

Quantitative Influence and Performance Analysis of VR Laparoscopic Surgical Training System

Peng Yu

Beihang University <https://orcid.org/0000-0002-8652-2744>

Junjun Pan (✉ pan_junjun@buaa.edu.cn)

Beihang University

Zhaoxue Wang

Beijing Normal University

Yang Shen

Beijing Normal University

Jialun Li

Beihang University

Aimin Hao

Beihang University

Haipeng Wang

Beijing General Aerospace Hospital

Research article

Keywords: VR, Laparoscopic surgical skills training, Cognitive load, Flow experience

Posted Date: April 27th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-462363/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background VR surgery training becomes a trend in clinical education. Many research papers validate the effectiveness of VR based surgical simulators in training surgeons. However, most existing papers employ subjective methods to study the residents' surgical skills improvement. Few of them investigate how to substantially improve the surgery skills on specific dimensions.

Methods In this paper, we resort to physiological approaches to objectively research quantitative influence and performance analysis of VR laparoscopic surgical training system for medical students. 41 participants were recruited from a pool of medical students. They conducted four pre and post experiments in the training box. In the middle of pre and post experiments, they were trained on VR laparoscopic surgery simulators (VRLS). When conducting pre and post experiments, their operation process and physiological data (heart rate and electroencephalogram) are recorded. Their performance is graded by senior surgeons using newly designed hybrid standards for fundamental tasks and GOALS standards for colon resection tasks. Finally, the participants were required to fill the questionnaires about their cognitive load and flow experience.

Results The results show that the VRLS could highly improve medical students' performance ($p < 0.01$) especially in depth perception and enable the participants to obtain flow experience with a lower cognitive load.

Conclusion The performance of participants is negatively correlated with cognitive load through quantitatively physiological analysis. This might provide a new way of assessing skill acquirement.

1. Background

As one of the modern minimally invasive procedures, laparoscopic surgery has become popular primarily due to its small wounds and rapid recovery. However, residents generally have a long training period (6 years at least) to be trained as qualified laparoscopic surgeons [1]. In the narrow operation space, mistakes may easily occur if laparoscopic surgeons do not handle procedures properly. Traditional laparoscopic surgical training usually chooses the training box within vitro animals or corpses organs, which could give rise to negative effects, such as high cost, low reusability, and related ethical issues [2]. The advent of VR surgical simulators has changed the surgeons learning mode. It can simulate the surgery from the visual, auditory, and tactile aspects. It not only reconstructs the real surgical environment and procedures but also can be reused for a variety of designed training tasks without surgical risk [3]. Thus, training on laparoscopic surgical simulator based on virtual reality (VR) has gradually becoming a standard in Europe at present. There are many pieces of research that validate the effectiveness of VR based laparoscopic surgery simulators (VRLS) on training surgeons [4, 5, 6]. To our knowledge, most of them employ subjective methods to study the improvement of medical students' surgical skills through VRLS [7]. But few papers investigate how to substantially improve the surgery skills on specific dimensions, such as physiological and psychological perspectives.

In this paper, we resort to physiological approaches to quantitatively measure the influence (cognitive load and flow experience) of VRLS on medical students.

Cognitive load theory (CLT) builds on established models of memory, including subsystems of sense, work, and long-term memory. John Sweller first put forward a systematic study in 1988 and established a theoretical hypothesis [8]. He believed that cognitive load refers to the total amount of mental activities exerted on an individual's cognitive system during a specific time of operation. CLT is an important learning theory, which is paid more and more attention in medical education. The medical field is a complex knowledge field. Medical workers need to integrate a variety of knowledge, skills and behaviors at a specific time and place at the same time, and make quick responses and decisions, which is prone to excessive cognitive load or even overload phenomenon. For novice physicians, complex tasks such as surgery can lead to excessive cognitive load (CL), which can have a negative impact on learning [9]. Rasiah assessed the relationship between cognitive load and dexterity parameters and found that reduced cognitive load significantly affected learning outcomes [10].

Flow experience is the state of mind in which a person is fully engaged in some activity and reaches an extreme level of pleasure [11]. It was first proposed by Csikszentmihalyi in the 1960s. He found that when people maximize their physical, mental, and mental states, they often produce the ultimate optimal experience [11]. In simulated laparoscopic surgery, Flow experience is expressed as "go all-out work". Studies have shown that improving flow experience during laparoscopic surgery can improve the effectiveness of the operation, thereby improving patient safety [12].

To our knowledge, there is no research to directly and quantitatively explore the relationship between the flow and total cognitive load. However, some studies have shown that learners with good academic performance have higher flow experience, lower external cognitive load, and higher related cognitive load [13, 14]. Chang's research has shown that flow experience is related to three different cognitive loads, confirming that media richness and game interaction can improve learners' flow experience, reduce external cognitive load, and promote closely related cognitive load [15].

In the medical field, the methods of measuring the cognitive load of medical workers basically follow the measurement technology of cognitive load or cognitive load [16]. The measurement methods for measuring cognitive load can generally be classified into three categories: subjective measurement, task performance, and physiological measurement. These three measurement approaches were all utilized in our study. At present, the main cognitive load measurement method is subjective measurement [17]. Some studies have found that combining subjective mental effort indicators and objective behavioral performance indicators to form a comprehensive indicator that can reveal some important information about cognitive load [18, 19, 20]. At present, the more classic scale is the Pass scale [20], SWAT scale [21] and NASA-TLS scale [22].

The main advantages of the psychophysiological measurement of cognitive load are the objectivity of the measurement, the sensitivity to different cognitive processes, the non-interference of the program, and their implicitness and continuity [23]. EEG is considered a physiological indicator, which can be used as

an online and continuous cognitive load measurement method to detect subtle fluctuations in instantaneous load. Measuring the changes in alpha and theta brainwave rhythms reflects what happens in the participant's information processing situation, even if the participant does not know these changes or cannot express them in words [24, 25]. As for the cognitive load, α is gradually suppressed. As the difficulty of the task increases, so does the amount of θ activity [26, 27, 28].

The most commonly used methods to measure flow experience are retrospective questionnaires and interviews [29, 30]. The Flow State Scale compiled by Jackson is mainly developed from the nine elements proposed by Csikszentmihalyi [31]. Many scholars develop a new flow scale based on the characteristics and needs of VR learning situations, but a sense of control, immersion, clear goals and feedback are still indispensable dimensions in the measurement of flow experience [32, 33, 34]. At present, there are few studies on evaluating flow experience in virtual reality based on physiological indicators [30]. Flow experience can be expressed in physical and physiological characteristics, as an objective flow indicator [15]. In terms of EEG, some studies indicate that the correlation between EEG and flow experience under peak performance conditions, and the induction of flow experience can improve the performance of workplaces, sports fields. Frederick [35] proposed EEG is an objective method of measuring the flow, it is more accurate than subjective behavioral measures.

With the rapid development of wearable devices, we could obtain different modality physiological data. The most easily acquired physiological data is the heart rate. Besides, we measure the cognitive load using EEG. Three hypotheses are designed as follows:

H1: Training on VRLS could improve the performance of medical students in some dimensions.

H2: Training on VRLS could improve the flow experience and lower the cognitive load for medical students in some dimensions.

H3: The performance is positively related to flow experience and negatively related to the cognitive load.

These hypotheses are validated through the following experiments and user studies. 41 Participants were recruited using a pool of medical students which contains both undergraduates and graduates. They conducted four pre and post experiments in the training box. In the middle of pre and post experiments, they were trained on VRLS. When conducting pre and post experiments, their operation process and physiological data (heart rate and electroencephalogram) are recorded. Their performance is graded by senior surgeons using newly designed hybrid standards for fundamental tasks and Global Operative Assessment of Laparoscopic Skills (GOALS) standards [36] for colon resection tasks. Finally, the participants were required to fill the questionnaires about their cognitive load and flow experience.

There are two main contributions provided by this work: (1) Using multimodal sensing data (EEG and heart rate), we design a physiological approach to quantitatively measure the influence of VRLS on medical students; (2) Our experiments reveal the negative correlation between the skill performance of trainees and their cognitive load, after correlation analysis. This research can identify the potential

benefits of VRLS and its improvement opportunities in laparoscopic procedure training. The remainder of this paper is organized as follows. We introduce our experiments, including participants, platform and procedure in Sect. 2. The results are documented and analyzed in Sect. 3. Section 4 provides the discussion. Section 5 concludes the paper with future work.

2. Methods

2.1 Participants

In this study, we recruited 41 participants between 17 and 27 years old (21.10 ± 2.79). Our experiments were approved by our hospital ethics committee. There are 15 male and 26 female medical students. 2 persons are left-handed and the others are right-handed. 31 participants never experienced VR and 10 participants have played VR or AR games once or twice before this study. 14 persons (34.15%) played games on PC or mobile phone every day. Just 4 (12.20%) participants rarely play games in their daily life. All participants never experienced VR based laparoscopic surgery simulator before this study.

2.2 Platform

The pre-test and post-test were carried out using a training box. The commercial laparoscopic physical training box (38×27×27 cm) is illustrated in Fig. 2 (a). There is a high-resolution camera on the inner top of the training box. All the operation process of participants could be recorded using the camera. The participants were required to conduct four tests (three fundamental laparoscopic surgery skill training tasks and one colon resection task) illustrated in Fig. 3. Three fundamental surgery skill training tasks are peg transfer, picking beans and threading skill practice. The electroencephalography (EEG) data were collected using a four-channel dry electrode headset (Muse 2, InteraXon Inc.) shown in Fig. 2 (b). The sensors are located at TP9 (left ear), AF7 (left forehead), AF8 (right forehead), TP10 (right ear). Compared with traditional EEG devices such as fMRI, PET and 20/32/64 electrode headsets, Muse 2 is portable and easy to use. [37] and [38] pointed out that the accuracy of Muse is good enough for research, even though there are only four electrodes.

The VR laparoscopic simulator (Fig. 4) was developed by the State Key Lab of VR Tech & Syst in Beihang University. The simulator consists of two major components. The first component is the computation module, which is a high-performance PC connected with a touch-screen monitor (1920×1080 dpi). The hardware parameters are as follows: Intel(R) Core (TM) i5-8500 CPU @3.00GHZ with 6 cores and NVIDIA GeForce GTX 1060 with 6GB memory. The software is running on Windows 10 64 bits professional version. The second component is the simulation module, which contains two surgical handlers connected with haptic devices (Geomagic touch, 3D System Company, US) and a navigation camera in a box. Two-foot pedals were utilized to activate the electrosurgical coagulation during surgery training.

The whole experiment consists of three main steps. Firstly, the participants were required to conduct a pre-test on a training box. Before this step, we explained the purpose of this study and how to operate. The pre-test contains three fundamental surgery skills tasks (Pre-FT) and one colon resection task (Pre-

CRT). During the operation, we recorded the heart rate and EEG data, using Polar H10 heart rate monitor chest strap and Muse 2 brain wave monitor respectively. Besides, the whole procedure was recorded as videos. Secondly, the participants were asked to conduct the same kind of tasks on VRLS. Everyone had to complete 4 trials within a week, and each trial lasts about 30 minutes. Finally, we required our participants to conduct the post-test. The post-test procedure (Post-FT and Post-CRT) is the same as the pre-test. After finishing all experiments, participants were asked to complete four questionnaires regarding the cognitive load and flow experience. Each questionnaire reflects the experimental results with different dimensions. To explore more information about the influence of VRLS on participants, we utilize three scales (Pass [39], NASA-TLX [40], WP Scale [41]) to measure the cognitive load. To measure the flow experience during the experiments, we combine two scales from EGame scale [34] and Cheng's scale [42] and redesigned the questions according to our experiments.

2.3 Data Processing

2.3.1 Data Collection

In this study, we obtain three types of data. The first is the performance scores computed from recorded videos according to the GOALS standards for colon resection task and our designed measure rules (e.g. completion time, number of mistakes, etc.) for fundamental surgery skill tasks. The second is the self-reported scores including cognitive load scores and flow experience scores computed from questionnaires. The third is the physiological data extracted from heart rate data and EEG. The performance scores and physiological data need to be processed before getting meaningful information. For performance scores, the fundamental surgery skill tasks and colon resection tasks are measured from different dimensions. We measured the performance of fundamental surgery skills from 7 dimensions including the completion time, the number of failures in peg transfer and picking beans, the number of times rope dropped, motion smoothness, depth perception and bimanual dexterity. The scores of each item are normalized and scaled to [0,10], then we obtain the final performance score by the sum of all items' scores. The GOALS standard measured one's laparoscopic skills from 4 aspects: depth perception, bimanual dexterity, efficiency, tissue handling and autonomy. Each dimension is in the range of [0,10].

For physiological data, we filtered the data at first. During the experiment, we found that the heart rate monitor chest strap could steadily and robustly measure the heart rates. However, the brain sensing headband is sensitive to head motion and the contacts between skin and electrodes. Thus, we filtered the EEG data according to four data quality indicators (1 = Good, 2 = Medium, 4 = Bad). Besides, the electrooculography (EOG) signal has strong disturbing effects. Fortunately, it can be measured and recorded using AF7 and AF10. Thus, we could easily delete these interference signals such as blink and jaw clench.

After that, we could process physiological data from the time domain, frequency domain and nonlinear domain. For heart rate data, we compute each participants' average, minimum, maximum heart rate. Then, we compute training load energy [43]. For EEG data, the exported data contains brain wave values

that are absolute band powers, based on the algorithm of the Power Spectral Density (PSD) of the EEG data for each channel. There are five frequency spectrums: Delta (1-4Hz), Theta (4–8 Hz), Alpha (7.5-13Hz), Beta (13–30 Hz), Gamma (30-44Hz). With four sensors (TP9, AF7, AF8, TP10), we could get 20 numbers for each frame. The raw EEG data in microvolts(μV), the range of which is [0,1682] for each sensor is also recorded. To study the cognitive load change between pre and post-test, we compute the participants' cognitive load score [44].

2.3.2 Statistical Analysis

After pre-processing all raw data, we utilize SPSS (Statistics V.25) to analyze our computed scores. We use the classical paired samples t-test to test the differences between pre and post tasks. A p-value of < 0.05 was recognized as statistically significant.

3. Results

In this section, we investigate the influence of VRLS on participants from two aspects: performance (Sec. 3.1) and physical-psychological (Sec. 3.2, Sec. 3.3). Besides, we attempt to reveal their correlations in Sec. 3.4.

3.1 Performance and VRLS

We could easily measure one's surgical skill proficiency using task completion time. We found that, after training on VRLS, the time to complete the same task could drop sharply ($p < 0.01$). The first column of Table 1 shows the time required for each participant to complete each task. The left upper of Fig. 7 is the intuitive presentation using a histogram. The efficiency of participants improved 1.6 and 5.4 times for fundamental surgery task and colon resection task respectively. Besides, the performance scores for fundamental surgery skill task (FT) and colon resection task (CRT) are shown in Fig. 5 and Fig. 6. The pre-fundamental surgery skill task score is significantly lower than the post fundamental surgery skill task score ($p < 0.01$). We could obtain the same results for the colon resection task. In summary, we could be sure that training on VRLS would increase the performance of medical students.

Furthermore, we inspect the performance change of colon resection task in each dimension according to GOALS standards. We found that the participants' performance is enhanced in all four dimensions: depth perception ($p < 0.001$), bimanual dexterity ($p < 0.001$), efficiency ($p < 0.001$), tissue handling and autonomy ($p < 0.001$). For depth perception, the average score increased 1.8 (pre: $M = 4.60$, $SD = 1.27$, post: $M = 6.40$, $SD = 1.14$). For bimanual dexterity, the average score increased 0.88 (pre: $M = 5.30$, $SD = 1.14$, post: 6.18 , $SD = 1.48$). For efficiency, the average score increased 1.22 (pre: $M = 5.33$, $SD = 1.51$; post: 6.55 , $SD = 1.95$). For tissue handling and autonomy, the average score increased 1.27 (pre: $M = 4.78$, $SD = 1.56$; post: $M = 6.05$, $SD = 2.46$).

3.2 Flow Experience

The flow experience is related to moderate heart rate (HR) [45]. Thus, we investigated the flow experience of the participants through self-reported psychological flow questionnaires and heart rates. In Figs. 7 and 1, we could learn about the general change of heart rates between the pre-test and the post-test. We observe that the average heart rate decrease significantly ($p < 0.05$). Especially, the maximum heart rate decreases at post-test procedure ($p < 0.05$). Interestingly, the minimum heart rate did not increase significantly. The average heart rate of post-test fundamental surgery skill task even increases a little. We think this might demonstrate that the participants had a more steady heart rate change after VRLS training.

The results of the self-reported flow experience are shown in Table 2. We could find that the goals of tasks are clear (4.02/5), the participants had a positive attitude toward the whole experiment, and they had a great sense of involvement (4.19/5). From the aspect of flow dimension, the score is $3.95/5.0 \pm 0.96$ which is a relatively high score. We might conclude that the participants obtain a good experience during the whole experiment.

3.3 Cognitive Load

The overall cognitive load is measured using Pass scale [39] ($M = 6.09/10$, $SD = 1.32$). The overall task difficulty is relatively low ($M = 3.80/10$, $SD = 1.44$). Table 3 (NASA-TLX) shows the self-reported mental workload of our designed tasks. The overall cognitive load of 41 participants is lower than the midpoint of the full range (0 – 10). From the table, the mental demand ($M = 5.00$, $SD = 1.67$), physical demand ($M = 5.34$, $SD = 2.44$) and effort ($M = 6.00$, $SD = 1.48$) dimensions were the mainly components that affect participants' cognitive load. Most participants thought they could accomplish the task well ($M = 2.27$, $SD = 1.27$) and had a pleasant experience ($M = 3.05$, $SD = 1.80$). That means they were confident with their surgery skills after training.

Besides, we study the participants' cognitive load from five perception dimensions shown in Table 3 (WP Scale). The results demonstrated that the most important ability needed in laparoscopic surgery skills is depth perception ($M = 8.12$, $SD = 1.82$). The following requirements are visual processing ability ($M = 7.98$, $SD = 1.51$) and haptic sensing ability ($M = 7.80$, $SD = 1.86$). The audio processing ability is the least used perception capability ($M = 4.32$, $SD = 2.56$). And to complete the tasks, it requires considerable attention from participants ($M = 7.19$, $SD = 1.44$). In addition, we could compute the cognitive load from EEG [41]. As they pointed that the α and θ brain waves are more competitive than others when computing cognitive load. When calculating cognitive load, we only considered α and θ frequency spectrums. Figure 8 shows the cognitive load scores comparison between pre ($M = 0.17$, $SD = 0.11$) and post ($M = 0.14$, $SD = 0.09$) fundamental surgery skills tasks. The post cognitive load is significantly lower than cognitive load of Pre-FT ($p = 0.04 < 0.05$). Figure 9 shows the cognitive load scores are also significantly decreased in colon resection tasks ($p < 0.01$).

3.4 Correlation Analysis

Conrad et.al observed that there is a negative relationship between average EEG alpha and flow experience by conducting a pilot study [46]. However, the limiting participants in their study might not

result in a convincing conclusion. In our study, we tried to reveal the correlation between three kinds of scores.

Firstly, the relation between performance and corresponding cognitive load extracted from EEG is shown as Table 4. In four tasks, we found that the cognitive load has negative influence on the participants' performance. For the Pre-FT, the cognitive load score is negatively related with the performance score ($R^2 = 0.79$, $p = 0.1$). For the Post-FT task, it also shows negative relation but not significant ($R^2 = 0.74$, $p = 0.3$). For Pre-CRT, the cognitive load score is significantly negatively related with the performance score ($R^2 = 0.74$, $p < 0.001$). The negative relation is also shown in Post-CRT ($R^2 = 0.61$, $p = 0.05$).

Secondly, we evaluate the relation between the cognitive scores extracted from scales (CLS) and cognitive load scores computed from EEG (CLE). We found that CLS and CLE have positive relations. The regression R square of CLS and pre-colon resection CLE is 0.68, $p < 0.05$. The regression R square of CLS and post colon resection CLE is 0.70, $p < 0.1$. Furthermore, we explore the correlation between performance scores with flow experience. We found that there is no significant correlation between self-reported flow experience score ($p = 0.34 > 0.05$). The correlation between performance scores is also not significantly related to heart rates ($p = 0.06 > 0.05$).

4. Discussion

Training medical students on VRLS has been considered a promising direction since its relatively low cost, risk-free and high reusability. Besides, VRLS could simulate physiological phenomena (e.g. breathing) and provide interactive guides to increase goal clarity. In this study, we quantitatively explore the influence of VRLS on medical students from surgery skill acquisition and physical-psychological aspects. In addition, we also examine the correlation between them.

4.1 Surgery Skill Acquisition

The results of performance scores indicated that VRLS could significantly improve the acquisition of surgical skills. After training on VRLS, the participants could accomplish the same tasks using a shorter time. That means the proficiency of surgical skills has been substantially enhanced. At the pre-test phase, the worst-performing dimension is depth perception. After VRLS training, the depth perception becomes the best among the four dimensions. The least change dimension is bimanual dexterity.

4.2 Cognitive Load and Flow Experience

From the feedback of the participants, the principal source of cognitive load is the false perception of depth during the operation. This result is consistent with the observation in Sec. 4.1. Through the depth perception is significantly improved, it is still a major influence factor. Introducing an immersive training environment might be an alternative scheme. However, that could rise the dizziness problem and cause uncomfortable training experience [47].

The results demonstrate that the participants' surgery performance has a relation with their physical-psychological state. Developed skills might indicate lower cognitive load, moderate heart rate and flow experience. This provides us additional options to quantitatively measure one's task skills. We could evaluate one's performance by monitoring physiological data such as EEG, heart rate. Compared with the traditional subjective method evaluating after experiments, this approach could identify the features that increase the participants' cognitive load in real-time. When integrated with a physiological data detection device, the training course designer could optimize the experiment environment setting and the experiment procedure adjustment. Besides, when medical students suffer from a high cognitive load, the system could help them with guiding information.

However, the flow experience has no clear relation between heart rate and the performance of participants. We just observed a little negative relation between heart rate and flow experience score ($p = 0.06$). The improvement of skills proficiency may increase the automation of operation. Then the participants might regard our tasks more easily to achieve (frustration = 3.05 in NASA-TLX). This would hinder the acquirement of flow experience. We think the flow experience might a state that is not easy to obtain in our study. To obtain flow experience, we should optimize our training system with more attractive elements (e.g. guidance information with AR glass).

5. Conclusion

Training surgeons on VRLS has been considered a promising direction due to its relatively low cost, risk-free and high reusability. In this paper, we quantitatively investigate the influence of VRLS on medical students from three aspects: performance evaluation, physiology (heart rate and EEG) and self-reported cognitive load and flow experience. 41 Participants were recruited to conduct pre and post experiments in training boxes. In the middle of pre and post experiments, they were trained on VRLS. Their operation video and physiological data (heart rate and electroencephalogram) were recorded. Then their performance was graded by senior surgeons using a number of rules. Finally, the participants filled questionnaires about their cognitive load and flow experience. The experimental results demonstrate that the VRLS could highly improve medical students' performance and enable the participants to obtain flow experience with a lower cognitive load.

Nevertheless, our work is not without limits. Currently, we just reveal the correlation between performance and cognitive load. We have not investigated their exact functional relation such as linearity, non-linearity or exponential. Many researchers utilized machine learning to measure and classify one's cognitive load [48]. This could be a potential research topic in the future.

Abbreviations

Virtual reality (VR)

Virtual reality based laparoscopic surgery simulators (VRLS)

Global operative assessment of laparoscopic skills (GOALS)

Cognitive load (CL)

Cognitive load theory (CLT)

Pre/post-test of three fundamental surgery skills tasks (Pre/Post-FL)

Pre/post-test of colon resection tasks (Pre/Post-CRT)

Electroencephalography (EEG)

Cognitive scores extracted from scales (CLS)

Cognitive load scores computed from EEG (CLE)

Heart rate (HR)

Declarations

Ethics approval and consent to participate

Ethics approval for the conduct of the research was gained from the Beijing Normal University's Research Ethics Committee in November 2020. The participants (medical students) are volunteered to take part in our study. They granted us to freely distribute the results of the experiments. However, the identifiable data including personal images, physiological data are not allowed to publish. All information related to an identifiable person was anonymized.

Consent to publish

Yes, the participants and authors are consent to publish the content in this manuscript.

Availability of data and materials

The materials described in the manuscript, including all relevant experimental data, will be freely available to any scientist wishing to use them for non-commercial purposes, without breaching participant confidentiality. The analysis data is available in our manuscript. The raw data such as operation videos and physiological data involves much personal information. Thus, we couldn't publish it in the public.

Competing interests

Peng Yu, Junjun Pan, Zhaoxue Wang, Yang Shen, Jialun Li, Aimin Hao, Haipeng Wang have no conflicts of interest or financial ties to disclose.

Funding

National Key R&D Program of China (No. 2018YFC0115102).

National Natural Science Foundation of China Nos. 61872020, U20A20195).

Authors' Contributions

Peng Yu designed the experiments and analyzed the experiment data, wrote the article.

Junjun Pan took part in the design of this study and the writing of this manuscript.

Zhaoxue Wang took part in the design of the experiment, recruitment of participants, reference checking and raw data processing.

Yang Shen took part in the design of the experiment and reference checking.

Jialun Li participant in the conduction of this experiment and was responsible for arranging the experiment schedule.

Aimin Hao took part in the design of this experiment.

Haipeng Wang was in charge of the recruitment of participants and medical guidance during our experiment.

Acknowledgement

This research is supported by National Key R&D Program of China (No. 2018YFC0115102), National Natural Science Foundation of China Nos. 61872020, U20A20195). We also thank the Faculty of Media and Communication, Bournemouth University (UK) with its support of Global Visiting Fellowship for Dr. Junjun Pan.

References

1. D. Ota, B. Loftin, T. Saito, R. Lea, and J. Keller (1995). Virtual reality in surgical education. *Comput Biol Med*, 25(2):127–137.
2. E.-Y. Jung, D. K. Park, Y. H. Lee, H. S. Jo, Y. S. Lim, and R. W. Park (2012). Evaluation of practical exercises using an intravenous simulator incorporating virtual reality and haptics device technologies. *Nurse Educ Today*, 32(4):458–463.

3. E. Yiannakopoulou, N. Nikiteas, D. Perrea, and C. Tsigris (2015). Virtual reality simulators and training in laparoscopic surgery. *Int J Surg*, 13:60–64.
4. T. Grantcharov, J. Rosenberg, E. Pahle, and P. Fench-Jensen (2008). Virtual reality computer simulation. *Surg Endosc*, 15(3):242–244.
5. T. P. Grantcharov, L. Bardram, P. Funch-Jensen, and J. Rosenberg (2003). Learning curves and impact of previous operative experience on performance on a virtual reality simulator to test laparoscopic surgical skills. *Ame J Surg*, 185(2):146–149.
6. S. Suebnukarn, P. Haddawy, P. Rhiemora, P. Jittimane, and P. Viratket (2010). Augmented kinematic feedback from haptic virtual reality for dental skill acquisition. *J Dent Educ*, 74(12):1357–1366.
7. M. M. Keehner, F. Tendick, M. V. Meng, H. P. Anwar, M. Hegarty, M. L. Stoller, and Q.-Y. Duh (2004). Spatial ability, experience, and skill in laparoscopic surgery. *Ame J Surg*, 188(1):71–75.
8. J. Sweller (1988). Cognitive load during problem solving: Effects on learning. *Cogn Sci*, 12(2):257–285.
9. S. A. W. Andersen, L. Konge, and M. S. Sørensen (2018). The effect of distributed virtual reality simulation training on cognitive load during subsequent dissection training. *Med Teach*, 40(7):684–689.
10. R. Bharathan, S. Vali, T. Setchell, T. Miskry, A. Darzi, and R. Aggarwal (2013). Psychomotor skills and cognitive load training on a virtual reality laparoscopic simulator for tubal surgery is effective. *Eur J Obst Gyn Reprod Bio*, 169(2):347–352.
11. M. Csikszentmihalyi and M. Csikszentmihalyi (1990). *Flow: The psychology of optimal experience*. Harper & Row New York.
12. L. Ahlborg, M. Weurlander, L. Hedman, H. Nisell, P. G. Lindqvist, L. Fellaänder-Tsai, and L. Enochsson (2015). Individualized feedback during simulated laparoscopic training: a mixed methods study. *Int J Med Educ*. 6–93.
13. J. Cheon and M. M. Grant (2012). The effects of metaphorical interface on germane cognitive load in web-based instruction. *Educ Technol Res Dev*, 60(3):399–420.
14. R. Zheng, M. McAlack, B. Wilmes, P. Kohler-Evans, and J. Williamson (2009). Effects of multimedia on cognitive load, self-efficacy, and multiple rule-based problem solving. *Br J Educ Technol*, 40(5):790–803.
15. J. Keller, H. Bless, F. Blomann, and D. Kleinböhl (2011). Physiological aspects of flow experiences: Skills-demand-compatibility effects on heart rate variability and salivary cortisol. *J Exp Soc Psychol*, 47(4):849–852.
16. Y. Q. Qiao, J. Shen, X. Liang, S. Ding, F. Y. Chen, L. Shao, Q. Zheng, and Z. H. Ran (2014). Using cognitive theory to facilitate medical education. *BMC Med Educ*, 14(1):1–7.
17. L. M. Naismith, J. J. Cheung, C. Ringsted, and R. B (2015). Cavalcanti. Limitations of subjective cognitive load measures in simulation-based procedural training. *Med Educ*, 49(8):805–814.

18. F. Paas, A. Renkl, and J. Sweller (2003). Cognitive load theory and instructional design: Recent developments. *Educ Psychol*, 38(1):1– 4.
19. F. G. Paas and J. J. Van Merriënboer (1993). The efficiency of instructional conditions: An approach to combine mental effort and performance measures. *Hum Factors*, 35(4):737–743.
20. J. E. Tuovinen and F. Paas (2004). Exploring multidimensional approaches to the efficiency of instructional conditions. *Instr Sci*, 32(1-2):133–152.
21. G. B. Reid and T. E. Nygren (1988). The subjective workload assessment technique: A scaling procedure for measuring mental workload. *Adv Psychol*, vol. 52, pp. 185–218.
22. S. G. Hart and L. E. Staveland (1988). Development of nasa-tlx (task load index): Results of empirical and theoretical research. *Adv Psychol*, vol. 52, 139–183.
23. A. C. Dirican and M. Göktürk (2011). Psychophysiological measures of human cognitive states applied in human computer interaction. *Proce Comp Sci*, 3:1361–1367.
24. E. Başar (2012). Brain function and oscillations: volume II: integrative brain function. *Neurophysiol Cog Proc*.
25. W. Klimesch, B. Schack, and P. Sauseng (2005). The functional significance of theta and upper alpha oscillations. *Exp Psychol*, 52(2):99–108.
26. A. Gevins, M. E. Smith, L. McEvoy, and D. Yu (2001). High-resolution eeg mapping of cortical activation related to working memory: effects of task difficulty, type of processing, and practice. *Cereb Cortex*, 7(4):374–385, 1997.
27. L. K. McEvoy, M. E. Smith, and A. Gevins (1998). Dynamic cortical networks of verbal and spatial working memory: effects of memory load and task practice. *Cereb Cortex*, 8(7):563–574.
28. M. B. Serman, C. A. Mann, D. A. Kaiser, and B. Y. Suyenobu (1994). Multi-band topographic eeg analysis of a simulated visuomotor aviation task. *Int J Psychophysiol*, 16(1):49–56.
29. L. Phillips, V. Interrante, M. Kaeding, B. Ries, and L. Anderson (2012). Correlations between physiological response, gait, personality, and presence in immersive virtual environments. *Presence: Teleop Virt Env*, 21(2):119–141.
30. D.-H. Shin, F. Biocca, and H. Choo (2013). Exploring the user experience of three-dimensional virtual learning environments. *Behav Info Technol*, 32(2):203–214.
31. S. A. Jackson and R. C. Eklund. Assessing flow in physical activity: the flow state scale-2 and dispositional flow scale-2. *J Sport Exe Psychol*, 24(2).
32. L.-K. Cheng, M.-H. Chieng, and W.-H. Chieng (2014). Measuring virtual experience in a three-dimensional virtual reality interactive simulator environment: a structural equation modeling approach. *VR*, 18(3):173–188.
33. B. Cowley, D. Charles, M. Black, and R. Hickey (2008). Toward an understanding of flow in video games. *CIE*, 6(2):1–27.
34. F.-L. Fu, R.-C. Su, and S.-C. Yu (2009). Egameflow: A scale to measure learners' enjoyment of e-learning games. *Comp Educ*, 52(1):101–112.

35. F. G. De Kock (2014). The neuropsychological measure (EEG) of flow under conditions of peak performance. PhD thesis, University of South Africa.
36. M. Vassiliou, L. Feldman, C. Andrew, S. Bergman, K. Leffondre, D. Stanbridge, and G. Fried (2005). A global assessment tool for evaluation of intraoperative laparoscopic skills. *Am J Surg*, 190:107–13.
37. F. M. Garcia-Moreno, M. Bermudez-Edo, M. J. Rodriguez-Fortiz, and J. L. Garrido (2020). A CNN-LSTM Deep Learning Classifier for Motor Imagery EEG Detection Using a Low-invasive and Low-Cost BCI Headband. *Int Conf Intell Environ*, 84–91.
38. O. E. Krigolson, C. C. Williams, A. Norton, C. D. Hassall, and F. L. Colino (2017). Choosing muse: Validation of a low-cost, portable eeg system for erp research. *Front Neuro*, 11:109.
39. Pass and Fred (1992). Training strategies for attaining transfer of problem solving skills in statistics: a cognitive-load approach. *J Educ Psychol*, 84(4):429–434.
40. S. Hart (2006). Nasa-task load index (nasa-tlx); 20 years later. *Proc Hum Factors Ergon Soc Annual Meet*, 50:904 –908.
41. P. Tsang and V. VELAZQUEZ (1996). Diagnosticity and multidimensional subjective workload ratings. *Ergonomics*, 39:358–81.
42. W. C. LiKeng Cheng, MingHua Chieng (2014). Measuring virtual experience in a three-dimensional virtual reality interactive simulator environment: a structural equation modeling approach. *Virtual Reality*, 18 (9):173–188.
43. R. H. Morton, J. R. Fitz-Clarke, and E. W. Banister (1990). Modeling human performance in running. *J Appl Physiol*, 69(3):1171– 1177.
44. E. W. Anderson, K. C. Potter, L. E. Matzen, J. F. Shepherd, G. A. Preston, and C. T. Silva (2011). A user study of visualization effectiveness using eeg and cognitive load. *CGF*, 30: 791–800.
45. Y. Tian, Y. Bian, P. Han, P. Wang, F. Gao, and Y. Chen (2017). Physiological signal analysis for evaluating flow during playing of computer games of varying difficulty. *Front Psychol*, 8:1121.
46. C. Conrad and M. Bliemel (2016). Psychophysiological measures of cognitive absorption and cognitive load in e-learning applications. *Int Conf Info Syst*.
47. M. Li, S. Ganni, J. Ponten, A. Albayrak, A. Rutkowski, and J. Jaki-mowicz (2020). Analysing usability and presence of a virtual reality operating room (vor) simulator during laparoscopic surgery training. *IEE VR*, 566–572.
48. C. Herff, O. Fortmann, C. Tse, X. Cheng, F. Putze, D. Heger, and T. Schultz (2015). Hybrid fnirs-eeg based discrimination of 5 levels of memory load. *NER*, 5–8.

Tables

Table 1

The statistical results of heart rates (HR) for two kinds of experiments' pre-test and post-test respectively.

	Avg Time(s)	Avg HR	Avg HR SD	Min HR	Min HR SD	Max HR	Max HR SD
Pre-FT	1316.93	94.96	11.14	79.40	10.24	112.33	10.65
Pre-CRT	292.07	96.51	11.48	85.49	10.39	107.30	12.05
Post-FT	844.31	92.71	11.84	81.30	10.35	108.10	12.52
Post-CRT	155.62	92.67	11.67	85.10	10.43	101.26	12.70

Table 2

Self-reported flow scale after all experiments (1-5, the higher score means higher flow experience).

	Mean Value	SD
VR Interactivity	3.85	0.89
Challenge	3.55	1.25
Skill	3.68	0.92
Goal Clarity	4.02	0.78
Knowledge Improvement	3.72	1.03
Positive Affect	4.29	0.77
Vividness	3.15	1.13
Involvement	4.19	0.75
Flow	3.95	0.96

Table 3

Self-reported mental workload (NASA-TLX scale) after all experiments (0-10, the higher score means higher mental workload)

	Measure Dimension	Mean	SD
NASA-TLX	Mental Demand	5.00	1.67
	Physical Demand	5.34	2.44
	Temporal Demand	4.07	2.16
	Performance	2.27	1.27
	Effort	6.00	1.48
	Frustration	3.05	1.80
WP Scale [46]	Attention	7.19	1.44
	Depth Perception Ability	8.12	1.82
	Visual Processing Ability	7.98	1.51
	Haptic Sensing Ability	7.80	1.86
	Audio Processing Ability	4.32	2.56

Table 4

The correlation of performance and cognitive load computed by EEG data. (The downward arrow ↓ means negative correlation, **: 0.01 level significant, *: 0.05 level significant, NA: Not Available)

	Pre-test Performance	Post-test Performance
Pre-FT	↓ *	NA
Pre-CRT	↓ **	NA
Post-FT	NA	↓
Post-CRT	NA	↓**

Figures



Figure 1

The participants operated on a training box during pre-test and post-test (Left). The heart rate and EEG were recorded and quantitatively analyzed. During pre-test and post-test, they were trained on VRLS (Middle: screenshot of picking small balls. Right: operation illustration).

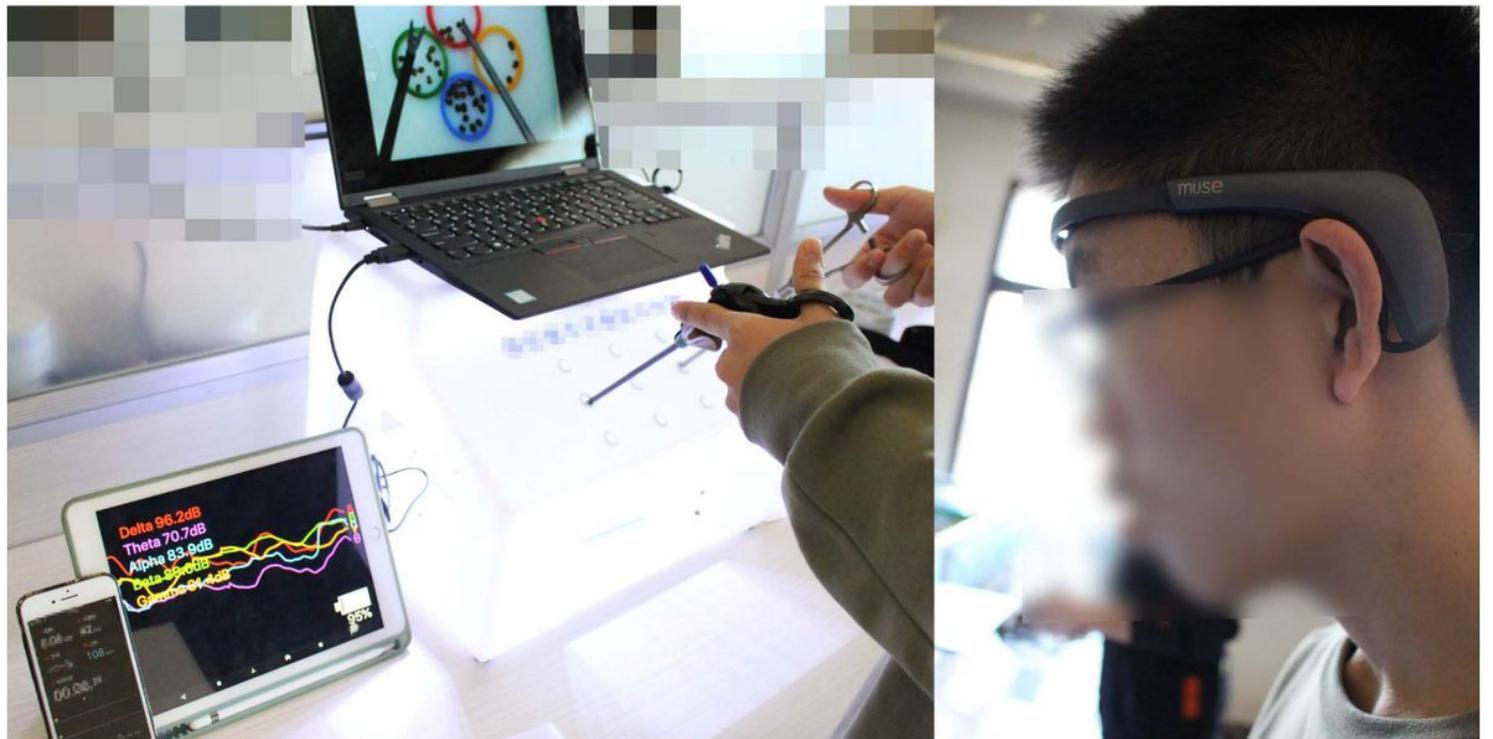


Figure 2

(Left) The participants are training on a box. The training process is recorded and saved as videos. The EEG is recorded using an iPad and the heart rate is recorded using an iPhone. (Right) A participant wore a Muse 2 headband to measure his EEG when operating the surgery skill task.



Figure 3

Three fundamental surgery tasks including peg transfer, picking beans a threading skill practice (Left) and the colon resection task (Right).

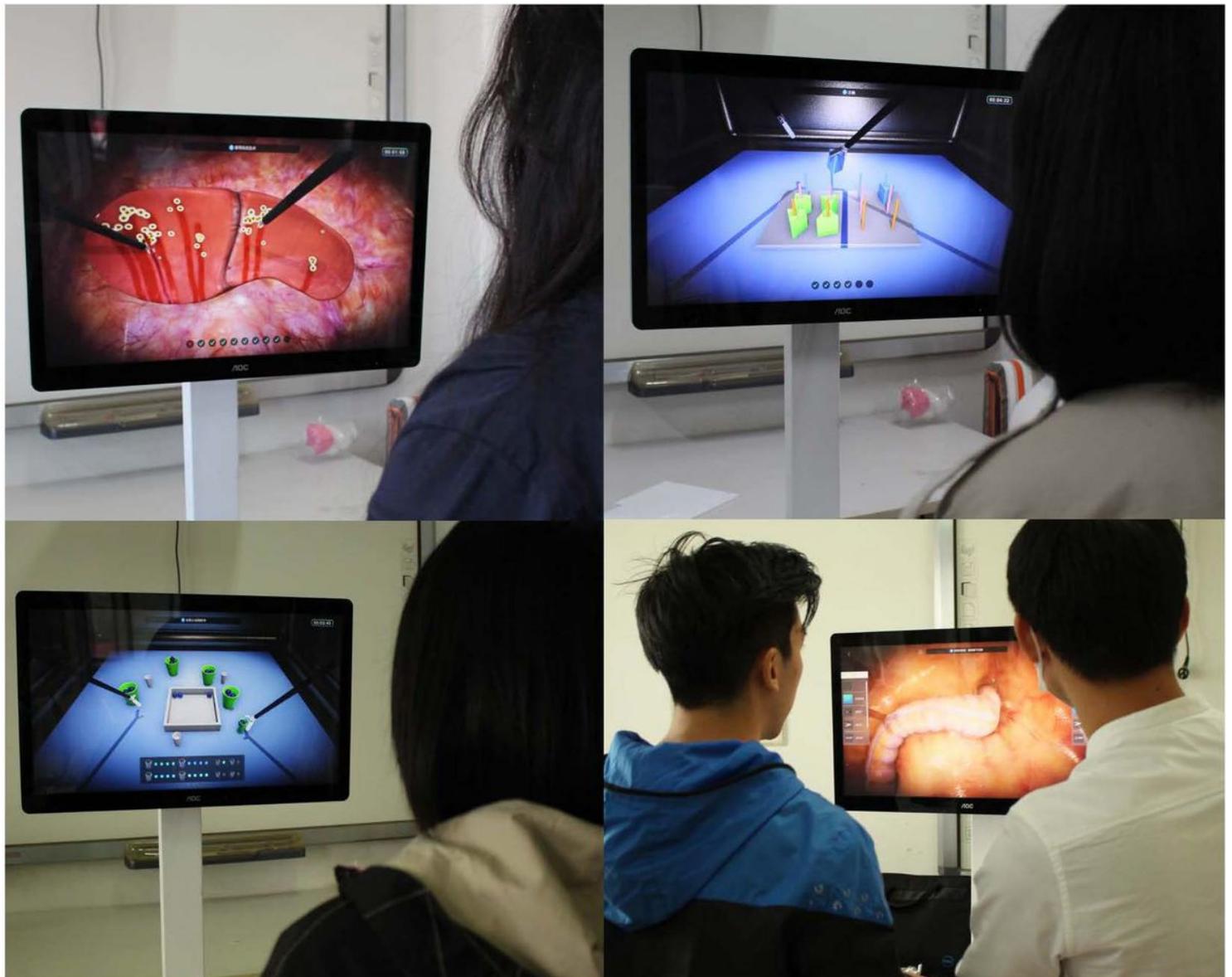


Figure 4

Training on VR laparoscopic surgery simulator.

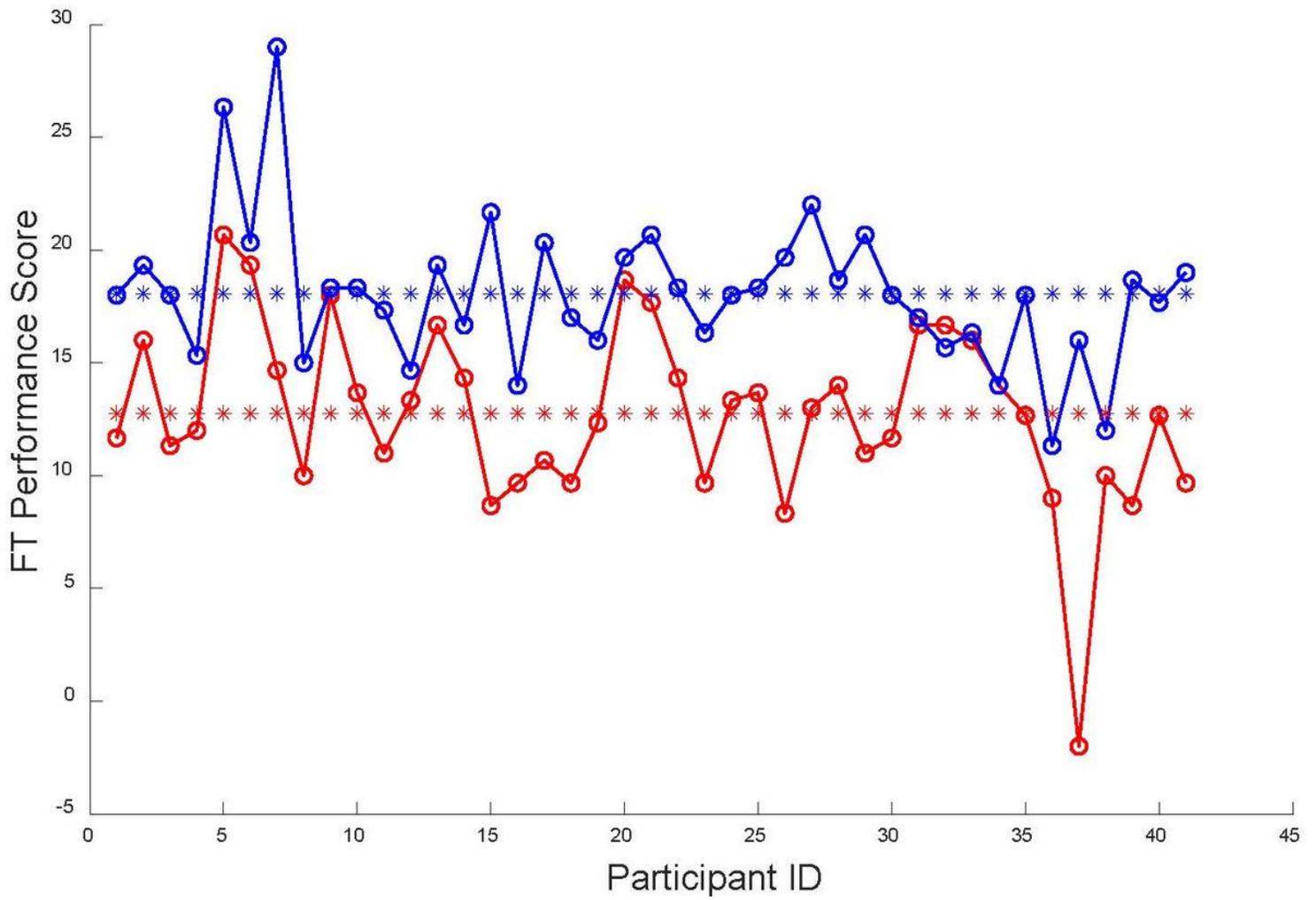


Figure 5

Pre-FT and Post-FT performance comparison. Pre-FT: 12:76_3:96, Post-FT: 18:07_3:24; $p < 0.01$.

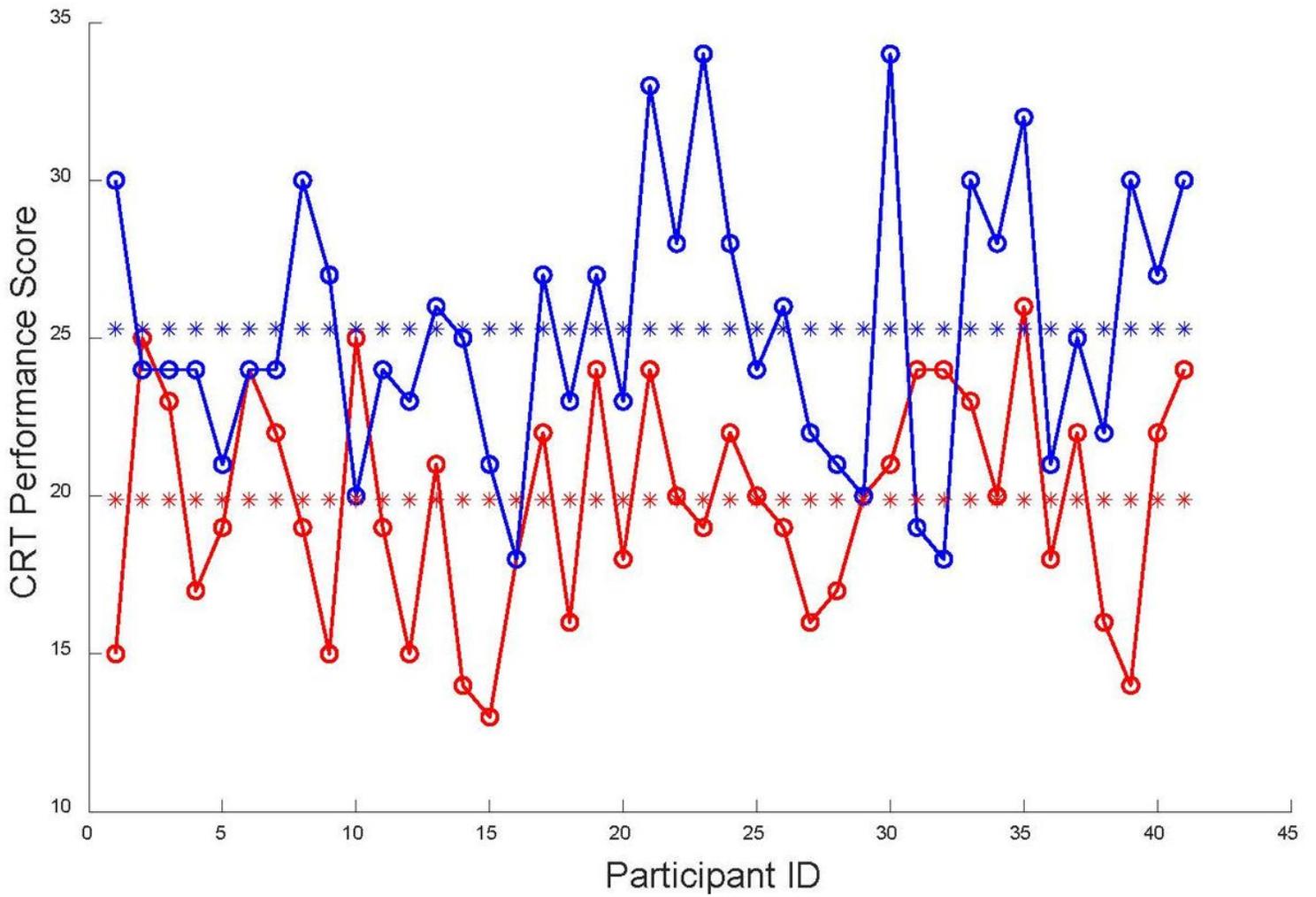


Figure 6

Pre-CRT and Post-CRT performance comparison. Pre-CRT: 19:88_12:51, Post-CRT: 25:30_18:16; $p = 4.0E-08 < 0.01$.

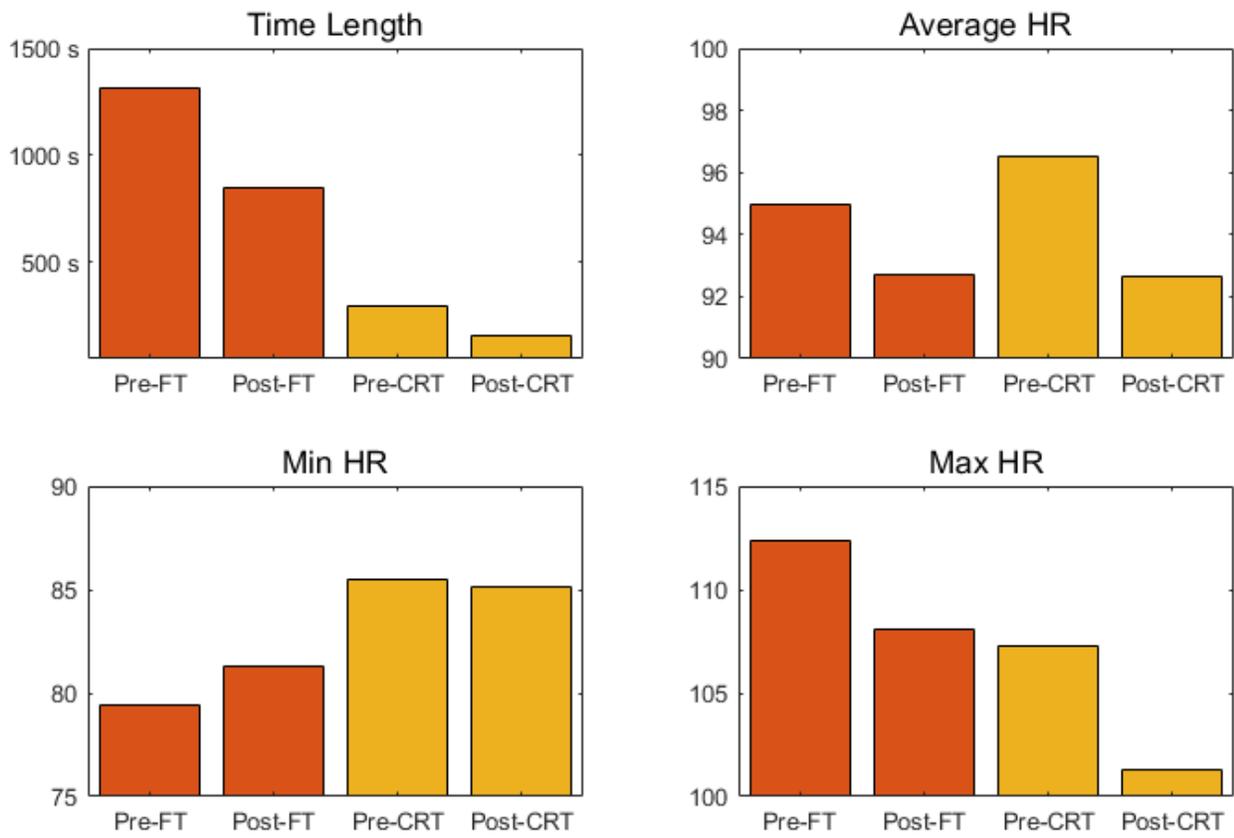


Figure 7

Heart rates statistical results comparison.

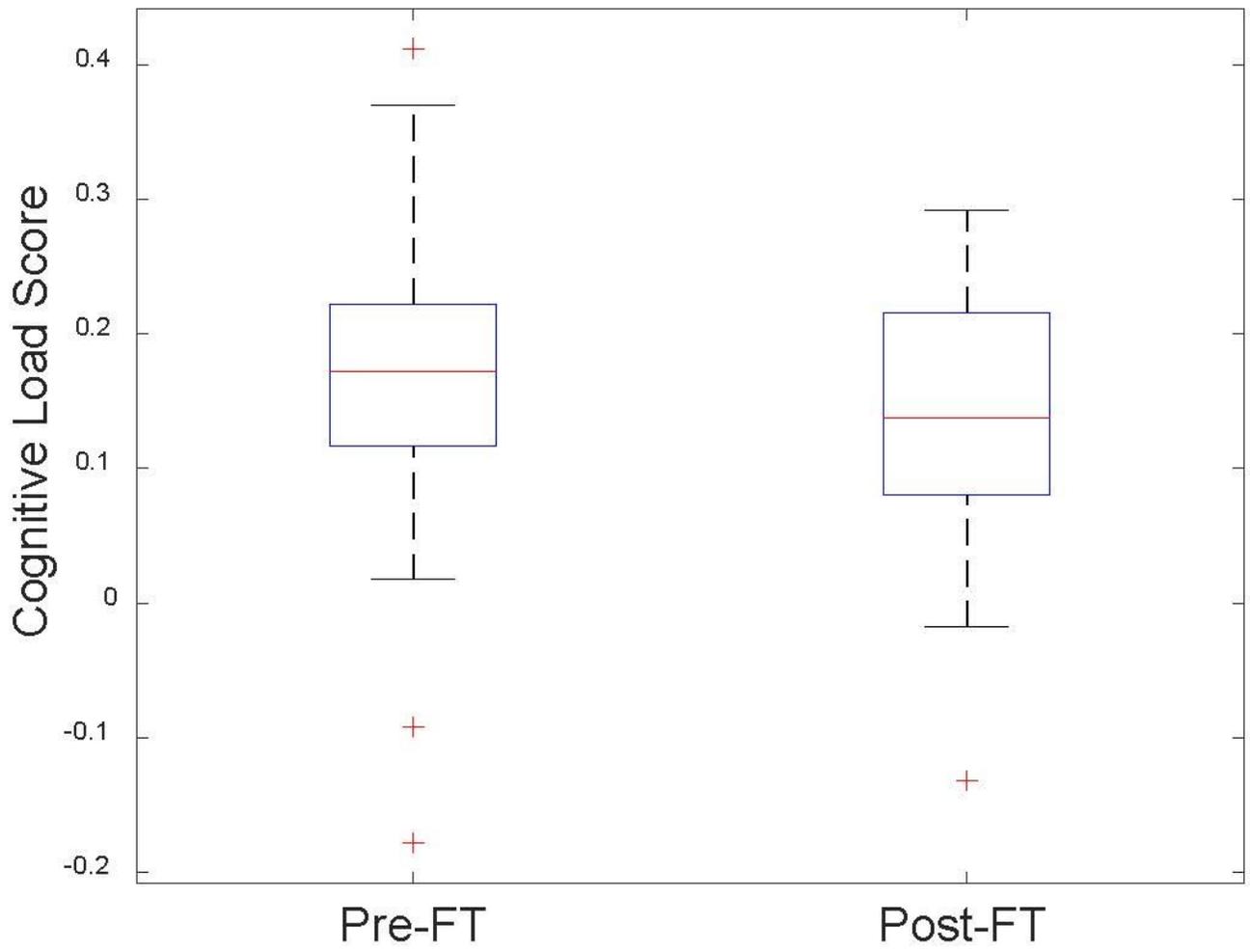


Figure 8

Pre-FT and Post-FT cognitive load score comparison. Pre-FT: 0:17_0:11, Post-FT:0:14_0:09, t-test: $p = 0.04 < 0.05$.

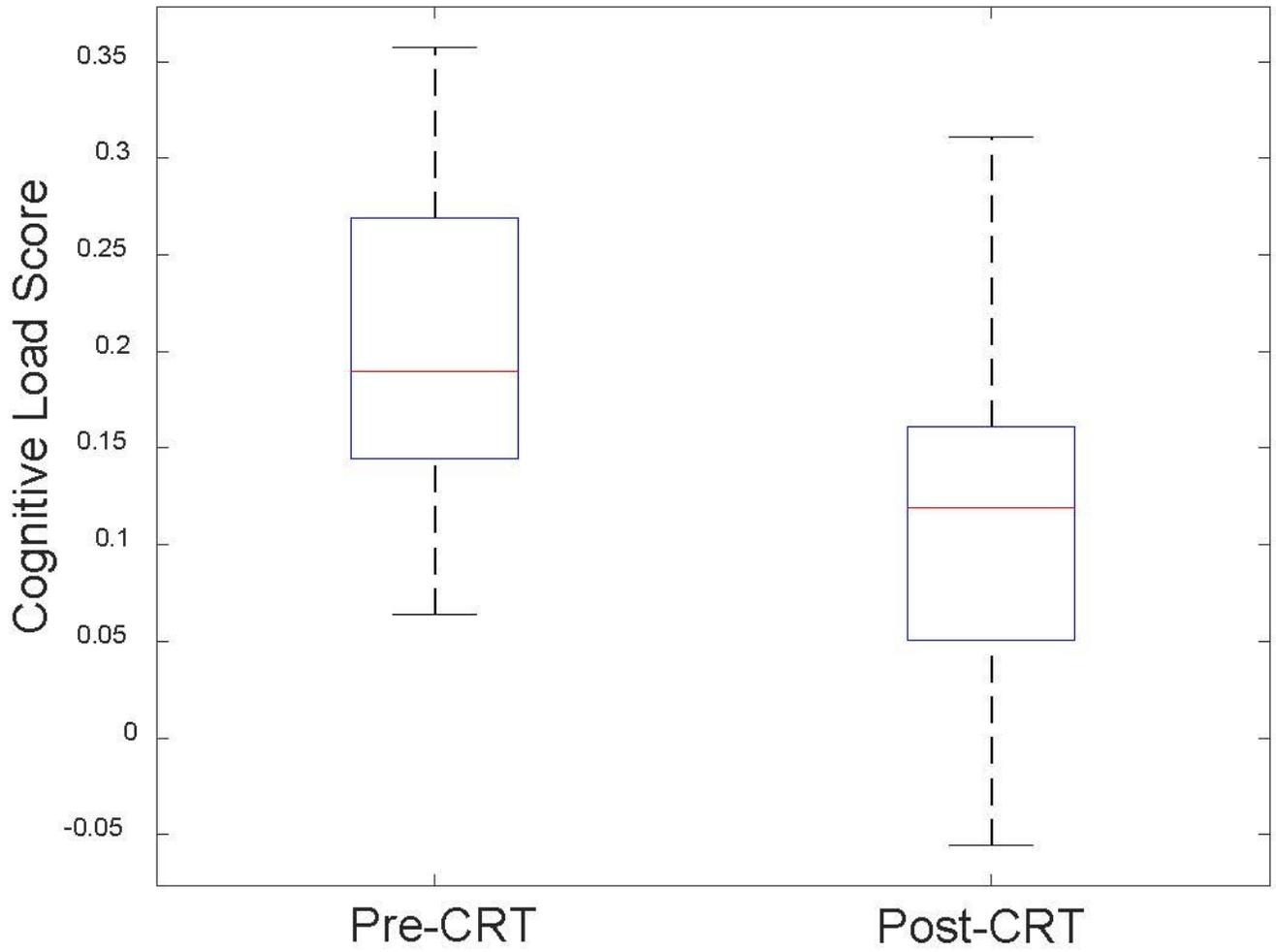


Figure 9

Pre-CRT and Post-CRT cognitive load score comparison. Pre: 0:19_0:01, post:0:11_0:01, t-test: $p = 0.0004 < 0.05$.