Face and Construct Validation of a Next Generation Virtual Reality (Gen2-VR©) Surgical Simulator

Ganesh Sankaranarayanan[a], **Baichun Li**[a,b], **Kelly Manser**[c], **Stephanie B. Jones**[d], **Daniel B. Jones**[d], **Steven Schwaitzberg**[c], **Caroline G. L. Cao**[e], and **Suvranu De**[a]

[a]Center for Modeling, Simulation and Imaging in Medicine, Rensselaer Polytechnic Institute, Troy, NY, USA

[b]School of Mechanical Engineering & Automation, Northeastern University, Shenyang, PRC

[c]Cambridge Health Alliance, Cambridge, MA, USA

[d]Beth Israel Deaconess Medical Center, Boston, MA, USA

[e]Wright State University, Dayton, OH, USA

Abstract

Introduction—Surgical performance is affected by distractors and interruptions to surgical workflow that exist in the operating room. However, traditional surgical simulators are used to train surgeons in a skills lab that does not recreate these conditions. To overcome this limitation, we have developed a novel, immersive virtual reality (Gen2-VR©) system to train surgeons in these environments. This study was to establish face and construct validity of our system.

Methods and Procedures—The study was a within-subjects design, with subjects repeating a virtual peg transfer task under three different conditions: **CASE I:** traditional VR; **CASE II:** Gen2-VR© with no distractions and **CASE III**: Gen2-VR© with distractions and interruptions.. In Case III, to simulate the effects of distractions and interruptions, music was played intermittently, the camera lens was fogged for 10 seconds and tools malfunctioned for 15 seconds at random points in time during the simulation. At the completion of the study subjects filled in a 5-point Likert scale feedback questionnaire. A total of sixteen subjects participated in this study.

Results—Friedman test showed significant difference in scores between the three conditions ($p < 0.0001$). Post hoc analysis using Wilcoxon Signed Rank tests with Bonferroni correction further showed that all the three conditions were significantly different from each other (Case I, Case II, p

< 0.001), (Case I, Case III, p < 0.001) and (Case II, Case III, p = 0.009). Subjects rated that fog (mean= 4.18) and tool malfunction (median = 4.56) significantly hindered their performance.

Conclusion—The results showed that Gen2-VR[©] simulator has both face and construct validity and it can accurately and realistically present distractions and interruptions in a simulated OR, in spite of limitations of the current HMD hardware technology.

Keywords

Immersive Virtual Reality; Surgery Simulator; Head Mounted Display; Cognitive Simulator; Face and Construct validation; Gen2-VR[©]

Introduction

Existing surgical simulators, both part-task (1,2) and procedural (3), are primarily focused on the development of psychomotor skills, i.e., training for hand-eye coordination and motor skills necessary for tasks such as tool movement, cutting, suturing, etc. The simulators offer little training in cognitive skills (4) associated with higher level mental functions related to workload management, planning, communication, decision-making and problem-solving. Moreover, these "Gen1" simulators are used in a laboratory setting that hardly represents the actual conditions in an operating room (OR).

There are multiple factors within the OR that influence the performance of a surgeon. Distractions and interruptions that arise from the sounds of pagers, phones, doors opening and closing, random and patient relevant conversations, and frequent movement behind the video monitor are known to affect the attention of the surgeons in the OR (5–7). Feuerbacher et al. (8) has shown that in the presence of distractions and interruptions, the error rates of subjects performing a simulated laparoscopic cholecystectomy procedure in the Lapsim simulator were significantly higher. Moreover, patient issues such as sudden drop in blood pressure or unexpected bleeding have to be dealt with within the stressful environment of the OR. In order to train in this realistic environment, the surgical simulator should not only be capable of providing realistic simulation but should also be able to immerse the subjects in this stressful environment. This is supported by modern adult learning theories, such as the "situated learning" model of Lave and Wenger (9) that argue that learning is best achieved by actually engaging in the process in its proper social context. To overcome the limitations of traditional Gen1 surgical simulators, we have developed a novel fully immersive next generation virtual reality (Gen2-VR[©]) system. In the Gen2-VR[©] system, the user is immersed in a virtual model of the OR using a head mounted display (HMD) and interacts with a virtual patient. Haptic interface devices are used to provide force feedback. It is expected that such a system will provide the opportunity to generate realistic conditions of the OR in a controlled manner and aid in both cognitive and psychomotor skill training. While HMDs have been used to generate interactive ORs (10), they have not been utilized for immersive operating environments with integrated force feedback for surgical simulation. Using a Gen1 simulator inside a mock OR (11) is an alternative, but requires expensive physical facilities and dedicated staff time. In this paper we present face and construct validation of the Gen2-VR[©] system.

Methods

In this study we have used a virtual peg transfer simulator (VBLaST-PT$^©$) which has been shown to reproduce the corresponding task in the Fundamentals of Laparoscopic Surgery (FLS) toolbox (2,12–14). The VBLaST-PT$^©$ has both hardware and software components. The hardware consists of two force feedback Phantom devices with surgical tool interfaces. The software is based on physics-based interactive simulation technology developed by our group. Further technical details regarding the VBLaST-PT$^©$ simulator may be found in reference (12). The VBLaST-PT$^©$ was originally developed as a Gen1 simulator with a computer monitor for visual display. To convert the Gen1 simulator to the Gen2-VR system, the user wears a VR2000 HMD from Virtual Realities Inc. and interacts with the haptic interface as shown in the left hand side image of Figure 1. The HMD effectively replaces the computer monitor. The software technology has been altered to immerse the user in a virtual room standing in front of a virtual monitor, interacting with the virtual pegs, as shown in the right hand side image of Figure 1. The virtual avatar in the scene represents the user. It is to be noted that even though the user sees and experiences everything in 3D with the HMD, the surgical task displayed on the monitor is 2D. The computer monitor seen in Figure 1 is seen by the user only in the Gen1 mode. It is also used by the research team member to monitor the progress of the study. While the user is immersed in the Gen2-VR$^©$ simulator, both distractions and interruptions are generated. Distraction can be both visual (movement behind monitor) and auditory (music, real sounds of telephone ringing, door opening and closing) in nature. Interruptions are in the form of tool malfunction and camera lens fogging. The movement of the user's head while wearing the HMD was tracked using the in-built inertial tracker in the VR2000 HMD. A standard stereo headphone was used to play auditory distractions at a constant level of 50 dB (A). An USB foot pedal was used to lock the camera view so that the frequent movements of the head while performing the procedure do not change the view.

Experiment Design

The Gen2-VR$^©$ simulator was tested for face and construct validity. For this institutional review board approved study, subjects were recruited at the annual Minimally Invasive Surgery Fellowship Conference held in August 2013 in Sedona, AZ and at the Beth Israel Deaconess Medical Center, Boston, MA. The study was a within-subjects design, with the same subjects repeating the peg transfer task twice under three different conditions:

CASE I: traditional Gen1 VR, i.e. user interacts with simulation scenario presented on a computer monitor.

CASE II: Gen2-VR $^©$ with the user interacting with the simulation scenario within a HMD, but without distractions and interruptions.

CASE III: Gen2-VR$^©$ with the user interacting with the simulation scenario within a HMD with distractions and interruptions.

After signing the consent form and completing a demographic questionnaire, the subjects were shown an instructional video of the VBLaST-PT$^©$ that explained the peg transfer task. After watching the video, the subjects were given one practice trial on the VBLaST-PT$^©$

(CASE I) simulator to get accustomed to the system. They then proceeded to perform the task twice on three difference conditions. The order in which the three conditions were presented was randomized. Their performance score on the peg transfer task was automatically calculated by the simulator and recorded for analysis. After completing the procedure, the subjects filled in a subjective feedback questionnaire on a 5-point Liker scale to evaluate both the realism of the representative task, immersive experience and the effects of the distractions and interruptions.

In Case III, to simulate the effects of distractions and interruptions, one auditory and two interruptions were presented while the subjects performed the task. The auditory distraction was music selected from a list of five genres listed in Table 1. Subjects were asked to rate the five genres of music in the pre-questionnaire with 1 being the most and 5 being the least preferred. The songs for each category were selected based on the highest views for each of the category on the YouTube video sharing website (Google Inc.). The least rated music category was then played intermittently while the subjects performed the study. The reason for not playing the music continuously is to prevent the subjects from becoming acclimated to the music. For interruption to the procedure, the camera lens was fogged for 10 seconds (Figure 2) and tool tip opening/closing was disabled for 15 seconds at random times within the simulation. For disabling the tool function, both right and left tools were used at random. Figure 3 shows the sequence of the auditory distraction and interruptions applied for a subject. It is to be noted that during the calculation of the normalized VBLaST-PT$^©$ score, the duration of the interruptions were subtracted from the total time so that only the time spent on the actual task was used for the final score.

Data Analysis

For data analysis, the IBM SPSS 22 software was used. Descriptive statistics were first computed for all the three conditions. Due to smaller sample size, non-parametric tests were used for analysis. In order to test the effect of the three conditions on the performance of the peg transfer task, the Friedman test was used. Post hoc multiple comparisons between the three conditions were performed using the Wilcoxon Signed Rank test with Bonferroni correction. Since there were three post hoc comparisons, a p value less than 0.017 was considered significant.

Results

Demographics

A total of 16 subjects (11 males and 5 females) participated in this study. 14 subjects were right-handed and two subjects were left-handed. All subjects were FLS certified. A total of 8 fellows, 4 attending and 4 PGY5's were recruited. Of the total 16 subjects, six were recruited at the MIS Fellowship Conference and the rest at the Beth Israel Deaconess Medical Center. Surgical experience of subjects post residency ranged from less than a year to more than 15 years.

Face Validity

The mean and standard deviation of the subjects rating on a 5-point Likert scale on the various aspects of the Gen2-VR© is shown in Table 2. The face validity results were tabulated under two different categories. The first category is on the rating of the VBLaST-PT© system including its simulation software and its hardware. The second is on the feedback of the immersive experience in the Gen2-VR© simulator and the effects of distractions and interruptions. All questions relateing to the VBLaST-PT© system received a score above 3 on a 5-point Likert scale with the highest mean score of 3.812 for the realism of the peg transfer tasks in the VBLaST-PT© system followed by the question on the usefulness of the simulation in learning ambidexterity skills which got a mean score of 3.625. For the second category, subject rated both interruptions very high in their ability to affect their performance with a mean score of 4.562 and 4.375 for tool malfunction and fogging, respectively. For the question on difference in performing the task between the VBLaST-PT© and the Gen2-VR© simulator, the mean rating was 4.375. The effect of music on performance received the lowest mean rating of 2.0. Subject's means rating for the questions related to immersion and quality of HMD were between 2 and 3.

Construct Validity

Figure 4 shows the box plot of the normalized score for the three conditions. The median scores for CASE I, CASE II and CASE III were 83.38, 73.25 and 67.37 respectively. The Friedman test on the data showed a significant effect on the conditions ($p < 0.0001$). Table 3 shows the results from the post hoc comparisons using the Wilcoxon Signed Rank test.

Checking for significance after Bonferroni correction ($p < 0.017$), all three cases, showed statistically significant difference indicating performances differed significantly under all the three conditions.

Discussion and conclusions

Though performing surgery requires focus, attention and fine motor skills, the operating room environment is typically noisy with intermittent sounds of up to 108 dB from machines, patient relevant and irrelevant conversations and alarms (15). Hasfeltdt et al. (16) reviewed several works that reported on noise in the operating room and reported that the sounds ranged from 51–75 dB(A) with peak noise levels of 80–119 dB(A).

The performance of surgeon in an operating room is affected by distractions that arise from noises and interruptions to the surgical workflow. Healy et al. (5) defined the distractions as break in attention and interruptions as break in task activity. In their study on observing 50 cases, both open and laparoscopic, in an operating theatre (29 laparoscopic and 21 open), they observed more interference in laparoscopic surgery (0.35/min) compared to open surgery (0.18). Equipment (laparoscopic tool not working, instrument not at hand, endoscope not working etc.) and work environment (misting of the camera lens, poor image focus, spotlight needs redirection etc.) related issues were most frequent in laparoscopic cases resulting in higher interference. Zheng et al. (6) did a study of disruption in the operating room during laparoscopic antireflux surgery. The disruptive activities recorded were: instrument change, position change between surgeons, nursing shift change,

conversation, phone or pager answering and extraneous interruptions. For a total of 12 cases that were observed, they found that the most common event that significantly suspended surgical workflow was instrument change during the procedure. In a study using the Xitact LC 3.0 simulator, Pluyter at al. (17) studied the effect of distractions arising from music, conversation and non optimal handling of the scope on 12 interns performing a laparoscopic cholecystectomy task. Overall, the task performance measured in terms of task completion time, score and error decreased significantly with distractions. In another study on endourological procedures, Persoon et al. (18) observed 78 cases and reported that distractions occur frequently (one event every 1.8 minutes). They also reported equipment problems and procedure related and medically irrelevant communications as the leading contributors for the distractions. Performance degradation due to the presence of distractions in robot-assisted laparoscopic surgery procedures have also been reported (19).

The effect of music on the performance of surgical task has been studied by many researchers (20–24). The studies have shown varying results with Moorthy et al. (22) showing no effect of music and Siu et al. (24) showing improvement in robot-assisted laparoscopic procedure.

Goodel et al. (25) studied the effect of cognitive overload while performing surgical tasks. Subjects performed six tasks on the MIST-VR simulator while simultaneously performing mental arithmetic problems. Their study showed that the cognitive overload increased completion time in the MIST-VR tasks. Hsu et al. (26) performed a similar study on subjects performing the peg transfer task of the Fundamentals of Laparoscopic Surgery (FLS) tasks. Solving a set of five mathematical addition problems were added along with the task for cognitive overload. Their study showed that the ability to do addition problems decreased significantly for the novice subjects, whereas experienced surgeons performed both tasks without any significant change in performance. The study showed that experienced surgeons were able to overcome the bottleneck of the psychological refractory period (PRP) (27) with the achievement of greater task automaticity.

In this study, for the first time, an immersive virtual reality simulator was used to reproduce the distraction and interruptions that occur in an operating room environment. The camera lens fogging and tool malfunctions were selected as the interruptions based on observations reported in the literature. Music was selected as the distraction since it is present most of the time in the operating room environment. The music from the least preferred category was played in order to increase the effect of distraction. The construct validity results showed that performance on the peg transfer task on the VBLaST-PT$^©$ simulator decreased with added distraction and interruptions. It can also be seen that performing the task with HMD (CASE II) had a significant effect on the subjects performance when compared to the case without the HMD (CASE I). Correspondingly, it can be seen that the subjects rating on the feedback questionnaire were lowest for questions related to immersion and HMD comfort. Due to restrictions in device drivers, we used the side-by-side method to render the virtual room in stereo, which resulted in reduction in horizontal resolution by half. Moreover the VR2000 HMD had only 45 degrees field of view that resulted in limited immersion. We have already integrated a higher resolution Zsight (Sensics Inc.) and higher field of view Oculus Rift (Facebook Inc.) to the Gen2-VR$^©$ simulator which are expected to improve

comfort and realism. Even though in this study the individual effect of the distractor and interruptions cannot be separated from the subjective feedback score, it can be seen that subjects were not significantly affected by the music even though their least preferred category was played. It may be due to the task simplicity and subjects already achieving automaticity in the task since they were all FLS certified. A more complex procedural task may show definitively the effect of music. The interruptions to the procedure were rated high by the subjects on their ability to hinder their performance. This may be due to the frustration of the subjects on their inability to do anything during that period. Heart rate and skin conductance measurements may provide further insights, which are planned as part of our future research.

Acknowledgments

References

1. Fraser SA, Klassen DR, Feldman LS, Ghitulescu GA, Stanbridge D, Fried GM. Evaluating laparoscopic skills: setting the pass/fail score for the MISTELS system. Surg Endosc. 2003 Jun; 17(6):964–7. [PubMed: 12658417]

2. Arikatla VS, Sankaranarayanan G, Ahn W, Chellali A, De S, Cao CGL, et al. Face and construct validation of a virtual peg transfer simulator. Surg Endosc. 2012 Dec 14.

3. Duffy AJ, Hogle NJ, McCarthy H, Lew JI, Egan A, Christos P, et al. Construct validity for the LAPSIM laparoscopic surgical simulator. Surg Endosc. 2004 Dec; 19(3):401–5. [PubMed: 15624062]

4. Shiralkar, U. Smart Surgeons Sharp Decisions: Cognitive Skills to Avoid Errors & Achieve Results. 1. Harley, UK: Tfm Pub Ltd; 2011. p. 136

5. Healey AN, Sevdalis N, Vincent CA. Measuring intra-operative interference from distraction and interruption observed in the operating theatre. Ergonomics. 2006 May 15; 49(5–6):589–604. [PubMed: 16717011]

6. Zheng B, Martinec DV, Cassera MA, Swanström LL. A quantitative study of disruption in the operating room during laparoscopic antireflux surgery. Surg Endosc. 2008 Oct; 22(10):2171–7. [PubMed: 18622567]

7. Wiegmann DA, ElBardissi AW, Dearani JA, Daly RC, Sundt TM 3rd. Disruptions in surgical flow and their relationship to surgical errors: an exploratory investigation. Surgery. 2007 Nov; 142(5): 658–65. [PubMed: 17981185]

8. Feuerbacher, RL. Effects of realistic distractions and interruptions on simulated surgical task performance [Internet]. 2010. [cited 2012 May 30]. Available from: http://ir.library.oregonstate.edu/xmlui/handle/1957/16171

9. Lave, J.; Wenger, E. Situated Learning: Legitimate Peripheral Participation. 1. Cambridge England; New York: Cambridge University Press; 1991. p. 138

10. Kim, M-H.; Park, J.; Rhee, S-M.; Park, S-W. An Interactive Virtual Operating Room. In: Koyamada, K.; Tamura, S.; Ono, O., editors. Systems Modeling and Simulation [Internet]. Springer; Japan: 2007. p. 342-6.[cited 2014 Jul 15]Available from: http://link.springer.com.libproxy.rpi.edu/chapter/10.1007/978-4-431-49022-7_69

11. Tsuda S, Scott D, Doyle J, Jones DB. Surgical skills training and simulation. Curr Probl Surg. 2009 Apr; 46(4):271–370. [PubMed: 19249439]

12. Arikatla VS, Ahn W, Sankaranarayanan G, De S. Towards virtual FLS: development of a peg transfer simulator. Int J Med Robot Comput Assist Surg MRCAS. 2013 Sep 13.

13. Arikatla VS, Sankaranarayanan G, Ahn W, Chellali A, Cao CGL, De S. Development and validation of VBLaST-PTc : a virtual peg transfer simulator. Stud Health Technol Inform. 2013; 184:24–30. [PubMed: 23400124]

14. Chellali A, Zhang L, Sankaranarayanan G, Arikatla VS, Ahn W, Derevianko A, et al. Validation of the VBLaST peg transfer task: a first step toward an alternate training standard. Surg Endosc. 2014 Apr 26.

15. Hodge B, Thompson JF. Noise pollution in the operating theatre. Lancet. 1990 Apr 14; 335(8694): 891–4. [PubMed: 1969991]

16. Hasfeldt D, Laerkner E, Birkelund R. Noise in the operating room--what do we know? A review of the literature. J Perianesthesia Nurs Off J Am Soc PeriAnesthesia Nurses Am Soc PeriAnesthesia Nurses. 2010 Dec; 25(6):380–6.

17. Pluyter JR, Buzink SN, Rutkowski A-F, Jakimowicz JJ. Do absorption and realistic distraction influence performance of component task surgical procedure? Surg Endosc. 2010 Apr; 24(4):902–7. [PubMed: 19789922]

18. Persoon MC, Broos HJHP, Witjes JA, Hendrikx AJM, Scherpbier AJJM. The effect of distractions in the operating room during endourological procedures. Surg Endosc. 2011 Feb; 25(2):437–43. [PubMed: 20734086]

19. Siu K-C, Suh IH, Mukherjee M, Oleynikov D, Stergiou N. The impact of environmental noise on robot-assisted laparoscopic surgical performance. Surgery. 2010 Jan; 147(1):107–13. [PubMed: 19879611]

20. Conrad C, Konuk Y, Werner P, Cao CG, Warshaw A, Rattner D, et al. The effect of defined auditory conditions versus mental loading on the laparoscopic motor skill performance of experts. Surg Endosc. 2010 Jun; 24(6):1347–52. [PubMed: 20033720]

21. Miskovic D, Rosenthal R, Zingg U, Oertli D, Metzger U, Jancke L. Randomized controlled trial investigating the effect of music on the virtual reality laparoscopic learning performance of novice surgeons. Surg Endosc. 2008 Nov; 22(11):2416–20. [PubMed: 18622551]

22. Moorthy K, Munz Y, Undre S, Darzi A. Objective evaluation of the effect of noise on the performance of a complex laparoscopic task. Surgery. 2004 Jul; 136(1):25–30. discussion 31. [PubMed: 15232535]

23. Rauscher FH, Shaw GL, Ky CN. Music and spatial task performance. Nature. 1993 Oct 14; 365(6447):611–611. [PubMed: 8413624]

24. Siu K-C, Suh IH, Mukherjee M, Oleynikov D, Stergiou N. The effect of music on robot-assisted laparoscopic surgical performance. Surg Innov. 2010 Dec; 17(4):306–11. [PubMed: 20817638]

25. Goodell KH, Cao CGL, Schwaitzberg SD. Effects of Cognitive Distraction on Performance of Laparoscopic Surgical Tasks. J Laparoendosc Adv Surg Tech. 2006 Apr; 16(2):94–8.

26. Hsu KE, Man F-Y, Gizicki RA, Feldman LS, Fried GM. Experienced surgeons can do more than one thing at a time: effect of distraction on performance of a simple laparoscopic and cognitive task by experienced and novice surgeons. Surg Endosc. 2008 Jan; 22(1):196–201. [PubMed: 17705087]

27. Pashler H, Johnston JC, Ruthruff E. Attention and Performance. Annu Rev Psychol. 2001; 52(1): 629–51. [PubMed: 11148320]

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Figure 1.
The Gen2-VR© simulator showing the physical interface (left) and the immersive virtual environment (right) with the user represented as an avatar.

Figure 2.
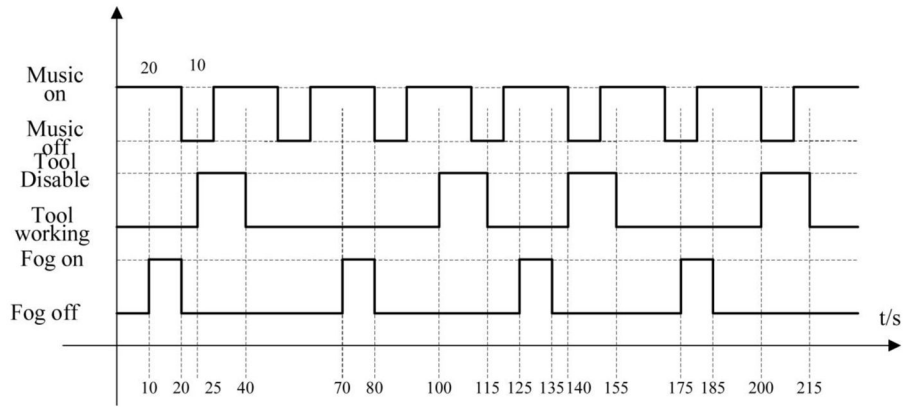Camera lens fogging view in the VR^2 simulator

Figure 3.
Sequence and duration of distraction and interruptions for a subject.
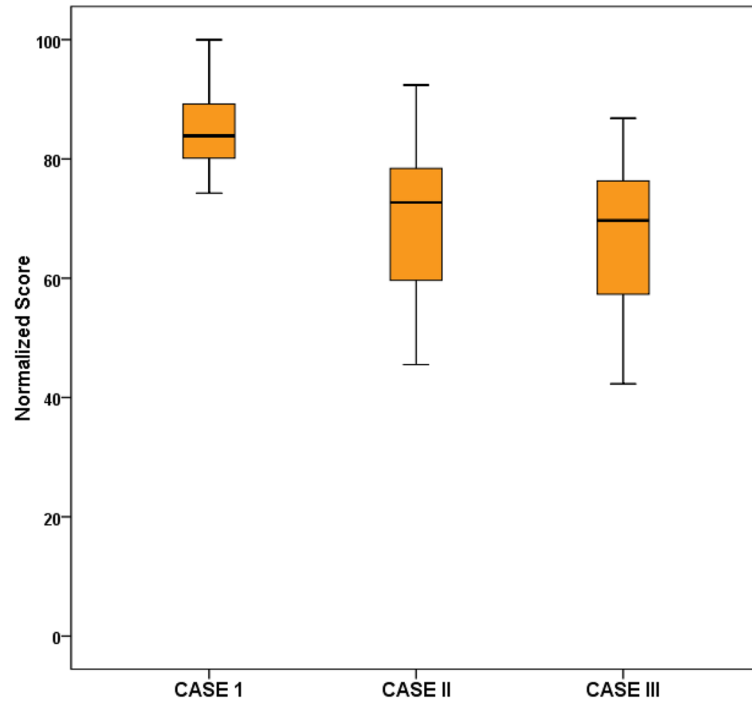
Figure 4.
Box plot of the normalized total scores for the three different cases

Table 1

Music genres used for auditory distraction

Number	Type
1	Classical
2	Metallica
3	Pop
4	Country
5	Jazz

Table 2

Face validity Likert scale data

Questionnaire	Ratings	
	Mean	**Std**
Hardware and Software of the VBLaST-PT© System		
1. Realism of task	3.812	1.223
2. Realism of instrument handling	3.25	1
3. Realism of the overall simulation	3.187	1.167
4. Realism of the force feedback	3.187	1.223
5. Degree of usefulness of the force feedback	3.437	1.152
6. Usefulness of the simulation in learning hand-eye coordination	3.562	0.892
7. Usefulness of the simulation in learning ambidexterity skills	3.625	1.147
8. Usefulness of the simulation in learning the fundamental laparoscopic technical skills	3.437	1.03
9. Assessment of how trustworthy the simulator is to quantify measures of performance	3.125	1.204
Gen2-VR© simulator immersive experience		
10. Comfort level in using the Head Mounted Display (HMD)	2.125	1.087
11. Realism of display in the HMD	2.625	0.957
12. In-room immersive experience	2.625	1.454
13. Did music hinder performance	2.0	0.894
14. Did fogging hinder performance	4.187	1.223
15. Did tool malfunction hinder performance	4.562	0.813
16. Difference in task with and without HMD	4.375	0.885

Table 3

Post hoc comparison of the three cases

Parameter	Post hoc comparisons		Wilcoxon signed rank test (Z, p)
Normalized score	CASE I	CASE II	(−4.084, < 0.0001)
	CASE I	CASE III	(−4.373, < 0.0001)
	CASE II	CASE III	(−2.619, 0.009)