of goals” (non-VR = 3.33, SD = 0.91; VR = 3.67, SD = 0.77), and “low perceived error rate” (non-VR = 3.35, SD = 1.03; VR = 3.48, SD = 0.92). The differences were not significant.

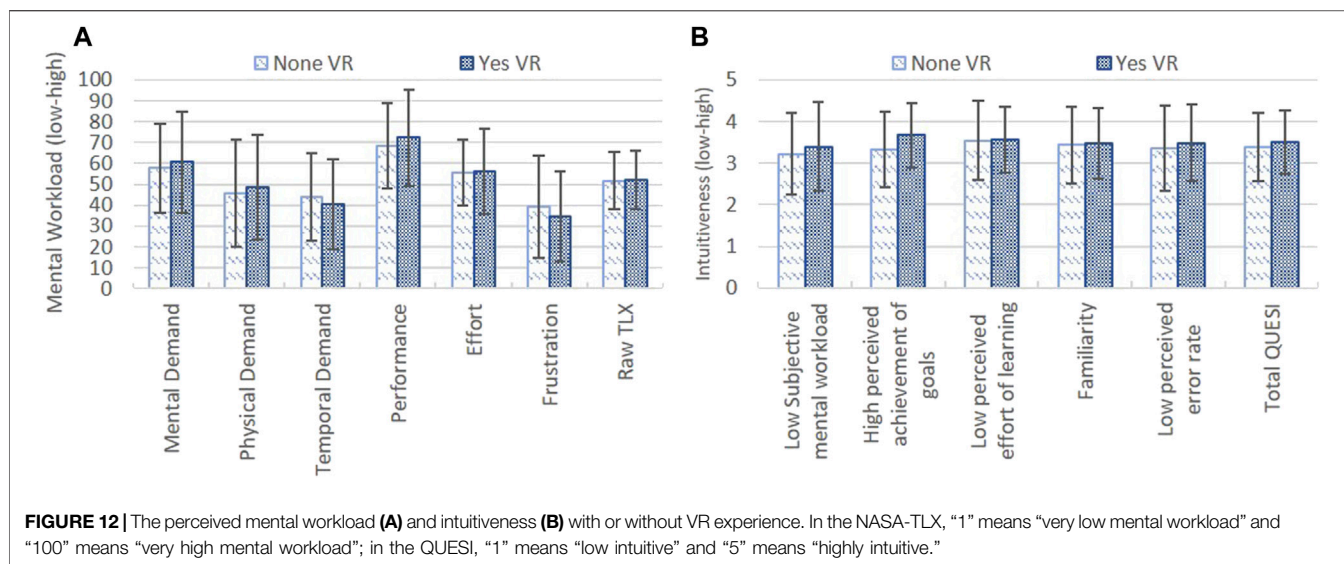
Main Effect and Interaction Effects Surgical Knowledge Versus Cultural Difference

There were obvious interaction effects between surgical knowledge and cultural difference except for mental workload, where the effect of surgical knowledge was dominant (Figure 13). The differences in the perception of presence, realism, and intuitiveness among the cross-culture novice groups tended to become larger, while the mental workload was slightly smaller. The main effects of the following factors were significant: culture on realism ($p < 0.001$), culture ($p < 0.001$), or surgical ($p < 0.01$) knowledge on mental workload, and culture on intuitiveness ($p < 0.05$). The interaction effects were not significant.

Surgical Knowledge Versus VR Experience

The interactive effect of surgical knowledge and VR experience was also distinct other than immersion (Figure 13). Surgical knowledge determined the level of perceived immersion. VR experience made the realism perception either drop or rise among the novices. Unlike the mental workload, the experts without VR experience might feel the VOR the most unintuitive, while the difference between novices with or without VR experience was smaller. The main effects and interaction effects were not significant.





Cultural Difference Versus VR Experience

The cultural difference and VR experience seemed to have almost no interactive effect; hence, the cultural difference was the determining factor (Figure 13). The different perceptions of immersion and mental workload became larger between the cross-cultural non-VR groups, while the different perceptions of intuitiveness became slightly smaller between the non-VR groups. The main effects were significant: culture on presence ($p < 0.05$), realism ($p < 0.05$), mental workload ($p < 0.01$), and intuitiveness ($p < 0.01$).

According to the interaction effects, the cultural difference was stronger than surgical knowledge and VR experience on the perception of immersion and usability. The strongest interaction effect seemed to appear between surgical knowledge and VR experience (interaction effect = 12.19), which seemingly was positively enhancing. The second large interaction effect was between surgical knowledge and cultural difference (interaction effect = 6.80). The main effects of cultural differences were significant on each aspect of immersion and usability, while surgical knowledge showed its influence on mental workload. Among the cross-culture expert groups, the differences in immersion, realism, and intuitiveness across the novice groups in a different culture increased at the same time.

Qualitative Feedback Experts Versus Novices

The qualitative feedback from the experts concentrated on the stiffness of the haptic interface and the rigidness of the surgical procedure. The majority of the experts were annoyed by the tool-changing interaction, which violated real tool-changing maneuvers. Besides, the correct OR layout and the availability of preferred instruments were also of concern to the experts. The novices more or less felt immersed by the sound simulation within the VOR, and the blurring lens was their focus.

The Dutch Versus Indian Participants

Both the Dutch and Indian participants expressed a strong need for localizing the communicating language as well as some surgical practices. The Indian surgeons pointed out a) the OR team would

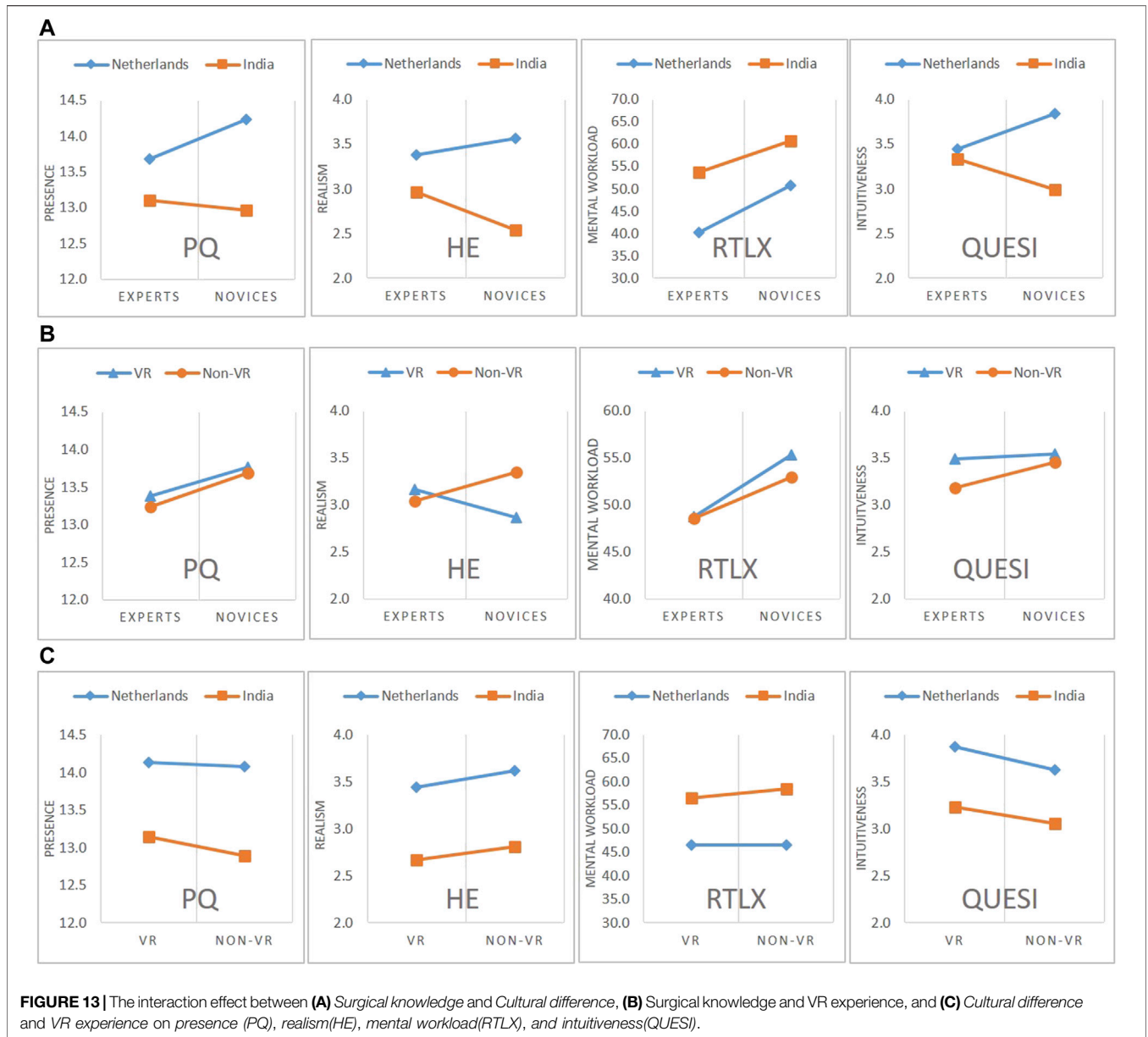
normally be located in a different placement other than the VOR, b) the team interaction was distracting and unrealistic, and c) the background sound was unfamiliar and unrelated. The Dutch surgeons commented a) the team communication was repetitive and disrupting, b) the camera assistant was missing, c) the team interaction was impersonal, and d) the background music was pleasant but needed to be personalized.

With Versus Without VR Experience

The experience VR user often mentioned was the “screen-door” effect of the VR headset and the game-like feeling during the training. They appeared to be more relaxed even for the first hands-on, while the non-experienced VR user tended to be more stressed when some errors appeared during the testing. Both groups tended to forget the time when they used the VOR system.

DISCUSSION

This study aimed at understanding the effect of surgical knowledge, cultural difference, and VR experience on the immersion and usability of a VOR system. Considering the surgical knowledge, when the authors combined the surgeons and surgical trainees from different cultures, interesting results were obtained. Unlike the results comparing the experts and the novices in the Netherlands or India, respectively, the only significant differences between the experts and the novices presented only on “mental demand.” It was also identified independently *via* Dutch and Indian studies (Li et al., 2020, 569; Ganni et al., 2020, 4). The scores from the Dutch novices were significantly higher on “perceived effort of learning” of QUESI, as well as the raw TLX and subscales like “physical demand,” “temporal demand,” and “effort” were disappeared (Li et al., 2020, 567). The higher rates from Indian experts on “subjective mental workload” and “perceived achievement of goals” of QUESI, as well as “prevent errors” and “reversible actions” of the heuristics, vanished alike (Ganni et al., 2020, 3–4). We might infer that the interaction between surgical knowledge and



the cultural difference would probably neutralize the significance of surgical proficiency, as we observed in *Surgical Knowledge Versus Cultural Difference*.

The most out-of-expectation finding of this study is the dominant effect of cultural difference. Despite the well-known phenomenon “WIRED” in academics, few pieces of evidence are showing the cultural differences in an immersive training setup. The people from paddy rice regions, such as east India, where Rajahmundry is located, tend to use more holistic thinking (55%) than analytic thinking (45%) (Henrich, 2020, 63, 253). The proportion of analytic thinking is 75% for the Dutch (Henrich, 2020, 63, 253). We would hence infer that the differences on the Presence Questionnaire and heuristic scale, for example, “consistent and standardized,” “visible,” and “matches with the real world”, might attribute to the holistic thinking from the Indian participants. Joseph Henrich found that a person

from intensive kinship culture would favor familiar relationships in teamwork, which was a rare case for those with weaker relational connections (Henrich 2020, 250–252). This might explain the significant differences in “familiarity.” Our brain needs more mental resources to handle the strange information, which might explain Indian participants wanted to “minimize memory load” and needed “informative feedback.” Using native language as a prominent symbol of local culture would increase the familiarity as pinpointed both in the questionnaires and the interview. People from rice paddy areas often have less “self-inflation” than the “WIRED” people (Henrich, 2020, 252). This might tell why the Indian participants had significantly lower scores on “self-evaluation of performance” and “perceived achievement of goals.”

The effect of VR experience was mainly on immersion where the low quality of interfaces were easily noticeable. The familiarity of the

technologies, such as desktop display vs. stereoscopic displays, might alter the focuses of the perception (Santos et al., 2009, 171–175). The knowledge of VR displays enabled the experienced VR to hold predefined focuses, for example, resolution and “screen-gate effect,” and recognize relate phenomena in the testing, as found in the interview as well. Another interesting finding is the interaction effects among surgical knowledge, cultural difference, and the VR experience. VR experience showed a strong confounding effect with surgical knowledge, especially on mental workload, while it had very little influence on cultural differences. The low intuitiveness of non-VR surgical experts indicated that domain proficiency sometimes might hinder the acquisition of new skills, especially those against their automated maneuvers.

Regardless of the effect from surgical knowledge, cultural difference, and VR experience, the participants concurrently experienced high mental demand, high performance challenge, and low frustration state. It seems to be impossible; however, it might imply that the participants were under a sort of “flow state” (Pilke, 2003, 348). Flow refers to the optimal experience, which is the state between frustration and boredom, where the mental state becomes an extremely rewarding concentration (Csikszentmihalyi, 1990). The enjoyment of immersion stems from the perception of “being in a complete absorption” with the unified novel narrative schema (Douglas and Hargadon, 2000, 154; Hekkert et al., 2003, 112–113). The pleasure of engagement appears to come from interactivity with an array of preestablished schemas (Douglas and Hargadon, 2000, 156). In medical education, immersive and interactive training such as simulation-based learning recently has gained significance (Taekman and Shelley, 2010, 102). The virtual environments provide all the prerequisites for flow. When it integrates team-based learning to stimulate team interaction, the enhanced immersion and engagement will merge into the flow (Douglas and Hargadon, 2000, 158; Taekman and Shelley, 2010, 116).

The following limitations of this study would open room for further research: 1) despite shedding light on cross-culture issues of VR-based training, the sample size was not large enough to explain these phenomena in depth. This evaluation protocol needs to be applied to other cultural contexts in healthcare education. 2) The immersive and engaging effect was discovered, but the understanding of how each type of distractors would influence the presence and usability is not reached yet. Systematic investigating, categorizing, and simulating distractors are the benchmarks of this work. 3) Subjective assessments are susceptible to personal bias and low replicability. Future studies should introduce “data-driven methods” that collect objective data, such as error rate and task completion time, and physiological data, such as eye-tracking, EEG, and facial EMG. 4) VR experience of seven novices is not available. Considering the sample size, further studies should verify the results with a larger sample size focusing on VR experience. To facilitate the future design of VR immersive training, a design guide is under development for a highly personalized experience.

CONCLUSION

This study explored the effects of surgical knowledge, cultural difference, and VR experience on the presence and usability of an

educational VR environment. The novelties of this study are as follows: 1) demonstrating the cultural differences of sixty-four novices and experts in presence, realism, mental workload, and intuitiveness; 2) the interaction effects of these main factors were shown, especially the strong interaction between surgical knowledge and cultural difference; 3) proposed “flow state” as a key feature for the future VR-based professional training. Despite the limited applications, VR-based immersive training is attracting attention both from academic and industrial fields. Integrating immersive technologies *via* human-centered design is opening a brand new horizon for health care and similar professional training. The numerous experts’ view collected in this study provides a solid base of a design guide for VR-based immersive training focusing on health care.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**; further inquiries can be directed to the corresponding author.

ETHICS STATEMENT

Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

AUTHOR CONTRIBUTIONS

ML and SG finished the experiment and wrote the first draft of this article. AA and AR provided the critical review of the questionnaires and research design, as well as the manuscript. DE and JJ organized the facilities of the research and reviewed the article critically.

FUNDING

ML doctoral research is sponsored by the China Scholarship Council (No. 201706280020).

ACKNOWLEDGMENTS

The authors want to express their appreciation to all the participants and the skill labs from both hospitals for their contribution to this study. We especially thank Jeroen Punten, Samir Ranjan Nayak, Bhashar Rao Ganni, and Gert de Vroom for their heartily help and support in organizing this study.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/frvir.2021.675334/full#supplementary-material>

REFERENCES

- Badash, I., Burt, K., Solorzano, C. A., and Carey, J. N. (2016). Innovations in Surgery Simulation: a Review of Past, Current and Future Techniques. *Ann. translational Med.* 4, 23. doi:10.21037/atm.2016.12.24
- Berguer, R., Chen, J., and Smith, W. D. (2003). A Comparison of the Physical Effort Required for Laparoscopic and Open Surgical Techniques. *Arch. Surg.* 138 (9), 967–970. doi:10.1001/archsurg.138.9.967
- Berguer, R., Smith, W. D., and Chung, Y. H. (2001). Performing Laparoscopic Surgery Is Significantly More Stressful for the Surgeon Than Open Surgery. *Surg. Endosc.* 15, 1204–1207. doi:10.1007/s004640080030
- Bowman, D. A., Gabbard, J. L., and Hix, D. (2002). A Survey of Usability Evaluation in Virtual Environments: Classification and Comparison of Methods. *Presence: Teleoperators & Virtual Environments* 11 (4), 404–424. doi:10.1162/105474602760204309
- Bowman, D. A., and McMahan, R. P. (2007). Virtual Reality: How Much Immersion Is Enough? *Computer* 40, 36–43. doi:10.1109/mc.2007.257
- Csikszentmihalyi, M. (1990). *Flow: Psychol. optimal experience*, Vol. 1990. New York: Harper & Row.
- Douglas, Y., and Hargadon, A. (2000). “The Pleasure Principle: Immersion, Engagement, Flow.” in Proceedings of the Eleventh ACM on Hypertext and Hypermedia, San Antonio TX, May 2000. (ACM), 153–160.
- Ganni, S., Meng, Li., Botden, S., Nayak, S. R., Ganni, B. R., Rutkowski, A. F., et al. (2020). Virtual Operating Room Simulation Setup (VORSS) for Procedural Training in Minimally Invasive Surgery—A Pilot Study. *Indian J. Surg.* 82 (6), 1–7.
- Hao, C. (2019). *Cultura: Achieving Intercultural Empathy through Contextual User Research in Design*. Ph.D. diss.: Delft University of Technology.
- Hart, S. G. (2006). NASA-task Load index (NASA-TLX); 20 Years Later. *Proc. Hum. Factors Ergon. Soc. Annu. Meet.* 50, 904–908. doi:10.1177/154193120605000909
- Hekkert, P., Snelders, D., and Wieringen, P. C. W. (2003). ‘Most Advanced, yet Acceptable’: Typicality and novelty as Joint Predictors of Aesthetic Preference in Industrial Design. *Br. J. Psychol.* 94 (1), 111–124. doi:10.1348/000712603762842147
- Henrich, J., Heine, S. J., and Norenzayan, A. (2010). The Weirdest People in the World? *Behav. Brain Sci.* 33 (2–3), 61–83. doi:10.1017/s0140525x0999152x
- Henrich, J. (2020). *The Weirdest People in the World: How the West Became Psychologically peculiar and Particularly Prosperous*. Farrar, Straus and Giroux.
- Hornbeck, R. G., and Barrett, J. L. (2013). Refining and Testing “Counterintuitiveness” in Virtual Reality: Cross-Cultural Evidence for Recall of Counterintuitive Representations. *Int. J. Psychol. Religion* 23 (1), 15–28. doi:10.1080/10508619.2013.735192
- Jakimowicz, J. J., and Buzink, S. (2015). Training Curriculum in Minimal Access Surgery. in *Training in Minimal Access Surgery*. London: Springer, 15–34. doi:10.1007/978-1-4471-6494-4_2
- Lee, G. I., Lee, M. R., Clanton, T., Sutton, E., Park, A. E., and Marohn, M. R. (2014). Comparative Assessment of Physical and Cognitive Ergonomics Associated with Robotic and Traditional Laparoscopic Surgeries. *Surg. Endosc.* 28, 456–465. doi:10.1007/s00464-013-3213-z
- Li, M., Albayrak, A., Zhang, Y., and van Eijk, D. (2018). Multiple Factors Mental Load Evaluation on Smartphone User Interface, in *Congress of the International Ergonomics Association*. Cham: Springer, 302–315. doi:10.1007/978-3-319-96059-3_33
- Li, M., Ganni, S., Ponten, J., Albayrak, A., Rutkowski, A. F., and Jack, J. (2020). “Analysing Usability and Presence of a Virtual Reality Operating Room (VOR) Simulator during Laparoscopic Surgery Training.” in 2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR), Atlanta, GA, March 2020. IEEE, 566–572.
- Lin, J., Zhu, R., Li, N., and Becerik-Gerber, B. (2020). Do people Follow the Crowd in Building Emergency Evacuation? A Cross-Cultural Immersive Virtual Reality-Based Study. *Adv. Eng. Inform.* 43, 101040. doi:10.1016/j.aei.2020.101040
- Munz, Y., Kumar, B. D., Moorthy, K., Bann, S., and Darzi, A. (2004). Laparoscopic Virtual Reality and Box Trainers: Is One superior to the Other? *Surg. Endosc.* 18, 485–494. doi:10.1007/s00464-003-9043-7
- Naumann, Anja., and Hurtienne, Jörn. (2010). “Benchmarks for Intuitive Interaction with mobile Devices.” in Proceedings of the 12th International Conference on Human Computer Interaction with mobile Devices and Services, Lisbon, Portugal, September 7–10, 2010, 401–402.
- Nisbett, R. (2004). *The Geography of Thought: How Asians and Westerners Think Differently—and Why*. New York: Simon & Schuster, Free Press.
- Parker, S. E. H., Laviana, A. A., Wadhwa, R. K., Wiegmann, D. A., and Sundt, T. M. (2010). Development and Evaluation of an Observational Tool for Assessing Surgical Flow Disruptions and Their Impact on Surgical Performance. *World J. Surg.* 34, 353–361. doi:10.1007/s00268-009-0312-z
- Pilke, E. M. (2003). Flow Experiences in Information Technology Use. *Int. J. human-computer Stud.* 61 (3), 347–357. doi:10.1016/j.ijhcs.2004.01.004
- Pluyter, J. R., Buzink, S. N., Rutkowski, A.-F., and Jakimowicz, J. J. (2010). Do absorption and Realistic Distraction Influence Performance of Component Task Surgical Procedure? *Surg. Endosc.* 24 (4), 902–907. doi:10.1007/s00464-009-0689-7
- PRISMA health (2021). *Becoming a General Surgeon*. Available at: <https://schealthsciencescenter.org/becoming-a-general-surgeon/> (Accessed March 1, 2021).
- Saalfeld, P., Mewes, A., Luz, M., Preim, B., and Hansen, C. (2015). Comparative Evaluation of Gesture and Touch Input for Medical Software. in *Mensch and Computer*. Berlin: Walter de Gruyter GmbH, 143–152.
- Santos, B. S., Dias, P., Pimentel, A., Baggerman, J. W., Ferreira, C., Silva, S., et al. (2009). Head-mounted Display versus Desktop for 3D Navigation in Virtual Reality: a User Study. *Multimedia tools Appl.* 41, 161–181. doi:10.1007/s11042-008-0223-2
- Schijven, M. P., Jakimowicz, J. J., Broeders, I. A. M. J., and Tseng, L. N. L. (2005). The Eindhoven Laparoscopic Cholecystectomy Training Course—Improving Operating Room Performance Using Virtual Reality Training: Results from the First E.A.E.S. Accredited Virtual Reality Trainings Curriculum. *Surg. Endosc.* 19, 1220–1226. doi:10.1007/s00464-004-2240-1
- Seymour, N. E., Gallagher, A. G., Roman, S. A., O’Brien, M. K., Bansal, V. K., Andersen, D. K., et al. (2002). Virtual Reality Training Improves Operating Room Performance. *Ann. Surg.* 236 (4), 458–464. doi:10.1097/0000658-200210000-00008
- Strachan, I. W. (2000). “Technology Leaps All Around Propel Advances in Simulators”. *Natl. Defense: Business Technol. Mag.*, November 1, 2000. Available at: <https://www.nationaldefensemagazine.org/articles/2000/10/31/2000november-technology-leaps-allaround-propel-advances-in-simulators>
- Taekman, J. M., and Shelley, K. (2010). Virtual Environments in Healthcare: Immersion, Disruption, and Flow. *Int. anesthesiology Clin.* 48 (3), 101–121. doi:10.1097/aia.0b013e3181eace73
- Van de Ven, J., Van Baaren, G. J., Fransen, A. F., van Runnard Heimeel, P. J., Mol, B. W., and Oei, S. G. (2017). Cost-effectiveness of Simulation-Based Team Training in Obstetric Emergencies (TOSTI Study). *Eur. J. Obstet. Gynecol. Reprod. Biol.* 216, 130–137. doi:10.1016/j.ejogrb.2017.07.027
- van Houwelingen, B. C. G., Rutkowski, A.-F., Ganni, S., Stapaniak, P. S., and Jakimowicz, J. J. (2019). Effects of Surgical Flow Disruptions on Surgeons’ Resources: A Pilot Study. *Surg Endosc* 34, 1–11. doi:10.1007/s00464-019-07239-2
- Wiegmann, D. A., ElBardissi, A. W., Dearani, J. A., Daly, R. C., and Sundt, T. M. (2007). Disruptions in Surgical Flow and Their Relationship to Surgical Errors: An Exploratory Investigation. *Surgery* 142 (5), 658–665. doi:10.1016/j.surg.2007.07.034
- Witmer, B. G., Jerome, C. J., and Singer, M. J. (2005). The Factor Structure of the Presence Questionnaire. *Presence: Teleoperators & Virtual Environments* 14 (3), 298–312. doi:10.1162/105474605323384654
- Witmer, B. G., and Singer, M. J. (1998). Measuring Presence in Virtual Environments: A Presence Questionnaire. *Presence* 7 (3), 225–240. doi:10.1162/105474698565686
- Zhang, J., Johnson, T. R., Vimla, P. L., Danielle, P. L., and Kubose, T. (2003). Using Usability Heuristics to Evaluate Patient Safety of Medical Devices. *J. Biomed. Inform.* 36 (1–2), 23–30. doi:10.1016/s1532-0464(03)00060-1
- Zheng, B., Jiang, X., Tien, G., Meneghetti, A., Panton, O. N. M., and Atkins, M. S. (2012). Workload Assessment of Surgeons: Correlation between NASA TLX and Blinks. *Surg. Endosc.* 26, 2746–2750. doi:10.1007/s00464-012-2268-6

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Li, Ganni, Albayrak, Rutkowski, van Eijk and Jakimowicz. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.