Enhancing Immersion in Virtual Reality Based Advanced Life Support Training: A Randomized Control Trial

Dilek Kitapcioglu, Mehmet Emin Aksoy, Arun Ekin Ozkan, Tuba Usseli, Dilan Cabuk Colak, Tugrul Torun

Submitted to: JMIR Serious Games
on: November 01, 2024

Table of Contents

Enhancing Immersion in Virtual Reality Based Advanced Life Support Training: A Randomized Control Trial

Dilek Kitapcioglu[1] MD, PhD, Prof Dr Med; Mehmet Emin Aksoy[1] MD, PhD, Prof Dr Med, MSc; Arun Ekin Ozkan[2]; Tuba Usseli[3]; Dilan Cabuk Colak[4] PhD; Tugrul Torun[4] MS

[1]Ac?badem University CASE (Center of Advanced Simulation and Education) Istanbul TR
[2]Bo?aziçi University Istanbul TR
[3]Ac?badem University Department of Anesthesia, Vocational School of Health Services Istanbul TR
[4]Ac?badem University Department of Psychology, Faculty of Humanities and Social Sciences Istanbul TR

Corresponding Author:
Mehmet Emin Aksoy MD, PhD, Prof Dr Med, MSc
Ac?badem University
CASE (Center of Advanced Simulation and Education)
Kayisdagi cad. No:32 Atasehir
Istanbul
TR

Abstract

Background: Serious game-based training modules are pivotal for simulation-based healthcare training. With advancements in artificial intelligence (AI) and natural language processing (NLP), voice command interfaces offer an intuitive alternative to traditional VR controllers in VR applications.

Objective: The aim of this study is to compare AI-supported voice command interfaces and traditional VR controllers in terms of user performance, exam scores, presence and confidence in Advanced Cardiac Life Support)ACLS trainings.

Methods: Sixty-two volunteer students from Acibadem Mehmet Ali Aydinlar University Vocational School for Anesthesiology, aged 20-22, participated in the study. All the participants completed a pretest consisting of 10 multiple-choice questions about Advanced Cardiac Life Support (ACLS). Following the pretest, participants were randomly divided into two groups: the Voice command Group (VG) (n=31) and the Controller Group (CG) (n=31). The VG Group members completed the VR-based ACLS serious game in training mode twice, using an AI-supported voice command as the game interface. The CG Group members also completed the VR-based ACLS serious game in training mode twice, but they used VR controllers as the game interface. The participants completed a survey for estimating the level of presence and confidence during game play. Following the survey, participants completed the exam module of the VR-based serious gaming module. At the final stage of the study, participants completed a post-test, which had the same content of the pretest. VR-based exam scores of VG group and CG group were compared using an independent samples t-test and linear regression analysis was conducted to examine the effect of presence and confidence rating.

Results: Both groups showed an improvement in performance from pretest to posttest, with no significant difference in the magnitude of improvement between the two groups. When comparing presence ratings, there was no significant difference between the VG and CG groups (VG: $M = 5.18$, $SD = 0.83$; CG: $M = 5.42$, $SD = 0.75$). However, when comparing VR-based exam scores, the CG group significantly outperformed the VG group (CG: $M = 80.47$, $SD = 13.12$; VG: $M = 66.70$, $SD = 21.65$), despite both groups having similar time allocations for the exam (VG: $M = 18.59$ min, $SD = 5.28$; CG: $M = 17.33$ min, $SD = 4.83$). Confidence levels were similar between the groups (VG: $M = 0.23$, $SD = 0.18$; CG: $M = -0.09$, $SD = 0.18$), but VG group displayed a significant overconfidence bias (VG: $M = 0.09$, $SD = 0.24$; CG: $M = -0.09$, $SD = 0.18$). Regression analysis revealed a marginally significant positive effect of presence on confidence scores ($B = 0.23$, $t = 1.858$, $p = .07$), suggesting that increased immersion in VR environments may enhance learners' confidence.

Conclusions: AI voice recognition advancements could enhance presence and learning outcomes in VR-based training. Further research should explore ways to optimize AI's role in education through VR. Clinical Trial: ClinicalTrials.gov NCT06458452

(JMIR Preprints 01/11/2024:68272)
DOI: https://doi.org/10.2196/preprints.68272

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?

✓ **Please make my preprint PDF available to anyone at any time (recommended).**

Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.

Only make the preprint title and abstract visible.

No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✓ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**

Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain v

Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <a href="http

Original Manuscript

Enhancing Immersion in Virtual Reality Based Advanced Life Support Training: A Randomized Control Trial

Dilek Kitapcioglu[1,3], Mehmet Emin Aksoy[2,3], Arun Ekin Ozkan[4], Tuba Usseli[5], Dilan Cabuk Colak[6,7], Tugrul Torun[6,7]*

[1]Department Medical Education, Medical Faculty, Acibadem Mehmet Ali Aydinlar University, Istanbul, Turkey

[2]Department of Biomedical Device Technology, Acibadem Mehmet Ali Aydinlar University, Istanbul, Turkey

[3]CASE (Center of Advanced Simulation and Education), Acibadem Mehmet Ali Aydinlar University, Istanbul, Turkey

[4]Institute of Biomedical Engineering, Bogazici University, Istanbul Turkey

[5]Department of Anesthesia, Vocational School of Health Services, Acibadem Mehmet Ali Aydinlar University, Istanbul, Turkey

[6]Department of Psychology, Faculty of Humanities and Social Sciences, Acibadem Mehmet Ali Aydinlar University, Istanbul, Turkey

[7]Psychology Laboratory for Cognitive and Behavioral Research, Department of Psychology, Faculty of Humanities and Social Sciences, Acibadem Mehmet Ali Aydinlar University, Istanbul, Turkey.

Corresponding Author:
Mehmet Emin Aksoy
emin.aksoy@acibadem.edu.tr
Address:
Acibadem Mehmet Ali Aydinlar University,
Department of Biomedical Device Technology,
Kayisdagi cad. No:32, 34752 Istanbul
Turkey
ORCID Information:
Dilek Kitapcioglu : 0000-0002-0537-0479
Mehmet Emin Aksoy : 0000-0003-3227-2660
Arun Ekin Ozkan : 0000-0002-8151-0672
Tuba Usseli : 0000-0001-5619-0235
Dilan Cabuk Colak : 0000-0002-6222-3985
Tugrul Torun : 0000-0002-4656-1079

Abstract

Background: Serious game-based training modules are pivotal for simulation-based healthcare training. With advancements in artificial intelligence (AI) and natural language processing (NLP), voice command interfaces offer an intuitive alternative to traditional VR controllers in VR applications. The aim of this study is to compare AI-supported voice command interfaces and traditional VR controllers in terms of user performance, exam scores, presence and confidence in Advanced Cardiac Life Support)ACLS trainings.

Materials and Methods: Sixty-two volunteer students from Acibadem Mehmet Ali Aydinlar University Vocational School for Anesthesiology, aged 20-22, participated in the study. All the participants completed a pretest consisting of 10 multiple-choice questions about Advanced Cardiac Life Support (ACLS). Following the pretest, participants were randomly divided into two groups: the Voice command Group (VG) (n=31) and the Controller Group (CG) (n=31). The VG Group members completed the VR-based ACLS serious game in training mode twice, using an AI-supported voice command as the game interface. The CG Group members also completed the VR-based ACLS serious game in training mode twice, but they used VR controllers as the game interface. The participants completed a survey for estimating the level of presence and confidence during game play. Following the survey, participants completed the exam module of the VR-based serious gaming module. At the final stage of the study, participants completed a post-test, which had the same content of the pretest. VR-based exam scores of VG group and CG group were compared using an independent samples t-test and linear regression analysis was conducted to examine the effect of presence and confidence rating.

Results: Both groups showed an improvement in performance from pretest to posttest, with no significant difference in the magnitude of improvement between the two groups. When comparing presence ratings, there was no significant difference between the VG and CG groups (VG: M = 5.18, SD = 0.83; CG: M = 5.42, SD = 0.75). However, when comparing VR-based exam scores, the CG group significantly outperformed the VG group (CG: M = 80.47, SD = 13.12; VG: M = 66.70, SD = 21.65), despite both groups having similar time allocations for the exam (VG: M = 18.59 min, SD = 5.28; CG: M = 17.33 min, SD = 4.83). Confidence levels were similar between the groups (VG: M = 0.23, SD = 0.18; CG: M = -0.09, SD = 0.18), but VG group displayed a significant overconfidence bias (VG: M = 0.09, SD = 0.24; CG: M = -0.09, SD = 0.18). Regression analysis revealed a marginally significant positive effect of presence on confidence scores (B = 0.23, t = 1.858, p = .07), suggesting that increased immersion in VR environments may enhance learners' confidence.

Conclusion**:** AI voice recognition advancements could enhance presence and learning outcomes in VR-based training. Further research should explore ways to optimize AI's role in education through VR.

Keywords: Artificial Intelligence; Voice Recognition; Serious Gaming; Immersion

Research highlights

The VR-based serious gaming module proved to be effective for Advanced Cardiac Life Support (ACLS) training. However, the AI-supported voice control mode needs improvement due to language barriers, which resulted in lower VR exam scores for the Voice Control (VG) group.

The level of presence was similar in both groups, which was unexpected since voice control is perceived as more realistic and akin to a real clinical environment.

Increased immersion in VR environments may enhance learners' confidence.

Introduction

Serious game based training modules have become one of the training modalities for simulation based healthcare training and have an important role for the training of healthcare professionals[1-5]. Training using serious gaming modules can be conducted through various platforms such as PCs, tablet PCs, virtual reality (VR), or mixed reality systems like augmented reality or augmented virtuality. The portability and increased affordability of VR systems in recent years have contributed to a significant growth in VR-based learning. By providing immersive and interactive experiences of virtual reality (VR) technology has revolutionized various fields including gaming, education, and training of healthcare professionals[1, 6-8].

Enhanced student engagement makes VR and/or mixed reality modules particularly favored for healthcare trainings, such as Advanced Cardiac Life Support (ACLS) training, which aims to train healthcare professionals in the management of cardiopulmonary arrest. More than a million healthcare professionals participate in Advanced Cardiac Life Support (ACLS) or Advanced Life Support (ALS) courses globally each year[9]. Since these courses must be repeated at specific intervals, as mandated by local regulations or institutional requirements, software-based training modules and VR-based serious games will allow trainees to refresh their knowledge and skills anytime and anywhere. [10].

VR based training modules have used hand-held controllers for navigation and interaction with the virtual environment so far. In parallel to the advancements in artificial intelligence (AI) and natural language processing (NLP) potential for voice command interfaces, offer an alternative method of control in VR applications by providing more intuitive and accessible means of interaction for users[11-15]. Voice commands, powered by NLP algorithms, enable users to navigate and manipulate virtual environments using natural language and have the potential to reduce the learning curve associated with VR controller inputs. Voice interfaces have also the potential to offer accessibility benefits for users with physical disabilities, for whom conventional controllers may be difficult to use[14].

Despite the potential advantages, using voice command systems in VR gaming has also its limitations like speech recognition accuracy, latency, the system's ability to understand context-specific commands, problems caused by environmental noise, user's accent or speech patterns[16, 17].

There has been limited research comparing the user performance and user satisfaction of AI-supported voice command interfaces against traditional VR controllers [11, 13, 15]. This study aims to compare the two control methods using various metrics like task completion time, participants' exam scores, level of confidence and level of presence. As level of presence gives insight into how deeply the learner is immersed in the game, while confidence ratings reveal how the learner perceives their own performance and decision-making, level of presence and confidence ratings during gameplay were also compared [18-20]. The hypothesis of this study is that AI-supported voice commands have the potential of offering a superior user experience in VR gaming by providing a natural form of interaction.

Material and Method

Recruitment

This study included 67 volunteer students from Acibadem Mehmet Ali Aydinlar University Vocational School for Anesthesiology. The participants, aged between 20-22 years, were in their fourth semester (Spring Semester 2023-2024) of the Anesthesiology program. Five students declined to participate in the study. The remaining participants were randomly assigned to two groups: 31 students using VR Controllers group for game interaction (CG) and 31 students in the using AI-supported Voice Recognition group for game interaction (VG). The exclusion criteria for the study included prior ACLS training, a history of VR-induced motion sickness, and medical conditions such as vertigo attacks or using medications causing vertigo-like symptoms. Three participants from the VG group were excluded for not meeting the inclusion criteria. The randomization was based on the participants' university ID numbers, with odd numbers assigned to one group and even numbers to the other. All participants read and signed an informed consent document detailing the study procedures. The study was conducted at the CASE (Center of Advanced Simulation and Education)

at Acibadem Mehmet Ali Aydinlar University.
This study was approved by the Scientific Ethical Committee of Acibadem Mehmet Ali Aydinlar University (Approval Number: ATADEK 2024-6).

Study Flow

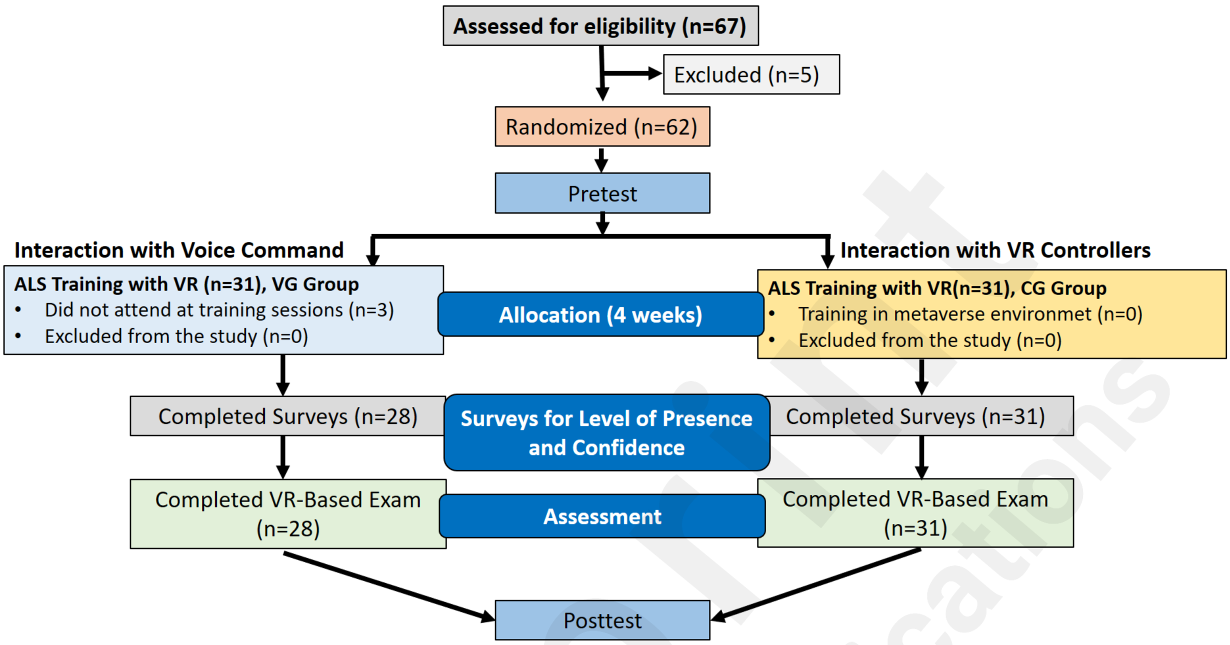


***Figure 1.** RCT flow diagram of the study*

All the participants completed a pretest consisting of 10 multiple-choice questions about ACLS, which served to assess the participants' initial knowledge level and the content is compatible with the latest version ACLS (Appendix 1)[21].
Following the pretest, participants were randomly divided into two groups: Voice command Group (VG) (n=31) and the VR controller Group (CG) (n=31) as seen in Figure 1. The VG Group members completed the VR-based ACLS serious game in training mode twice, using an AI-supported voice command as the game interface. The CG Group members also completed the VR-based ACLS serious game in training mode twice, but they used VR controllers as the game interface. Although having volunteered for the study, three participants from the VG group did not attend the study.
The participants from both groups completed the Presence Questionnaire (PQ) to estimate the level of presence during gameplay [22]. The PQ was first developed by Witmer and Singer to subjectively measure the level of presence in 3D virtual environments [23]. The scale was revised in a subsequent study conducted in 2005, which revealed its factor structure [24]. As a result of the principal components analysis, it was determined that the scale had a four-factor structure: Involvement, Sensory Fidelity, Adaptation/Immersion, and Interface Quality. The adaptation of the PQ to Turkish was conducted by Gokoglu and Cakiroglu [22]. The factor analysis of the Turkish version of the PQ found a five-factor structure, which was also confirmed by a confirmatory factor analysis.
In the involvement factor of the seven-point Likert-type scale, there are items for evaluating a situation obtained because of focusing the individual's attention and mental potential consistently or meaningfully on relevant stimuli, activities, or events. For the sensory fidelity factor, items related to perceiving the virtual reality scenario visually, auditorily, and haptically were included. The adaptation/immersion factor includes items related to being included in and interacting with a continuous flow of experiences and stimuli with a sense of being surrounded, while the interface quality factor includes items to evaluate the effect of visual and control interfaces in the virtual reality experience. Lastly, in the interaction factor, items related to interaction between individuals and the virtual environment were included. The adapted scale was found to be valid, reliable (α

= .84), and applicable with its 29 items under five factors [22].Participants' presence score was calculated as the mean of their responses to all items in the scale.

Following the survey, participants were asked to indicate how they believed they will perform in the upcoming exam from a scale of one to five, where one refers to very bad and 5 refers to very good, to assess confidence in future performance. After confidence assessments, participants completed the exam module of the VR-based serious gaming module, which evaluates technical and nontechnical of the participants. The VG group utilized a voice interface, while the CG group used VR controllers to complete the exam. At the final stage of the study, participants completed a posttest, which had the same content of the pretest.

Serious Gaming Module Used for The Study

The software used in this study is a virtual reality-based Advanced Life Support serious game that has been developed in accordance with the ACLS guidelines European Resuscitation Council (ERC) and American Heart Association (AHA) [21, 25]. The development team worked closely with clinicians to ensure compliance to these guidelines as well as crisis resource management (CRM) criteria [26-28].

This serious game works with a learning management system (LMS) and keeps user credentials as well as game results in a shared database with the help of a learning record store (LRS) [29-31]. The interactions of the players with the virtual world are tracked using a 3D visualization engine and experience API (xAPI) calls are created accordingly [30, 32]. These calls are sent to the LRS servers with the help of a Unity extension library that uses HTTP protocol [33]. This library also includes security measures and authentication methods included with HTTP protocols required for the safekeeping of user data.

The virtual reality-based Advanced Life Support serious game includes an artificial intelligence driven voice recognition feature that uses natural language processing algorithms of a service called Wit.ai [34]. Wit.ai is an open-source service that has existing Unity plug-ins which facilitates the implementation process. Through this service, the sentences of the user are sent to wit.ai servers, processed into game commands according to predefined keywords and returned to determine the action taken by the user.

There are two stages in this serious game: beginner training and VR-based exam. These stages can be played by either voice command or using controller of the VR headset. The simulated environment consists of a virtual hospital room and equipment to immerse the user as much as possible in this experience.

The beginner training stage is directed towards users that are willing to familiarize themselves with the ACLS algorithm and the VR environment. In this stage, the user is completely guided from the beginning to the end using visual and audio cues. Which objects to interact with or which voice commands to say are given to the users. Examples of guidance can be seen in Figure 2 and Figure 3 respectively. There are also no timing constraints in this mode which removes the stress element and allows an unexperienced user to focus on learning the algorithm [8].

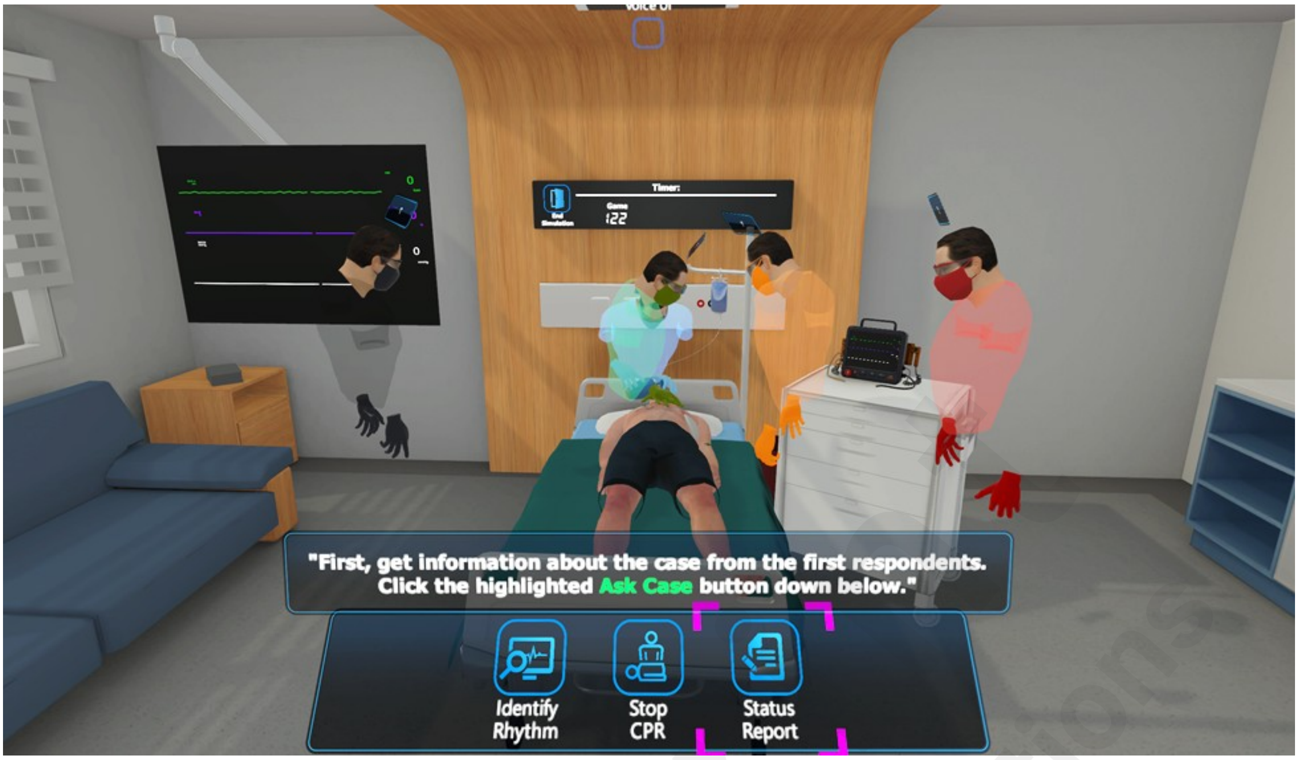


Figure 2. Screen capture of laser pointer-controlled mode of the ACLS serious game.

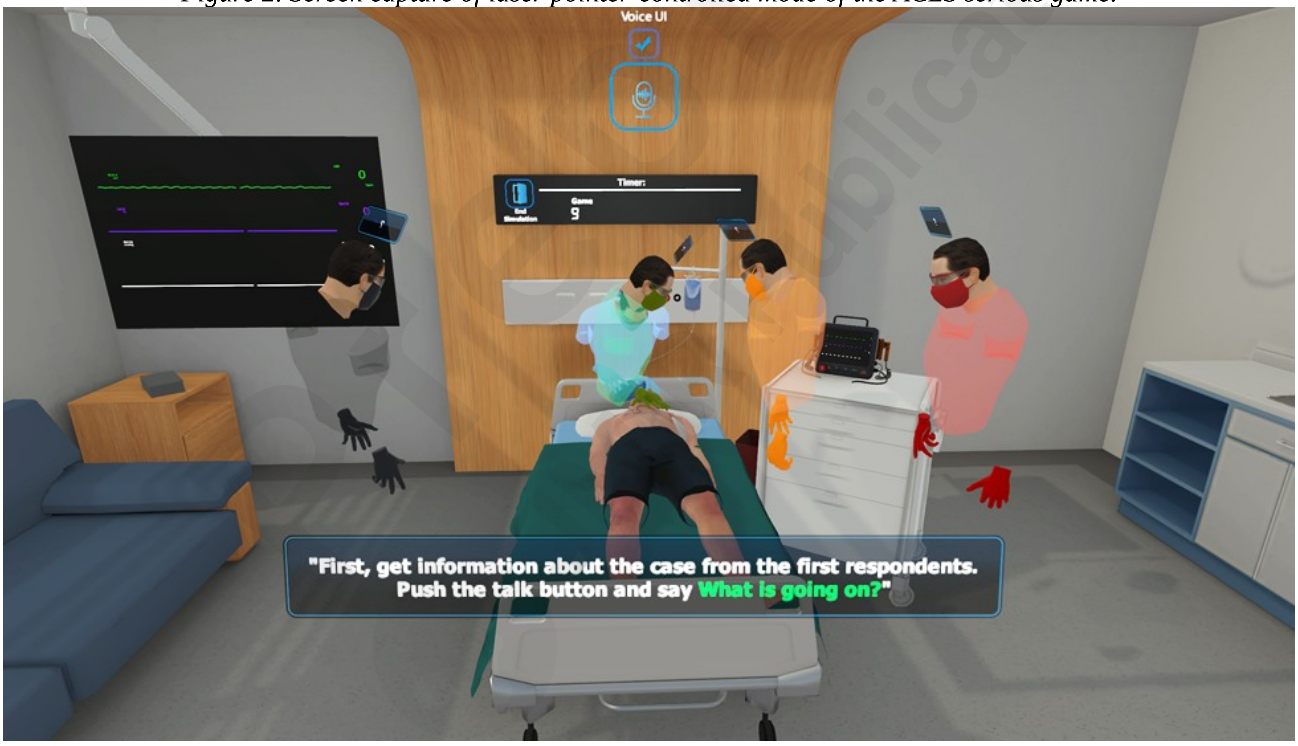


Figure 3. Screen capture of voice-control mode of the ACLS serious game.

The exam stage is designed for users that are confident in their abilities after spending some time in basic training stage. Each action of the user is scored according to its accuracy, order and timing. Each action taken by the user is graded based on its timing, order, and accuracy. The final score of VR-based exam is created from two categories: 70% (70/100) from ACLS assessments and 30% (30/100) from crisis resource management performances.

The scores of the exam mode were stored and could be accessed afterwards.

Data Analysis

Data from the comparison of pretest and post-test scores, VR-based exam scores, and time spent in training mode were evaluated to compare the two groups. Additionally, the relationship between VR-based exam scores and time spent in training was assessed. The correlation of these data with survey results on participants' presence and confidence ratings was also investigated.

Given the presence of multiple variables, a variety of statistical tests were employed for data analyses based on the nature of the addressed question. To compare pretest and posttest scores independent samples t-test was used. VR-based exam scores of VG group and CG group were compared using an independent samples t-test and linear regression analysis was conducted to examine the effect of presence and confidence rating. The details of statistical tests used for each analysis are reported in the Results section, along with the corresponding findings. Prior to analysis, the dataset was screened for outliers across all variables. No outliers were detected; thus, no data points were removed. The central tendency of the data was described using the mean.

Data was analyzed using JASP®, an open-source software for statistical analysis [35]. Visuals were created using both JASP and R Studio® [36].

Comparison of Pretest and Posttest Scores

An independent samples t-test showed that those in the VG group (M = 51.03, SD = 16.98) performed better than those in the CG group (M = 41.67, SD = 15.56) in the pretest, t (57) = -2.21, p = 0.03, d = -0.58. This difference was nearly significant in the posttest performance as well, t (57) = -1.97, p = .05, d = -0.51 (M_{VG} = 60.00, SD_{VG} = 15.35; M_{CG} = 51.67, SD_{CG} = 17.04). If the change in performance from the pretest to the posttest is different between groups was also examined using a 2 (*Group*: VG vs. CG) x 2(*Test*: pretest vs. posttest) mixed ANOVA where the first factor is between-subjects variable and the second is within-subjects variable. The results showed a main effect of group, such that VG group performed better than CG group in general, F (1, 57) = 6.43, p = .01, n^2_p = 0.10,. The main effect of test was also significant as the performance was higher in the posttest than the pretest for both groups, F (1, 57) = 15.69, p < 0.001, n^2_p = 0.22. This increase of performance from pretest to posttest, however, was not different for VG and CG groups, as the interaction between group and test was not significant, F (1, 57) = 0.05, p = 0.83, n^2_p = 8.18x10^{-4}. These findings indicate that while the voice command group had a higher performance in general, and both groups showed a performance increment from pretest to posttest, the increase in performance was not higher in one group than the other (see Figure 4).

CG Group VG Group

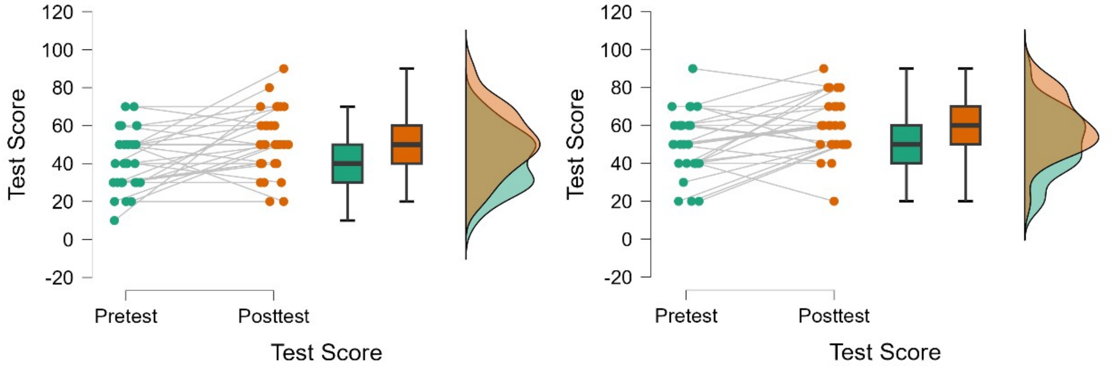


Figure 4. Pretest and posttest scores for CG and VG groups, respectively. Dots indicate each participants' test score. The boxes indicate interquartile ranges; the line in the middle indicates the median. The figures are created by using Jasp[35].

Comparison of Presence Ratings

In order to test if there was a difference between participants' ratings of presence across groups, an independent samples t-test with presence rating as the dependent variable and group as the independent variable was conducted. The results of the analysis showed that the presence ratings did not differ across groups, t (57) = 1.16, p = 0.25, d = 0.30, M_{VG} = 5.18, SD_{VG} = 0.83, M_{CG} = 5.42, SD_{CG} = 0.75.

Comparison of VR-based Exam Scores and Confidence Ratings

Crucial to the purpose of the study, VR-based exam scores of VG group and CG group were compared using an independent samples t-test with group as the independent variable and exam score as the dependent variable. The results showed that CG group's performance in the VR based exam (M = 80.47, SD = 13.12) was higher than VG group's performance (M = 66.70, SD = 21.65), t (57) = 2.97, p = 0.004, d = 0.77 (see Figure 5). Importantly, the performance advantage of the CG group can not be attributed to the time it took them to complete the test as an independent samples t-test showed that both groups spent similar amount of time in the exam, t (57) = -0.95, p = 0.35, d = -0.25, M_{VG} = 18.59, SD_{VG} = 5.28, M_{CG} = 17.33, SD_{CG} = 4.83 (Figure 6).

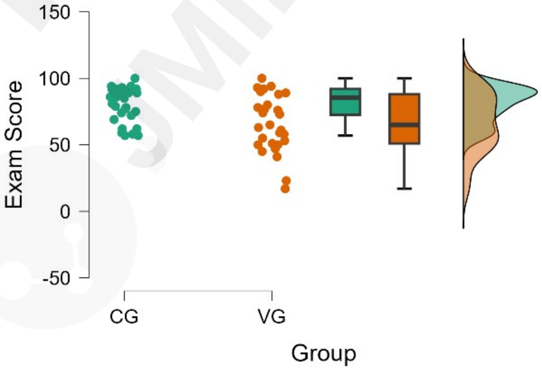


Figure 5. *Scores of VR-based exam mode for CG and VG group. Dots indicate each participants' exam score. The boxes indicate interquartile ranges; the line in the middle indicates the median. The figure is created by using Jasp [35].*

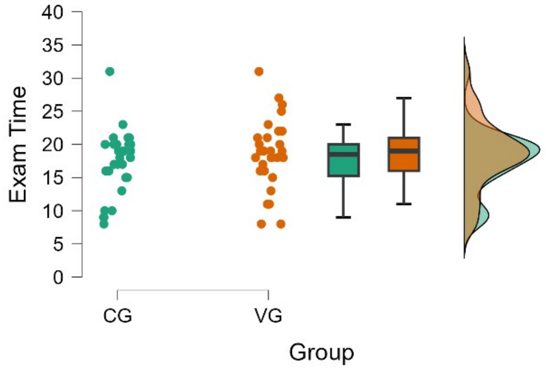


Figure 6. *Time spent on the advanced training stage for CG group and VG group in minutes. Dots indicate each participants' time. The boxes indicate interquartile ranges; the line in the middle indicates the median. The figure is created by using Jasp [35].*

In order to see if the observed difference in exam scores was also present in the confidence ratings as well, another independent samples t-test with group as the independent variable and confidence rating as the dependent variable was conducted. Results showed that there was no significant difference between groups' confidence ratings, t (57) = -0.99, p = 0.33, d = -0.26, M_{VG} = 3.79, SD_{VG}=0.77, M_{CG} = 3.60, SD_{CG} = 0.72. In general, there is a positive relationship between performance and confidence levels, such that higher performance is accompanied by higher confidence levels [18, 37]. Therefore, given the lower performance of VG group, one would expect to see lower confidence ratings as well. The absence of this expected difference points to a possible over-confidence bias in VG group, i.e., having higher confidence judgements compared to actual performance [38]. To see if this is the case, confidence bias for each participant was calculated by scaling their confidence ratings and advanced training performance over 1 and subtracting the scaled performance from the scaled confidence rating, a commonly used method for confidence bias quantification [39]. An independent samples t-test with confidence bias as the dependent variable and the group as the independent variable showed a significant difference with VG group's performance and confidence difference (M = 0.09, SD = 0.24) being bigger than the CG group's performance and confidence difference (M = -0.09, SD = 0.18), indicating an overconfidence bias in the VG group, t (57) = -3.21, p = 0.002, d = -0.83 (Figure 7).

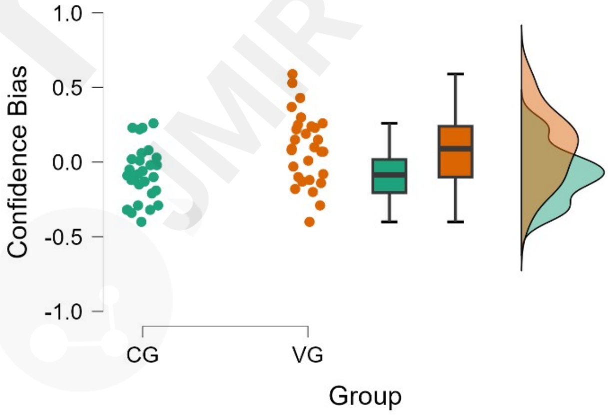


Figure 7. *Confidence bias for CG group and VG group. Dots indicate each participants' confidence bias. The boxes indicate interquartile ranges; the line in the middle indicates the median. The figure is created by using Jasp [35].*

A linear regression analysis was conducted to examine the effect of presence rating, group (VG), and their interaction on exam scores as shown in Table 1a. As it can be seen in Table 1b, the model was statistically significant, F (3, 55) = 4.10, p = 0.01, with an R^2 of 0.18. Examining variables in the model showed that being in the VG group had a positive but non-significant effect on exam scores (t = 1.215, p = .230), indicating that the previously found group effect on exam scores became insignificant when presence is included in the model. Indeed, the interaction between presence rating and group had a marginally negative effect on exam scores for the VG group (t = -

1.679, *p* = .099), suggesting that the effect of presence rating on exam scores was reduced by 9.99 units in the VG group compared to the CG group.

Table 1.a. Model summary for linear regression.

Model	R	R^2	Adjusted R^2	RMSE
H_0	0.00	0.00	0.00	18.99
H_1	0.43	0.18	0.14	17.63

Table 1.b. ANOVA results of the linear regression model.

Model		Sum of Squares	df	Mean Square	F	p
H_1	Regression	3822.79	3	1274.26	4.10	0.01
	Residual	17087.72	55	310.69		
	Total	20910.51	58			

Table 1.c. Results of linear regression analysis.

Model		Unstandardized	Standard Error	Standardized	t	p
H_0	(Intercept)	73.70	2.47		29.81	< .001
H_1	(Intercept)	62.00	24.03		2.58	0.01
	Presence Rating	3.41	4.39	0.14	0.78	0.44
	Group (VG)	38.79	31.93		1.22	0.23
	Presence Rating ✻ Group (VG)	-9.99	5.95		-1.68	0.099

Indeed, as it can be seen in Figure 8, the effect of presence on exam score presents a negative relationship for the VG group, while a positive one for the CG group. This finding requires further testing with a with a larger sample size, though, as the interaction effect was only marginally significant.

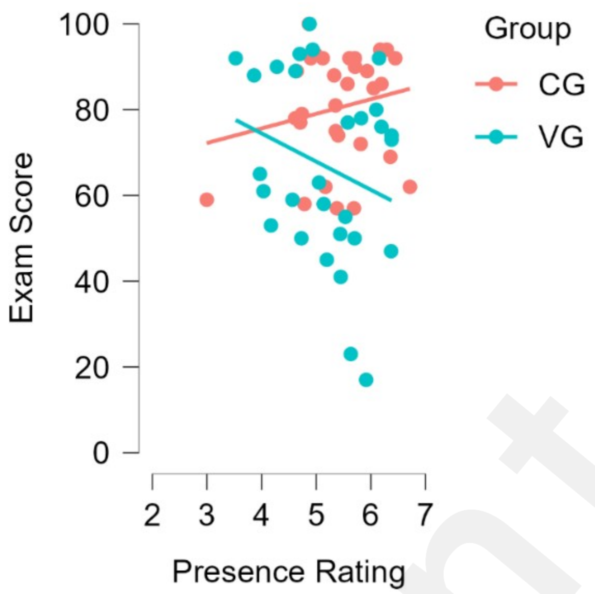


Figure 8. Relationship between presence rating and exam score for CG and VG groups. Dots indicate each participants' presence rating and exam score. The figure is created by using Jasp [35].

Confidence Prediction by Presence

A linear regression analysis was conducted to examine the effect of presence rating on confidence scores. As shown in Table 2.a and 2.b., the model was marginally significant, F (1, 57) = 3.45, p = .07, with an R^2 of .06. Specifically, presence had a marginal positive effect on exam scores with a B of 0.23 (t = 1.86, p = .07), indicating that participants became more confident in their future performance as their feelings of presence increased as seen in Figure 9. This finding also requires further testing with a larger sample size, though, as the model was only marginally significant.

Table 2.a. Model summary for linear regression.

Model	R	R^2	Adjusted R^2	RMSE
H_0	0.00	0.00	0.00	0.75
H_1	0.24	0.06	0.04	0.73

Table 2.b. ANOVA results of the linear regression model.

Model		Sum of Squares	df	Mean Square	F	p
H_1	Regression	1.86	1	1.86	3.45	0.068
	Residual	30.65	57	0.54		
	Total	32.51	58			

Model		Unstandardized	Standard Error	Standardized	t	p
H_0	(Intercept)	3.70	0.10		37.91	< .001

Model		Sum of Squares	df	Mean Square	F	p
H1	(Intercept)	2.50	0.65		3.82	< .001
	Presence Rating	0.23	0.12	0.24	1.86	0.068

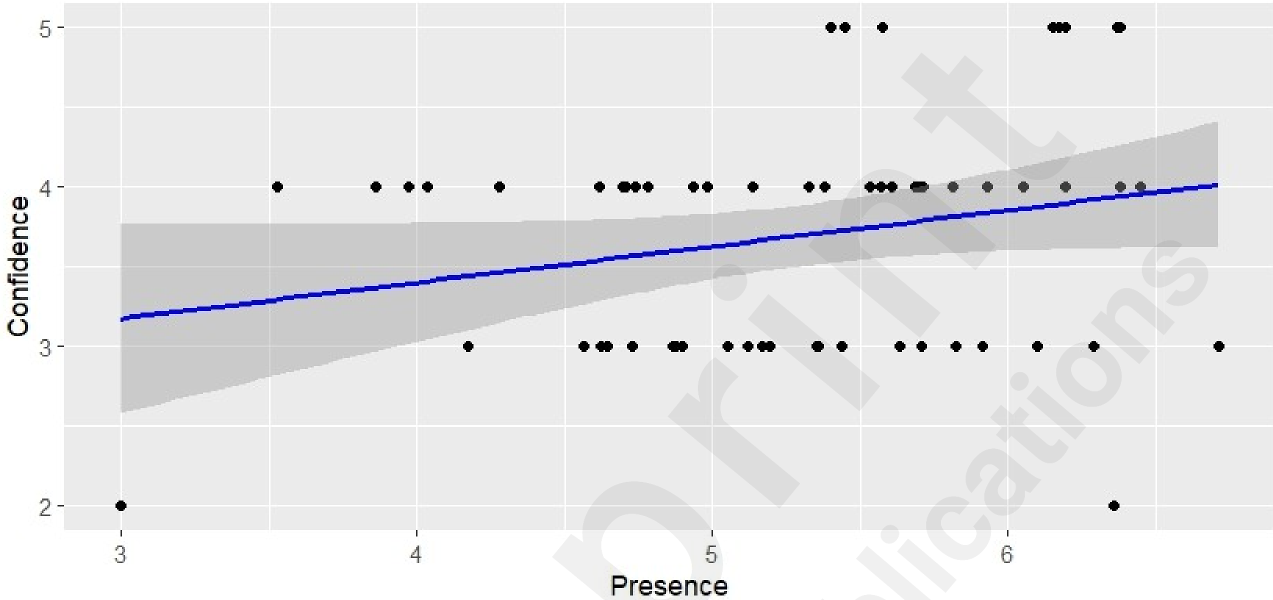


Figure 9. *Regression line for presence as predictor variable and confidence as the outcome variable. Dots indicate each participants' presence rating and confidence. The figure is created by using RStudio [40].*

Discussion

This study investigated the impact of AI-driven voice control on VR-based Advanced Cardiac Life Support (ACLS) training, focusing on performance outcomes, presence ratings, and confidence levels.

The findings of the study indicate that while the voice command group had a higher posttest scores and both groups showed a performance increment from pretest to posttest, the increase in posttest scores was not higher in one group than the other.

When the VR based exam scores of the two groups were compared, the CG group significantly outperformed the VG group in the exam mode (M = 80.47, SD = 13.12 vs. M = 66.70, SD = 21.65), despite similar time allocations for the exam (MVG = 18.59 min, SDVG = 5.28; MCG = 17.33 min, SDCG = 4.83). This finding underscores the potential problems of AI-guided voice recognition in terms of accuracy and performance in different languages [41, 42]. Since Turkish was selected as the language for AI-driven voice recognition in this study, there were minor difficulties due to the challenges of voice recognition, which caused participants in the VG group to lose time and achieve lower scores compared to the CG group.

When presence ratings were compared, our study revealed no significant difference between the VG and CG groups (MVG = 5.18, SDVG = 0.83; MCG = 5.42, SDCG = 0.75), diverging from expectations based on previous literature emphasizing AI's potential to enhance user engagement and immersion in virtual environments [18, 19]. Our study findings suggest that while AI-driven voice recognition may facilitate learning, its impact on presence may be nuanced and context-dependent, necessitating further research on this topic.

While they are related, confidence bias and task performance are dissociable [38],making it important to examine both to better understand the learning process. Although both groups reported comparable confidence levels (M_{VG} = 0.23, SD_{VG} = 0.18; M_{CG} = -0.09, SD_{CG} = 0.18), the VG group exhibited a significant overconfidence bias (M = 0.09, SD = 0.24) compared to the CG group (M = -0.09, SD = 0.18). This discrepancy highlights the importance of aligning perceived competence with actual performance outcomes in educational settings[43].

The regression analysis revealed a marginally significant positive effect of presence on confidence scores (B = 0.23, t = 1.858, p = .07), indicating that increased immersion in VR environments may enhance learners' confidence. This finding is expected, as confidence is influenced by cues such as fluency, which can be enhanced through higher levels of immersion [44]. However, the interaction effect between presence and exam scores was marginally negative for the VG group (t = -1.679, p = .099), suggesting a potential diminishing return of presence of the VG group's participants. Whether this negative effect is related to overconfidence bias observed in VG group remains an open question for future studies.

Conclusion

The findings of this study reveal that there is a potential connection between AI-driven voice recognition and educational outcomes in VR environments. In parallel to the advancements in AI voice recognition across various languages, the level of presence and learning outcomes in VR-based serious gaming modules have the potential to improve in the near future. Exploring alternative measures to enhance presence and examining the interaction effects of AI technologies on individual learners could provide deeper insights into optimizing AI's role in VR-based education. This could lead to improved learning outcomes and better preparation for real-world challenges in healthcare.

Clinical trial registration

ClinicalTrials.gov NCT06458452

Acknowledgments

The authors express their gratitude to the biomedical support team at the Center of Advanced Simulation and Education of Acibadem Mehmet Ali Aydinlar University for their invaluable support during this study.

Data Availability

The data sets generated during or analyzed during this study are available from the corresponding author on request.

Authors' Contributions

DK was responsible for study conceptualization, methodology, and reviewing and editing the manuscript.

MEA was responsible for study conceptualization, methodology, and resources and drafting, reviewing, correspondence.

AEO was responsible for the software and methodology.

TU was responsible for recruitment of the participants and data collection.

DCC was responsible for data analysis and survey preparation.

TT was responsible for running the surveys.

Disclosure

The authors did not use generative AI to write any portion of the manuscript.

Conflicts of Interest

None declared.

Funding

This research did not receive any specific grants from public, commercial, or not-for-profit funding agencies

Ethical Approval

This study was approved by the Scientific Ethical Committee of Acibadem Mehmet Ali Aydinlar University (Approval Number: ATADEK 2024-6).

Abbreviations

ACLS : Advanced Cardiac Life Support
AHA : American Heart Association
AI : Artificial Intelligence
ALS : Advanced Life Support
API : xAPI Experience
ERC : European Resuscitation Council
HTTP : HyperText Transfer Protocol
LMS : Learning Management System
LRS : Learning Record Store
NLP : Natural Language Processing
PQ : Presence Questionnaire
VR : Virtual Reality

References

1. Tudor Car, L., et al., *Outcomes, Measurement Instruments, and Their Validity Evidence in Randomized Controlled Trials on Virtual, Augmented, and Mixed Reality in Undergraduate Medical Education: Systematic Mapping Review.* JMIR Serious Games, 2022. **10**(2): p. e29594.
2. Reis, G., et al., *Mixed reality applications in urology: Requirements and future potential.* Annals of Medicine & Surgery, 2021. **66**.
3. Stepan, K., et al., *Immersive virtual reality as a teaching tool for neuroanatomy.* International Forum of Allergy & Rhinology, 2017. **7**(10): p. 1006-1013.
4. Aksoy, M.E., et al., *Assessing Correlation Between Virtual Reality Based Serious Gaming Performance and Cognitive Workload Changes via Functional Near Infrared Spectroscopy*.

2019. p. 375-383.

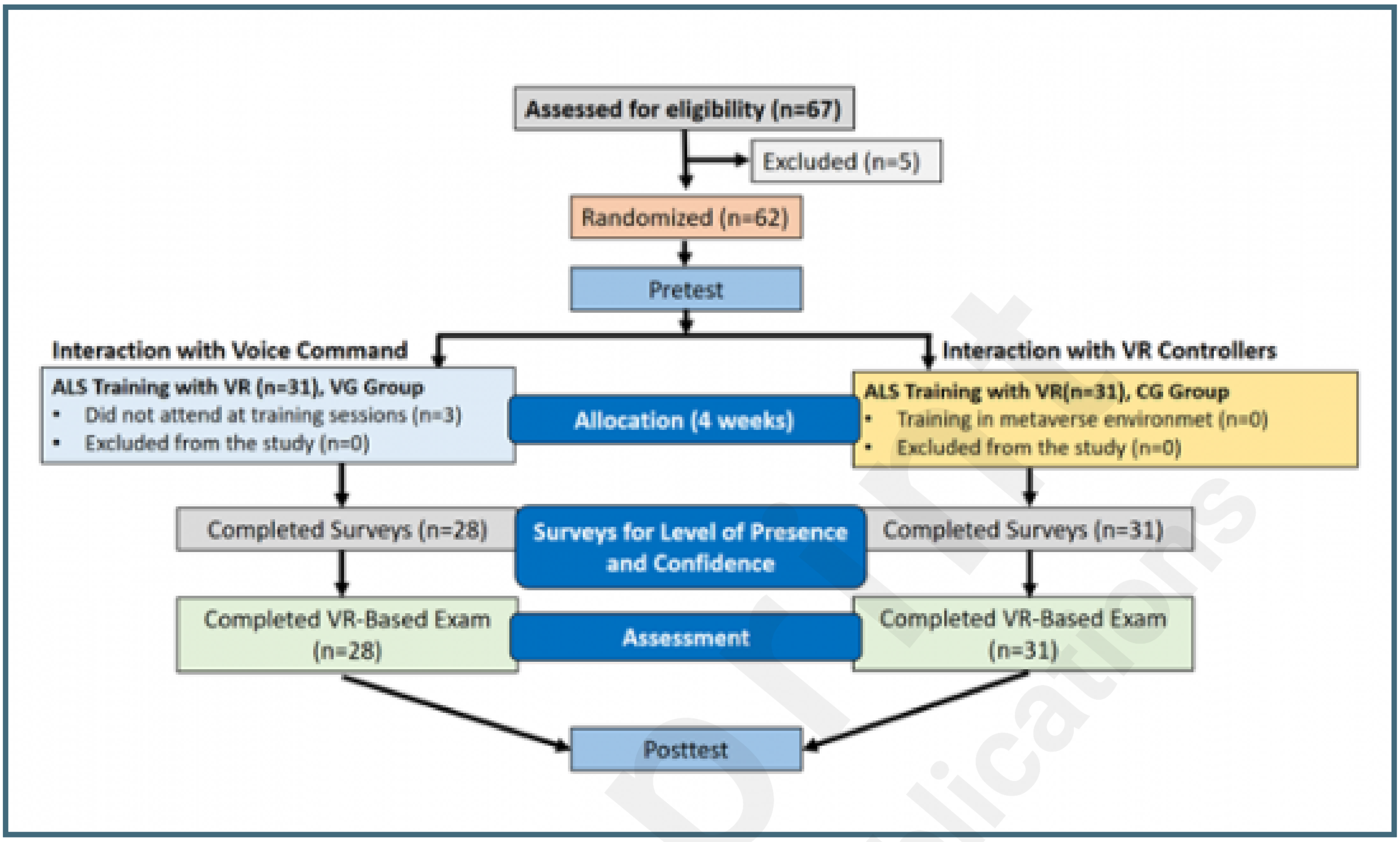
5. Aksoy, E., *Comparing the Effects on Learning Outcomes of Tablet-Based and Virtual Reality–Based Serious Gaming Modules for Basic Life Support Training: Randomized Trial.* JMIR Serious Games, 2019. **7**(2): p. e13442.
6. Liberatore, M.J. and W.P. Wagner, *Virtual, mixed, and augmented reality: a systematic review for immersive systems research.* Virtual Reality, 2021. **25**(3): p. 773-799.
7. Maas, M.J. and J.M. Hughes, *Virtual, augmented and mixed reality in K–12 education: a review of the literature.* Technology, Pedagogy and Education, 2020. **29**(2): p. 231-249.
8. Aksoy, M.E., et al., *Comparing the Outcomes of Virtual Reality–Based Serious Gaming and Lecture-Based Training for Advanced Life Support Training: Randomized Controlled Trial.* JMIR Serious Games, 2023. **11**: p. e46964.
9. Lockey, A., Y. Lin, and A. Cheng, *Impact of adult advanced cardiac life support course participation on patient outcomes—A systematic review and meta-analysis.* Resuscitation, 2018. **129**: p. 48-54.
10. Moore, N., et al., *Exploring User Needs in the Development of a Virtual Reality–Based Advanced Life Support Training Platform: Exploratory Usability Study.* JMIR Serious Games, 2020. **8**(3): p. e20797.
11. Ishigaki, S.A.K., et al., *Voice Commands with Virtual Assistant in Mixed Reality Telepresence.* 2024. **1983**: p. 148-158.
12. Cha, H.-S., W.-D. Chang, and C.-H. Im, *Deep-learning-based real-time silent speech recognition using facial electromyogram recorded around eyes for hands-free interfacing in a virtual reality environment.* Virtual Reality, 2022. **26**(3): p. 1047-1057.
13. Yang, J., et al., *Prototyping Virtual Reality Interactions in Medical Simulation Employing Speech Recognition.* 2020: p. 351-355.
14. Masina, F., et al., *Investigating the Accessibility of Voice Assistants With Impaired Users: Mixed Methods Study.* Journal of Medical Internet Research, 2020. **22**(9): p. e18431.
15. Song, Z., et al., *Application of Speech Recognition Technology to Virtual Reality System.* 2015. **12**.
16. Malmierca, M.S., et al., *Speech Recognition in Natural Background Noise.* PLoS ONE, 2013. **8**(11): p. e79279.
17. Li, D., et al., *Improving Speech Recognition Performance in Noisy Environments by Enhancing Lip Reading Accuracy.* Sensors, 2023. **23**(4): p. 2053.
18. McCurdy, L.Y., et al., *Anatomical Coupling between Distinct Metacognitive Systems for Memory and Visual Perception.* The Journal of Neuroscience, 2013. **33**(5): p. 1897-1906.
19. Osking, H. and J.A. Doucette, *Enhancing Emotional Effectiveness of Virtual-Reality Experiences with Voice Control Interfaces.* 2019. **1044**: p. 199-209.
20. de Bruin, A.B.H., J. Dunlosky, and R.B. Cavalcanti, *Monitoring and regulation of learning in medical education: the need for predictive cues.* Medical Education, 2017. **51**(6): p. 575-584.
21. Panchal, A.R., et al., *Part 3: Adult Basic and Advanced Life Support: 2020 American Heart Association Guidelines for Cardiopulmonary Resuscitation and Emergency Cardiovascular Care.* Circulation, 2020. **142**(16_suppl_2).
22. Gokoglu, S. and U. Cakiroglu, *Sanal GerÇeklİk Temellİ ÖĞrenme Ortamlarinda BulunuŞluk Hİssİnİn ÖlÇÜlmesİ: BulunuŞluk ÖlÇeĞİnİn TÜrkÇe'ye Uyarlanmasi.* Eğitim Teknolojisi Kuram ve Uygulama, 2019. **9**(1): p. 169-188.
23. Witmer, B.G. and M.J. Singer, *Measuring Presence in Virtual Environments: A Presence Questionnaire.* Presence: Teleoperators and Virtual Environments, 1998. **7**(3): p. 225-240.
24. Witmer, B.G., C.J. Jerome, and M.J. Singer, *The Factor Structure of the Presence*

Questionnaire. Presence: Teleoperators and Virtual Environments, 2005. **14**(3): p. 298-312.
25. Kane, A.D. and J.P. Nolan, *Changes to the European Resuscitation Council guidelines for adult resuscitation*. BJA Education, 2022. **22**(7): p. 265-272.
26. Krage, R., et al., *Relationship between non-technical skills and technical performance during cardiopulmonary resuscitation: does stress have an influence?* Emergency Medicine Journal, 2017. **34**(11): p. 728-733.
27. Peltonen, V., et al., *An observational study of technical and non-technical skills in advanced life support in the clinical setting*. Resuscitation, 2020. **153**: p. 162-168.
28. Saravana-Bawan, B.B., et al., *Evaluating Best Methods for Crisis Resource Management Education*. Simulation in Healthcare: The Journal of the Society for Simulation in Healthcare, 2019. **14**(6): p. 366-371.
29. Arancibia Muñoz, M.L. and C. Halal Orfali, *The Role of Learning Management Systems (Lms) in Educational Environments: Didactic and Technical Use*. 2018. **1**: p. 10473-10479.
30. García Clemente, F.J., et al., *Collecting Experience Data from Remotely Hosted Learning Applications*. 2018. **22**: p. 170-181.
31. Rustici Software. *Learning Record Store (LRS)*. 2023 [cited 2023 28.02.2023]; Available from: https://xapi.com/learning-record-store/.
32. Akbari, F., et al., *Comparison of the effects of virtual training by serious game and lecture on operating room novices' knowledge and performance about surgical instruments setup: a multi-center, two-arm study*. BMC Medical Education, 2022. **22**(1).
33. Technologies, U. *Unity Plugin*. 2023 [cited 2023 01.023.2023]; Available from: https://unity.com/.
34. AI, W. [cited 2024; Available from: https://wit.ai/.
35. *The Visual Modeling Module*. 2022; Available from: https://jasp-stats.org/2020/04/21/the-visual-modeling-module/.
36. *The R Project for Statistical Computing*. 2023 [cited 2023; Available from: https://www.r-project.org/.
37. Rouault, M., et al., *Human Metacognition Across Domains: Insights from Individual Differences and Neuroimaging*. Personality Neuroscience, 2018. **1**.
38. Fleming, S.M. and H.C. Lau, *How to measure metacognition*. Frontiers in Human Neuroscience, 2014. **8**.
39. Schraw, G., et al., *Does a general monitoring skill exist?* Journal of Educational Psychology, 1995. **87**(3): p. 433-444.
40. *Announcing RStudio Workbench*. 2024; Available from: https://posit.co/blog/announcing-rstudio-workbench/.
41. Fang, K. and J. Wang, *Interactive Design With Gesture and Voice Recognition in Virtual Teaching Environments*. IEEE Access, 2024. **12**: p. 4213-4224.
42. *A detailed survey of Turkish automatic speech recognition*. Turkish Journal of Electrical Engineering & Computer Sciences, 2020. **28**(6).
43. Schraw, G. and D. Moshman, *Metacognitive theories*. Educational Psychology Review, 1995. **7**(4): p. 351-371.
44. Finn, B. and S.K. Tauber, *When Confidence Is Not a Signal of Knowing: How Students' Experiences and Beliefs About Processing Fluency Can Lead to Miscalibrated Confidence*. Educational Psychology Review, 2015. **27**(4): p. 567-586.

Supplementary Files

Figures

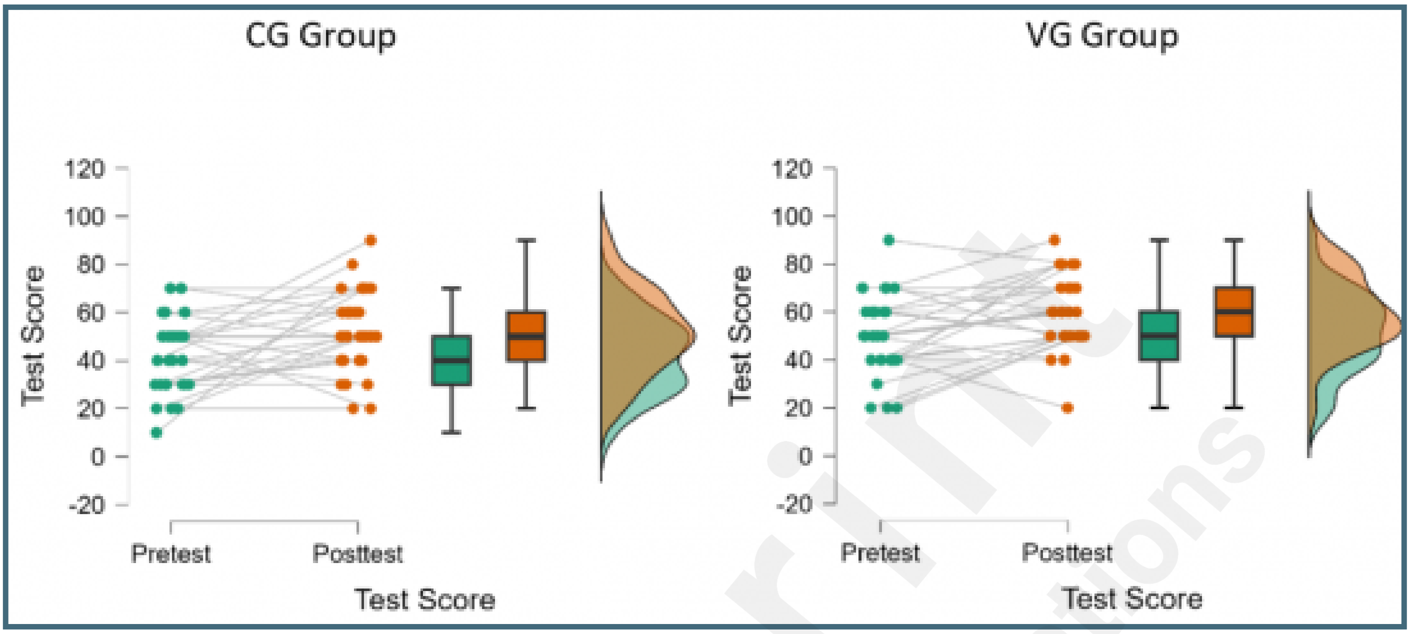
RCT flow diagram of the study.



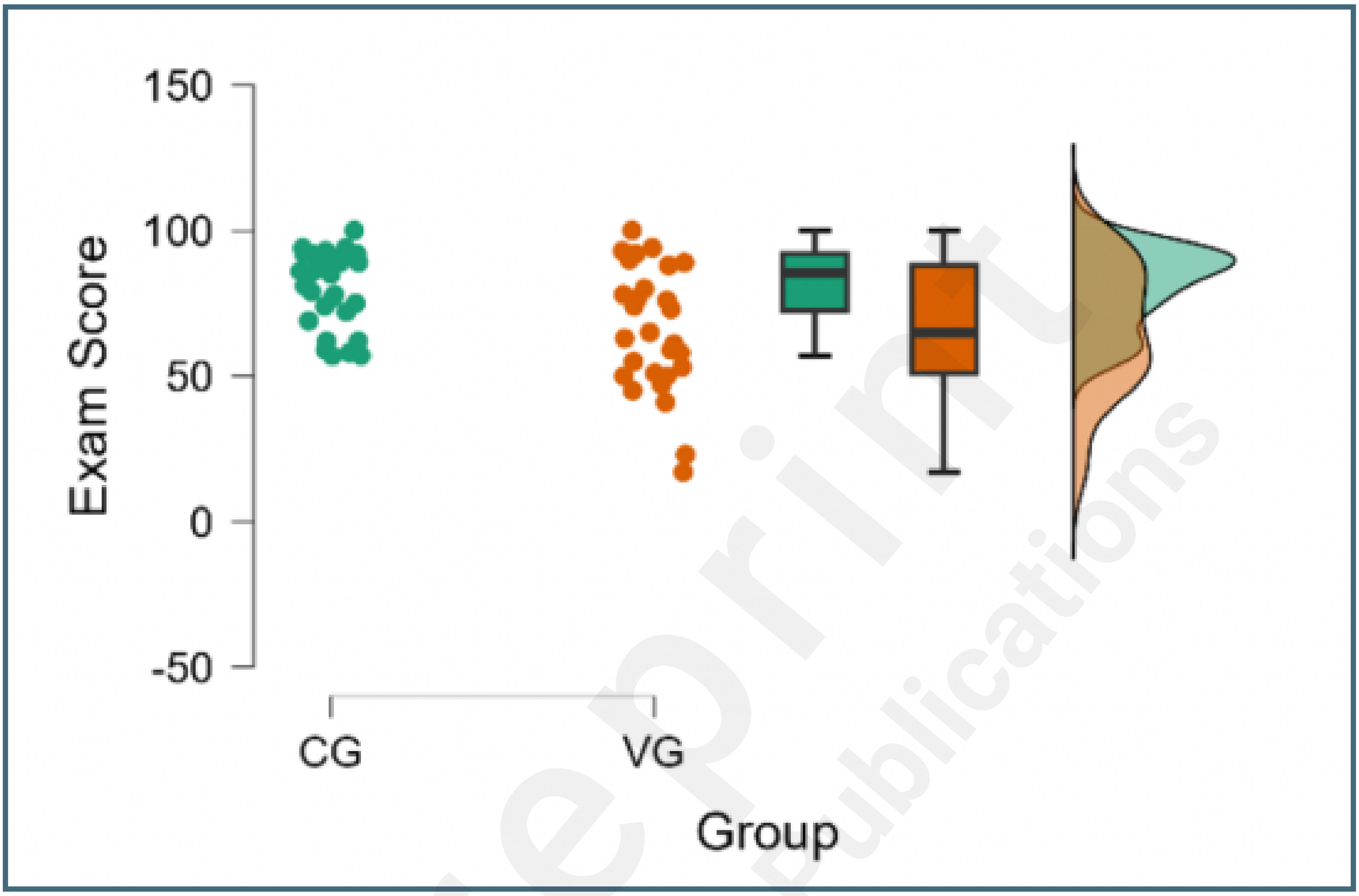
Screen capture of laser pointer-controlled mode of the ACLS serious game.

Screen capture of voice-control mode of the ACLS serious game.

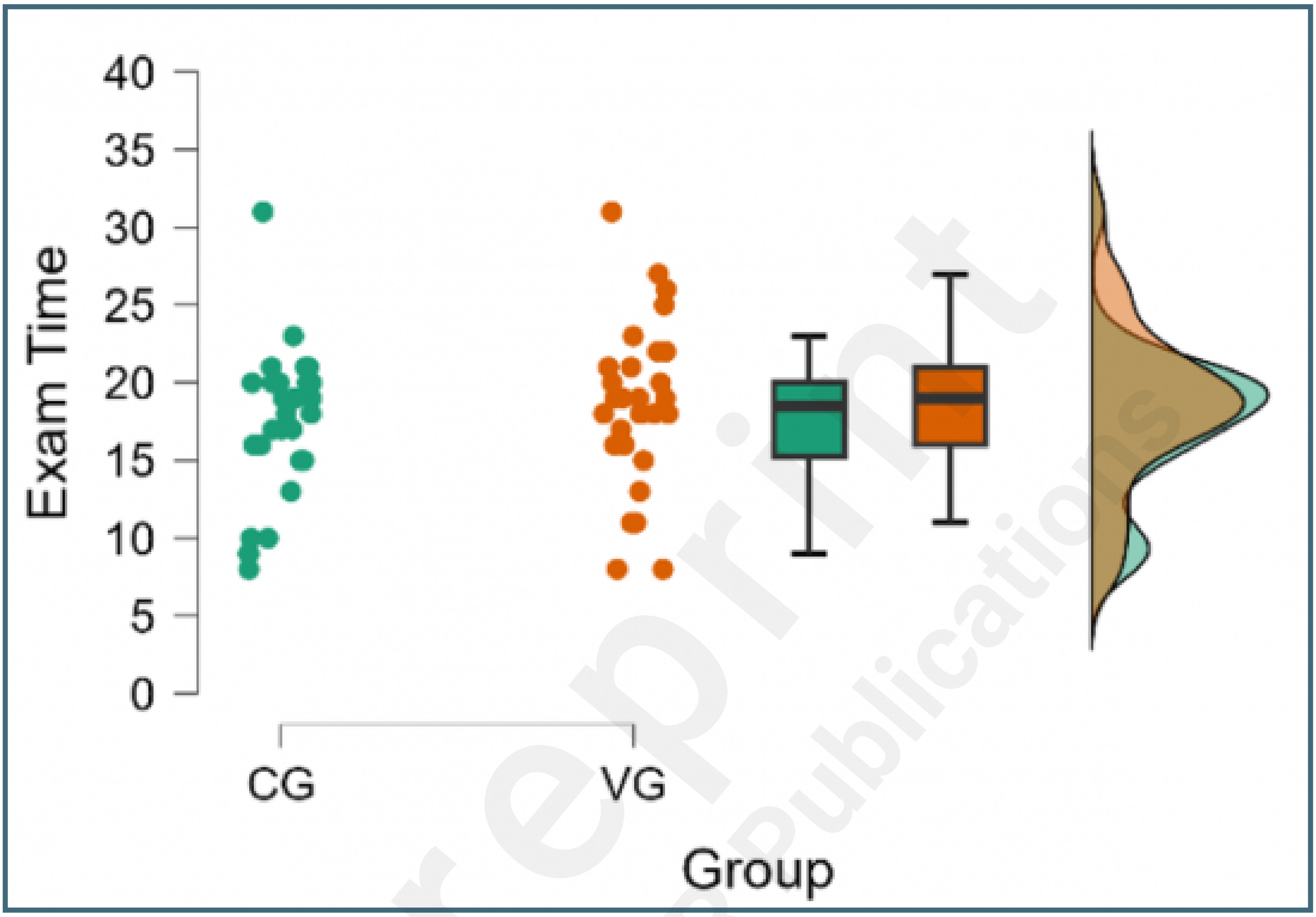
Pretest and posttest scores for CG and VG groups, respectively. Dots indicate each participants' test score. The boxes indicate interquartile ranges; the line in the middle indicates the median. The figures are created by using Jasp[35].



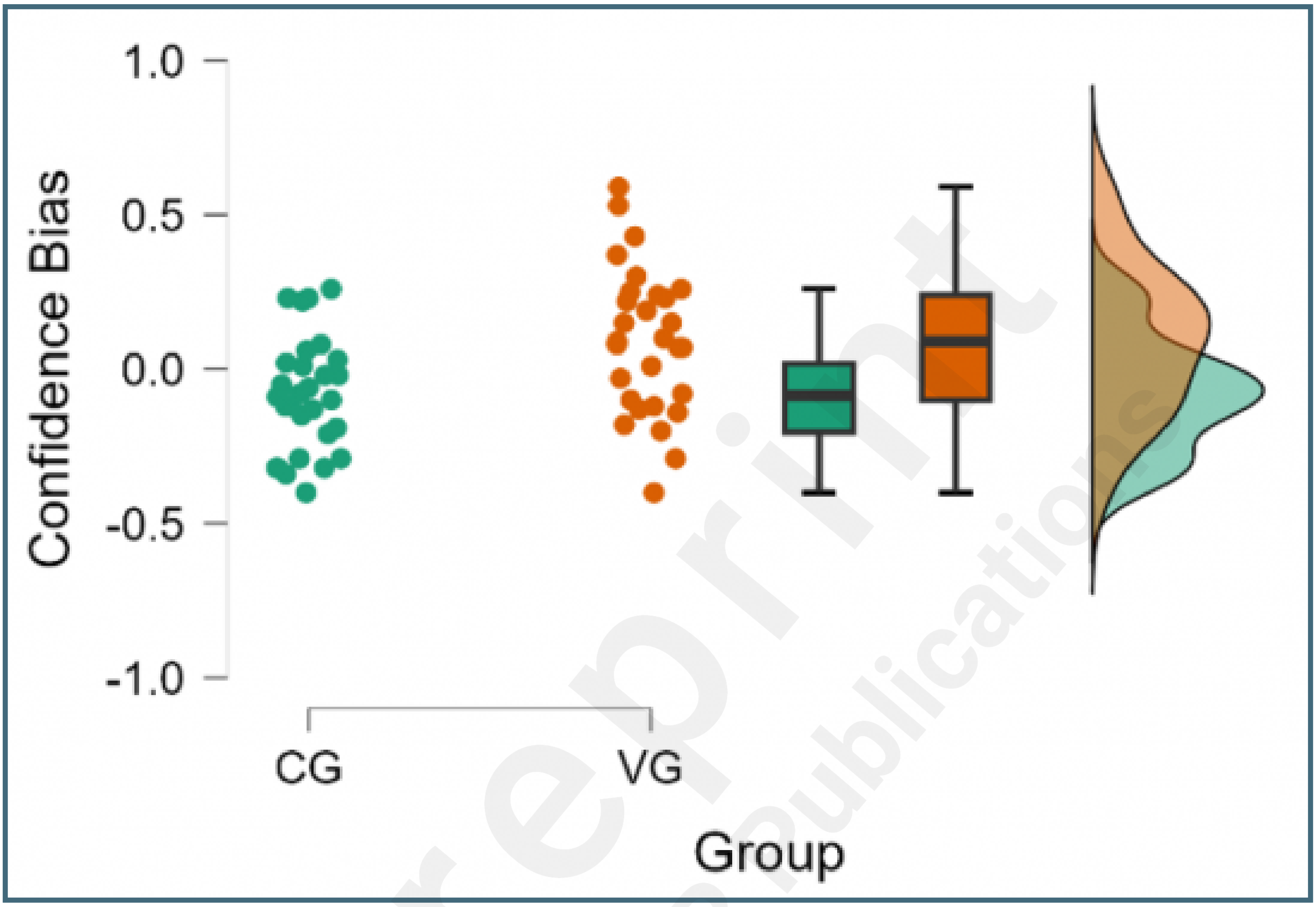
Scores of VR-based exam mode for CG and VG group. Dots indicate each participants' exam score. The boxes indicate interquartile ranges; the line in the middle indicates the median. The figure is created by using Jasp [35].



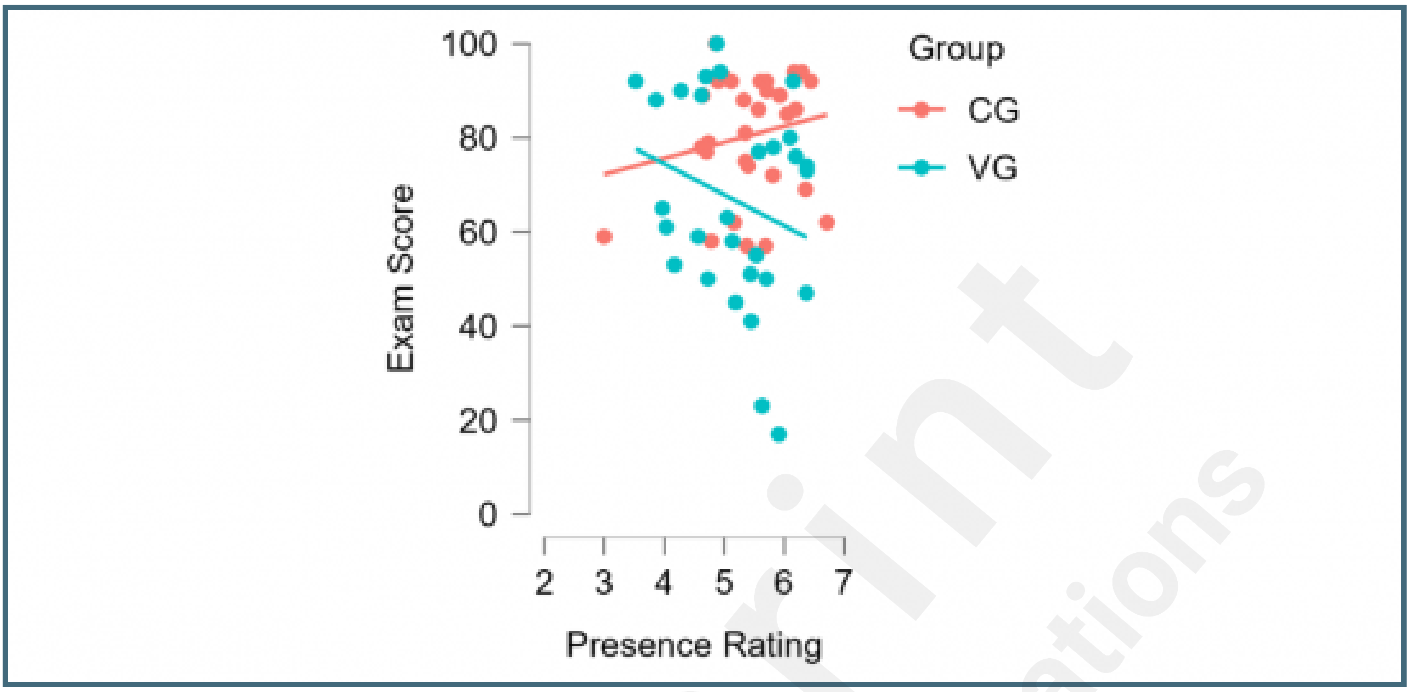
Time spent on the advanced training stage for CG group and VG group in minutes. Dots indicate each participants' time. The boxes indicate interquartile ranges; the line in the middle indicates the median. The figure is created by using Jasp [35].



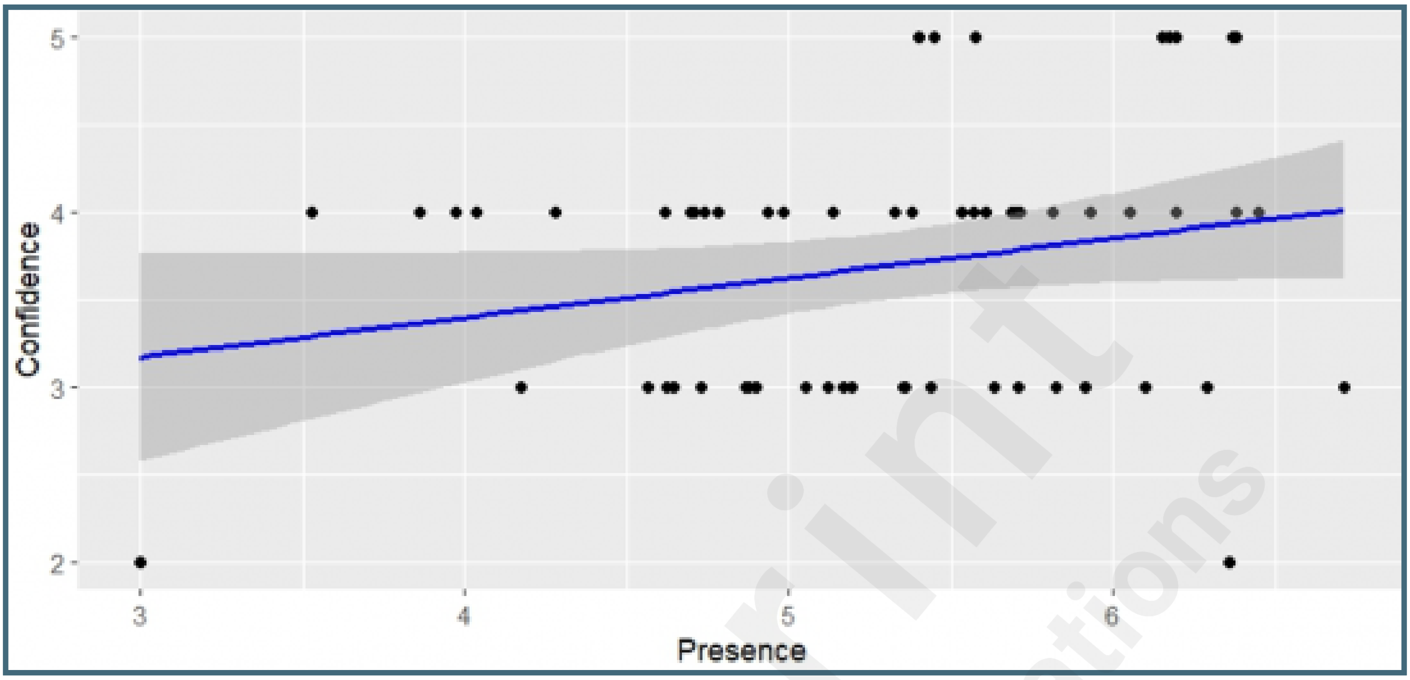
Confidence bias for CG group and VG group. Dots indicate each participants' confidence bias. The boxes indicate interquartile ranges; the line in the middle indicates the median. The figure is created by using Jasp [35].



Relationship between presence rating and exam score for CG and VG groups. Dots indicate each participants’ presence rating and exam score. The figure is created by using Jasp [35].



Regression line for presence as predictor variable and confidence as the outcome variable. Dots indicate each participants' presence rating and confidence. The figure is created by using RStudio [40].



CONSORT (or other) checklists

CONSORT-EHEALTH Checklist.
URL: http://asset.jmir.pub/assets/5f0cae15806d96b5031cca2bc2ea1bf2.pdf

TOC/Feature image for homepages

Voice Command vs. VR Controller.