Journal of Medical Education for Future Demands
Volume 23 · Issue 1 · e151030

OPEN PEER REVIEW

The Effect of Virtual Reality-Based Teaching (VRBT) on Nursing Students' Learning Performance and Engagement in Anatomy

Behnam Rasouli

SUBMITTED	REVISED	ACCEPTED	DOI
11 Aug 2024	14 Dec 2024	23 Dec 2024	10.5812/jme-151030

About this report

This open peer review (OPR) report shows the public comments made on this article during peer review, by the reviewers, the associate editor and the editor-in-chief, for every revision round, together with the authors' replies. Comments appear only with the consent of the people who wrote them. Reviewers who chose to sign their report are named; the others are shown as "Referee". Comments are edited for language only, and confidential remarks to the editors are not included.

More about open peer review: brieflands.com/knowledgebase/opr

Contents

1 **Original submission** Afagh Zarei, Referee 1 2
2 **Revision 1** Referee 1 4
Authors' reply to reviewers 5
Reviewer and editor comments 12
3 **Revision 2** Referee 1, Referee 2 13
Authors' reply to reviewers 14
Reviewer and editor comments 16
4 **Revision 3** Referee 1 17
Authors' reply to reviewers 18
Reviewer and editor comments 26

1 Original submission

2 comments · 13 Aug 2024 – 25 Aug 2024

Afagh Zarei Reviewer

MAJOR REVISION 13 Aug 2024

Thank you for your valuable and very interesting study. Also, thank you for the detailed reporting of the results of the evaluation of the effects of VR use on the learning, engagement, and attitudes of nursing students. I have reviewed this manuscript and provided some suggestions as follows. I hope they will be helpful.

1. Structure of manuscript: Structure the manuscript according to the journal's suggested structure.

2. Introduction: It is not evident why VR is used in the classroom. The use of technology in the classroom must be justified because technologies are diverse and each has its own capabilities. The fact that a technology is emerging cannot alone be a good reason to use it, especially if that technology is expensive.

3. What are the new methods that are used in anatomy education and have a lower cost, and what are the shortcomings that virtual reality can address?

4. The knowledge gap is not well explained in this study.

5. In the study design, define abbreviations when they first appear in the manuscript. For instance: TL on page number 4.

6. Although in this study, like most studies in the field of education, you had no choice but to use the students of one classroom, the sample size should still be calculated so that the reader can decide on the power of the study and the use of its results in another setting. Describe the population: the level of the students.

7. The procedure needs more clarification. 1. Describe how you assigned the students to the intervention and control groups. 2. How was the instruction through VR conducted? Describe what you summarized in figure 2. 3. What was the time sequence of instruction in the intervention and control groups? 4. Were the content and teacher the same in the two groups? 5. What were the subjects taught in both groups? 6. How did you make sure that the samples in the two groups were equivalent at the outset? 7. Did the instruction in each group last 8 weeks? And did you have 56 sessions of 60 minutes during the 8 weeks? 8. Did you use applications or VR glasses for the intervention group? Please clarify.

8. The information in both figure 1 and 2 can be reported in tables rather than figures.

9. The titles of figures and tables need more description.

10. Regarding the Academic Achievement Test, please provide more information about the similarity between the pretest and posttest. Were the pretest and posttest the same in the intervention and control groups? What is a single choice question? We have Multiple Choice Questions (MCQs); maybe you meant best answer questions. Did you use Cronbach's alpha for vacancy fillers and short answer questions too? How did you ensure that these Academic Achievement Tests assess all objectives of the course?

11. How was the Self-reported Perceptions of Engagement designed? Was it a self-developed or a standard questionnaire? What was the validity of the questionnaire? What were the questions about? What were the minimum and maximum scores of the questionnaire?

12. How was the Attitude Toward VRBT Questionnaire designed? Was it a self-developed or a standard questionnaire? What was the validity of the questionnaire? What were the questions about? What were the minimum and maximum scores of the questionnaire?

13. Interview: What were your questions in the interviews? What were the objectives of the interview? Where were the interviews conducted?

14. Analysis: 1. How were the qualitative data analyzed? 2. How was the normality of the quantitative data evaluated?

15. Results: Was there any difference between the intervention and control groups in terms of gender and age?

16. In table 3, please clarify which test was used to obtain each significance value.

17. Since the views of students toward Virtual Reality-Based Teaching were investigated through interviews, the results of these interviews are expected to be reported in a qualitative research format.
18. The discussion section needs further elaboration and explanation of the results.
19. What were the limitations of your study?
20. Has ethical permission been obtained to conduct this study? If so, please mention the ethical code.

Referee 1 Reviewer

MAJOR REVISION 25 Aug 2024

Overall, while the study addresses an important area in nursing education, several methodological issues must be addressed to strengthen the validity and reliability of the findings. We recommend that the authors revise the manuscript to incorporate these suggestions, providing a clearer rationale for their methods and a more thorough discussion of their results.

2

Revision 1

AUTHORS' RESPONSE TO REVIEWERS

Authors' reply to reviewers

The next pages contain the authors' point-by-point reply to the comments of this revision round, as submitted with the revised manuscript. It is published as the authors wrote it.

Dear Editor and reviewers:

Thank you for giving us the opportunity to submit a revised draft of our manuscript titled The Effect of Virtual Reality-Based Teaching (VRBT) on Nursing Students' Learning Performance and Engagement in Anatomy to Journal of Medical Education for Future Demands. I appreciate the time and effort that you and the reviewers have dedicated to providing your valuable feedback on my manuscript. I am grateful to the reviewers for their insightful comments on my paper. I have been able to incorporate changes to reflect most of the suggestions provided by the reviewers. I have highlighted the changes within the manuscript. Also, many of the things mentioned by the respected reviewers were in the original version of the article, but for the magazine, due to the limitation of the number of words, I had removed many of them.

Reviewer(s)' Comments to Author:

Comment 1. Structure of manuscript: Structure the manuscript according to suggested structure of journal.
Response: Thank you for pointing this out We agree with this and have incorporated your suggestion throughout the manuscript.

Comment 2. Introduction: it is not evident what is the reason of using VR in the classroom. There must be justification in using technology in the classroom. Because the technologies are diverse and each has its own capabilities. Emerging technology alone cannot be a good reason to use it, especially if that technology is expensive.
Response: Thank you for pointing this out We agree with this and have changed this point according below:
The use of VR in anatomy education creates a 3D representation of the human body. This immersive experience allows students to see anatomical systems from multiple angles and perspectives and to deeply reinforce the spatial relationships between body tissues. Also, in virtual reality, students manipulate 3D models and participate in virtual dissection, which leads to better learning and satisfaction from the educational experience [Page 2,line 12].

Comment 3. What are the new methods that are used in anatomy education and have a lower cost, and what are the defects that virtual reality can cover?
Response: Thank you for pointing this out we agree with this and have changed this point according below:
The use of virtual reality in anatomy education makes the learning experience accessible and affordable, and in contrast to the old and existing methods in anatomy education, it can cause deeper learning[Page 2,line 26].
Comment 4. Knowledge gap is not well explained in this study.
Response: Thank you for pointing this out we agree with this and have changed this point according below:
The challenge of the current research is that one of the major drawbacks of conventional methods is the limited access to corpse samples. Many universities are having trouble preparing corpse samples, and virtual reality can be a solution to the gap[Page 2,line 18].
Comment 5. In study design, describe abbreviation for the first time in the manuscript. For instance: TL in page number 4.
Response: Thank you for pointing this out we agree with this and have incorporated your suggestion throughout the manuscript for example Page 2,line 37.

Comment 6. Although in this study, like most studies in the field of education, you had no choice but to use the students of one classroom, but still the sample size should be calculated so that the reader can decide on the power of the study and the use of its results in another setting. Describe the population: the level of students,

Response: Thank you for pointing this out we agree with this and have changed this point according below:

The participants were 70 students enrolled in anatomy course that we choose 62 according to Krejcie & Morgan 's table. [Page 3,line 3]

Comment 7. The procedure needs more clarification. 1. Describe how assign the students in the intervention and control group. 2. How was the instruction through the VR done? Describe what you summarized in figure 2. 3. What time sequence was the instruction in the intervention and control groups? 4. Were the content and teacher same in two groups? 5. What were the subjects taught in both groups? 6. How did you make sure that the input samples in the two groups are equivalent? 7. Did the instruction in each group last 8 weeks? And did you have 56 session of 60 minutes during the 8 weeks? 8. Did you use applications or VR glasses for intervention group? please clarify.

Response :Thank you for pointing this out we agree with this and have changed this point according below:

Then, we tested VR devices and lessons for effective performance in VRBT. At the same time, the group that used VR had to familiarize themselves with the VR device and the classroom structure, which reduced the effects of the new technology on student learning. During six weeks from the second to the seventh week, the professor taught anatomy in the VRBT group. The professor in the VRBT group introduced the content and prerequisites of the lesson every week in 10 minutes. Then, the students performed the activities specified by the professor to learn scientific material and experience virtual scenes in VR. Due to the lack of enough VR devices for students, we randomly divided them into 10 subgroups of three people and they used Headset VR Box 105 while learning. The mobile phone was placed inside VR Box 105, and wired headset earphone were connected to it. The students used the VR applications (Table 2) through the VR Box 105 remote controller. At the same time, they viewed images and listened to audio with wired headset earphone. Each student experienced approximately 7 to 8 minutes of virtual scenes. Next, each subgroup performed face-to-face classroom activities where the checklist including educational activities (printed materials) was reviewed and completed (Figure 2). TL group were taught through traditional learning method. The students of this group also used the Moodle cloud and educational materials online. This group, which was randomly divided into six subgroups of 5 students, who learned the same subjects with face-to-face teaching methods along with the Moodle cloud in a regular classroom. In each session, the teacher introduced the lesson content using PowerPoint and videos. After that, each subgroup learned the material through the Moodle cloud and anatomy textbooks and did learning activities and assignments together. In the Moodle cloud, there was a video, a PDF, an assignment that the student used in each session. After the students' discussion, the teacher summarized the answers (Figure 2). In order to ensure the same educational content in VRBT and TL groups, two anatomy professors outside the research reviewed the content of the application programs in Table 1, then determined the educational content for the TL according to that, and that content in TL was taught. Also, for the confounding factors in the TL group, Moodle cloud was an online platform where educational materials were placed, and it was not a distinct method that could affect the performance of the TL. Therefore, this platform cannot be a confusing factor on the TL and the samples of both groups (TL and VRBT) were familiar with the Moodle Cloud environment and they were the same in terms of familiarity with this environment. [Page 3,line 9]

Comment 8. Both figure 1 and 2 can report in the tables not figures.
Response: Thank you for pointing this out we think that figures 1 and 2, if they are in the form of figures, will show the research process and method better.

Comment9. The titles of figures and tables need more description.
Response: Thank you for pointing this out we agree with this and have incorporated your suggestion throughout the manuscript.

Comment10. About Academic Achievement Test, please provide more information about the similarity of two pretest and posttest? Were the pretest and posttest the same in intervention and control groups? What is single choice question? We have Multiple Choice Questions (MCQs); maybe you meant best answer questions. Did you use Cronbach's alpha for vacancy fillers and short answer questions too? How did you ensure that these Academic Achievement Tests assess all objectives of the course?
Response: Thank you for pointing this out we agree with this and have changed this point according below:
To check the performance of the students, a test consisting of a pre-test and a post-test was developed with the help of two experienced anatomy professors. To reduce the effect of similarity between tests on learners, the content of pre-test and post-test was designed differently. The pre-test consisted of 25 items that focused on assessing that prior knowledge of pre-participation student anatomy was equivalent. This consisted of 10 true-false and 15 single-choice items (total score = 40). The post-test consisted of 24 items that measured anatomy knowledge and students' performance in learning units. These included eight single-choice questions, eight true-false questions, five vacancy fillers, and three short response items. The reliability of pre-test and post-test was calculated using Cronbach's alpha test, both tests were greater than 0.80 and their internal consistency was acceptable. The reliability of the pre-test and post-test was measured using Cronbach's alpha values, both of which were greater than 0.80 and had acceptable internal consistency. A group of experts including three teachers confirmed the content validity of the test. The last version of the pre-test and post-test was reviewed by both anatomy professors and a group of experts. We removed all obstacles to ensure high validity and reliability. The pre-test and post-test in this research were in Persian language. [Page 5,line 3]

Comment 11. How was the Self-reported Perceptions of Engagement designed? Was it a self-developed or a standard questionnaire? What was the validity of the questionnaire? What were the questions about? What was the minimum and maximum score of the questionnaire?
Response: Thank you for pointing this out we agree with this and have changed this point according below:
To measure students' engagement, the researchers designed a " Self-reported Perceptions of Engagement " based on Reeve's (33) research. To design the draft questionnaire, data was collected from 50 undergraduate students. The researchers analyzed the collected data and based on that they designed a draft questionnaire consisting of 45 questions. According to the opinion of experts from several universities, unrelated and similar items were removed from the draft version. The revised draft included 37 questions. It was graded with a 5-point Likert scale in five categories including completely agree (5), agree (4), partially agree (3), disagree (2), completely disagree (1). Kaiser Meyer-Olkin (KMO) and Bartlett sphericity tests were used for factor analysis (Fraenkel & Wallen, 2006). The value of KMO equal to 0.94 was obtained,

which was considered very good. Also, the result of Bartlett's test was significant and for factor analysis, the data were identified suitable ($\chi 2=8530,19$; $p<,000$). The internal consistency coefficient was calculated with Cronbach's alpha, which was 0.950, and this value was reliable. KMO value was obtained 0.93 and the researchers considered it very good. The result of Bartlett's test was significant ($\chi 2=7719,09$; $p<,000$). This questionnaire had four factors and explained 51.11% of the total variance. The factors had high internal consistency. (Behavioral (13 items) ,863; emotional (10 items),834; cognitive (7 items) ,825; agentic (6 items),711). [Page 5,line 16]

Comment12. How was the Attitude Toward VRBT Questionnaire designed? Was it a self-developed or a standard questionnaire? What was the validity of the questionnaire? What were the questions about? What was the minimum and maximum score of the questionnaire?
Response: Thank you for pointing this out we agree with this and have changed this point according below:
An attitude questionnaire was created by modifying an attitude questionnaire was created by modifying an existing scale(23). The attitude questionnaire with a five-point response scale (completely agree (5), agree (4) to completely disagree (0)) was developed in order to measure the attitude of the treatment group toward VRBT. The questionnaire included ten questions about system operation and learning activities in VRBT. The result of the KR20 attitude questionnaire was more than 0.81, which was higher than the reliability standard (0.7). [Page 5,line 33]
Comment 13. Interview: what were your questions in the interviews? what was the objectives of the interview? Where was the interviews done?
Response : Thank you for pointing this out We agree with this and have changed this point according below:
Semi-structured interviews were used to investigate participants' satisfaction with the process of using VR in the classroom. 11 students were randomly selected from the VRBT group. The semi-structured interview consisted of 11 open questions.

The participants in the research were briefly interviewed. Interview data was collected. To extract the views of students towards VRBT, two of the authors reviewed the interviews in a blind review and selected the most relevant sentences, and then the sentences that were common between both authors were selected. The interview questions included the following:

1- Have you worked with VR software before?
2- How did you feel about using VR glasses in class?
3- Do you learn more from this method or from the usual method?
4- What was the difference between this form of teaching and the classes you had so far?
5- Do you think you will learn more with this software?
6- Was there a technical problem while working with this software?
7- What was the difference between this type of training and regular training?
8- Does working with this software increase your motivation to learn?
9- Are you interested in using this software in other lessons?
10- Did you feel tired after using VR glasses several times?
11- Do you suggest using this method for other courses and other students? [Page 6,line 2]

Comment 14. Analysis: 1. How was the qualitative data analyzed? 2. How did the normality of quantitative data evaluate?
Response : Thank you for pointing this out we agree with this and have changed this point according below:
Thematic analysis was used to analyze the data. The students expressed that they felt excited, happy and valued by knowing about the use of VR in this course. Two students also stated that they were hesitant because they had no previous experience working with VR technology in the classroom. But interviews revealed that the motivation of students had increased:

Student1: “I was glad when I first heard. I hesitate a little bit. We were faced with things we didn’t know how to do them”.

Student3: “I felt nice things. I was excited and happy. I felt valuable in this course.”
Student4: “Access to the resources and samples any time kept me motivated.”
Student5: “It was very nice to me. I walked constantly around mobile. I continuously was mobile.”

Students said that VR had a positive effect on their learning and had effect on their motivation:

Student 2: “structures of anatomy was very clear and obvious which I got through VR device.”

Student 6: “I developed a strong interest in anatomy lessons after the VRBT experience.”

Student 7: “I filled parts with resources when I didn’t understand during course study and I tried to figure out the issues.”

Students had technical issues while researching internet connectivity, tablets, app and other technical issues. Also, in their opinion, VR can be used in other lessons by removing technical barriers. Students also gave advice on using VR in lesson:
Student2: “The information, which I encountered relevant and irrelevant, caused confusion. I think the persistence will be increased if it is used in practical courses.”
Student3: “I want to reach the course notes at the end of the course.”

Student8: “sometimes I was confused, because there many photos and information about topics”

Student 11: “After using the VR glasses several times, I felt tired” [Page 10,line 3]

Comment 15. Results: Was any difference between intervention and control groups in term of gender and age?
Response: there was not any difference between intervention and control groups in term of gender and age.
Comment 16. In table 3, please clarify with which test each significant value was obtained?

Response : Thank you for pointing this out we agree with this and have changed this point according below:
Students' previous knowledge of anatomy was examined with a pre-test before the learning activities. According to Table 3, the result of the independent t-tests (VRBT and TL groups: $t = 0.67$, $p = 0.82 > .05$) showed that there was no significant difference in academic achievement in the pre-test. This shows that the previous knowledge in the groups were probably equal. Also, the result of the independent t-tests (VRBT and TL groups: $t = 2.27$, $p = 0.05$) showed that

in the post-test there was a significant difference between the two groups and the students in the VRBT group had higher scores than the TL group in the achievement test (Table 3 and Figure 4). In addition, as shown in Table 3 based on paired t-tests there was significant difference in academic achievement in the pre-test and post-test of VRBT group, while there was no significant difference in academic achievement in the pre-test and post-test of TL group. [Page 7,line 1]

Comment 17. Since, the views of students towards Virtual Reality-Based Teaching investigated through interviews, it is expected to report the results of these interviews in format of qualitative research.

Response : Thank you for pointing this out We agree with this. But we think that if we mention the statements of the students themselves without coding, the results will be more effective for the reader as follows:

Student1: “I was glad when I first heard. I hesitate a little bit. We were faced with things we didn’t know how to do them”.

Student3: “I felt nice things. I was excited and happy. I felt valuable in this course.”

Student4: “Access to the resources and samples any time kept me motivated.”

Student5: “It was very nice to me. I walked constantly around mobile. I continuously was mobile.” [Page 10,line 3]

Comment.18. The discussion section needs further elaboration and explanation on the results.

Response : Thank you for pointing this out We agree with this and have changed this point according below:

In this study, we used mobile VR software for anatomy lessons. In order to evaluate the effectiveness of the intervention method, an experiment was conducted in four units of anatomy course on students in a medical school. The results showed that students in VRBT group had significantly higher academic achievement compared to those who were taught in TL. These findings were in line with the results of past studies showed that VR had significant effect on increasing the learning rate(1, 3, 5, 7, 8, 24, 27-29). According to this study, three specific reasons illustrate the benefits of VR. First, the greater success of the VRBT group may be result of the positive characteristics of VR. Also, secondly, face-to-face training using VR is better for solving complex problems and creating knowledge, while computer environments alone are more suitable for creating connections between ideas and interpretations(4-6).Therefore, collaborative learning activities in VR-based environment can provide rich and deep learning experience (4,5,6). Also, VR may increase the conceptual knowledge of anatomy through the visual representation of the hidden structure of the content and quick feedback to the learner in the virtual space(8-10, 29, 30).

The results also showed that students who took VRBT lessons did not show more engagement than those who were taught through TL methods. These results are contrary to previous studies showed that learning with VR is effective on engagement in anatomy learning as well as traditional methods(12, 13, 22)). More studies are needed in this field. In addition, it is suggested to consider a seven-week interval between the intervention and the conflict assessment to check the validity again and compare with longer time interval for the reliability test. To evaluate participation in the teaching method, there must be a certain period of time between the experiment and the test period. Researchers have suggested that a six-month

interval is necessary to obtain reliable results(12). In the present study, instead of six months, a time interval of eight weeks was used, because firstly, the study program of the research participant was heavy, the study was conducted in two months. Also, the foundation block lasted eight weeks. For this reason, the evaluation was done at the end of the block.

Also, the results showed that students in the VRBT group had favorable tendency to use VR in classroom. This implies that *VRBT* had a good user interface and increases interest in learning. This result is consistent with previous studies showed that the integration of VR in anatomical lessons can have positive effects on the attitude of learners(28-30) . In addition, the interview showed that the VRBT lessons were more interesting for the learners and confirmed that they were interested in learning more from the HMD than the conventional methods of learning in the classroom. The VR model stimulated students' motivation, learning and sense of participation by creating an inclusive environment, encouraged them to reflect on the content and engage in learning experiences. The VR model stimulated students' motivation, learning and sense of participation by creating an inclusive environment, encouraged them to reflect on the content and engage in learning experiences. [Page 11,line 10]

Comment 19. What were the limitations of your study?
Response : Thank you for pointing this out we agree with this and have changed this point according below:
There were several limitations in this study. First, the results of the research showed that VR is effective on the learning of fourth-year nursing students, which means that it cannot be generalized to other students with different ages and semesters. Designing or implementing VR for anatomy learning activities in the classroom should consider more characteristics of learners (Learning style, mental imagery ability and passion for learning). Cheng & Tsai (2020) suggested, future studies could examine immersion as a factor for attention and enjoyment. Also, we can refer to a study (22) that in designing VR-based learning environments, VR capabilities such as direct control should be considered. Furthermore, although the present study showed that VR can provide immediate feedback for student learning, can this type of VR design that includes feedback elements be as effective as many other virtual learning spaces or traditional learning environments? Therefore, in the future, VR feedback elements should be educationally and technologically designed to improve anatomy learning. In addition, since groups of students use the HMD in the classroom for a short period of time, future studies should provide the students with the HMD in advance and then examine their activities in the VR environment. Finally, for the effectiveness of using VR in anatomy classes, long-term research should be conducted. [Page 12,line 7]
Comment 20. Has ethical permission been obtained to conduct this study? if so, please mention the ethical code.
Response : Thank you for pointing this out we did not have an ethical code in this study but Students acknowledged their willingness to use the VR tools and participate in the study voluntarily and provided consent, knowing their information would be used anonymously. Written informed consent was obtained from each participant prior to his or her inclusion in the study.

2 Revision 1

1 comment · 19 Oct 2024

Referee 1 Reviewer

MAJOR REVISION 19 Oct 2024

Comment 3 has not been properly addressed. You did not introduce the new methods, and you only mentioned virtual reality (VR).
Please respond to each of the questions in Comment 10 again.
What exactly are the questions in the Self-Reported Perception Questionnaire? Please clarify this in the explanation of the questionnaire.
The Attitude Questionnaire is an adapted version of existing questionnaires. You need to specifically name the existing questionnaire.
Please have the translation reviewed by a native speaker and provide a certificate for it. I recommend the Radan English Institute for this.

3

Revision 2

AUTHORS' RESPONSE TO REVIEWERS

Authors' reply to reviewers

The next pages contain the authors' point-by-point reply to the comments of this revision round, as submitted with the revised manuscript. It is published as the authors wrote it.

Dear Editor and reviewers:

Thank you for giving us the opportunity to submit a revised draft of our manuscript titled The Effect of Virtual Reality-Based Teaching (VRBT) on Nursing Students' Learning Performance and Engagement in Anatomy to Journal of Medical Education for Future Demands. I appreciate the time and effort that you and the reviewers have dedicated to providing your valuable feedback on my manuscript. I am grateful to the reviewers for their insightful comments on my paper. I have been able to incorporate changes to reflect most of the suggestions provided by the reviewers. I have highlighted the changes within the manuscript. Also, many of the things mentioned by the respected reviewers were in the original version of the article, but for the magazine, due to the limitation of the number of words, I had removed many of them.

Reviewer(s)' Comments to Author:

Comment 3. What are the new methods that are used in anatomy education and have a lower cost, and what are the defects that virtual reality can cover?

Response: Thank you for pointing this out we agree with this and have changed this point according below:
3d anatomy software and applications, virtual dissection, augmented reality, online learning platforms, 3d printing and peer-to-peer learning are the new methods that are used in anatomy education. While VR can significantly enhance anatomy education by providing interactive and engaging learning experiences, its limitations compared to traditional methods emphasize the value of a blended approach. Combining both VR and traditional dissection methods can lead to a more comprehensive understanding of human anatomy, incorporating the strengths of each method. [Page 2,line 15].

Comment10. About Academic Achievement Test, please provide more information about the similarity of two pretest and posttest? Were the pretest and posttest the same in intervention and control groups? What is single choice question? We have Multiple Choice Questions (MCQs); maybe you meant best answer questions. Did you use Cronbach's alpha for vacancy fillers and short answer questions too? How did you ensure that these Academic Achievement Tests assess all objectives of the course?
Response: Thank you for pointing this out we agree with this and have changed this point according below:
Two anatomy professors assisted in developing a test, including a pre-test and a post-test, to evaluate student performance. The pre-test and post-test content were designed in a way to minimize the impact of test similarity on learners. The pre-test included 25 questions aimed at evaluating whether students had the same level of knowledge on pre-participation anatomy. This included 10 items for true or false and 15 items for multiple choice (total score = 40). The post-test included 24 items evaluating anatomy knowledge and students' progress in learning modules. This consisted of eight multiple-choice questions, eight true-false questions, five vacancy fillers, and three short response items. Cronbach's alpha was used to calculate the reliability of both the pre-test and post-test for the multiple-choice test, with values exceeding 0.80 indicating acceptable internal consistency. Additionally, we employed Kuder-Richardson Formula 21 for different types of inquiries, with all questions scoring higher than 0.80. A panel of experts, which included three educators, verified that the test has content validity. To guarantee that Academic Achievement Tests cover all course objectives, the most recent pre-test and post-test were examined by anatomy professors and a panel of experts. We eliminated every barrier to guarantee strong validity and reliability. The intervention and control groups had the same pre-test and

post-test results. The pre-test and post-test utilized in this study were conducted in the Persian language.[Page 4,line 3].

New Comment: The Attitude Questionnaire is an adapted version of existing questionnaires. You need to specifically name the existing questionnaire

Response: Thank you for pointing this out we agree with this and have changed this point according below:

An attitude questionnaire was created by modifying an attitude questionnaire was created by modifying attitude scale in research of *Li, Mayer, Malmberg* (23). The attitude questionnaire with a five-point response scale (completely agree (5), agree (4) to completely disagree (0)) was developed in order to measure the attitude of the treatment group toward VRBT. The questionnaire included ten questions about system operation and learning activities in VRBT. The result of the KR20 attitude questionnaire was more than 0.81, which was higher than the reliability standard (0.7).. [Page ,line 21].

3 Revision 2

2 comments · 21 Nov 2024 – 9 Dec 2024

Referee 1 Reviewer

MINOR REVISION 21 Nov 2024

Please review the comments again.

Referee 2 Reviewer

ACCEPT 9 Dec 2024

Hello dear author
I appreciate your efforts in addressing the feedback provided during the review process and believe that your manuscript has been significantly improved as a result. Congratulations.

4

Revision 3

AUTHORS' RESPONSE TO REVIEWERS

Authors' reply to reviewers

The next pages contain the authors' point-by-point reply to the comments of this revision round, as submitted with the revised manuscript. It is published as the authors wrote it.

Dear Editor and reviewers:

Thank you for giving us the opportunity to submit a revised draft of our manuscript titled The Effect of Virtual Reality-Based Teaching (VRBT) on Nursing Students' Learning Performance and Engagement in Anatomy to Journal of Medical Education for Future Demands. I appreciate the time and effort that you and the reviewers have dedicated to providing your valuable feedback on my manuscript. I am grateful to the reviewers for their insightful comments on my paper. I have been able to incorporate changes to reflect most of the suggestions provided by the reviewers. I have highlighted the changes within the manuscript. Also, many of the things mentioned by the respected reviewers were in the original version of the article, but for the magazine, due to the limitation of the number of words, I had removed many of them.

Reviewer(s)' Comments to Author:

Comment 1. Structure of manuscript: Structure the manuscript according to suggested structure of journal.
Response: Thank you for pointing this out We agree with this and have incorporated your suggestion throughout the manuscript.

Comment 2. Introduction: it is not evident what is the reason of using VR in the classroom. There must be justification in using technology in the classroom. Because the technologies are diverse and each has its own capabilities. Emerging technology alone cannot be a good reason to use it, especially if that technology is expensive.
Response: Thank you for pointing this out We agree with this and have changed this point according below:
The use of VR in anatomy education creates a 3D representation of the human body. This immersive experience allows students to see anatomical systems from multiple angles and perspectives and to deeply reinforce the spatial relationships between body tissues. Also, in virtual reality, students manipulate 3D models and participate in virtual dissection, which leads to better learning and satisfaction from the educational experience [Page 2,line 12].

Comment 3. What are the new methods that are used in anatomy education and have a lower cost, and what are the defects that virtual reality can cover?

Response: Thank you for pointing this out we agree with this and have changed this point according below:
3d anatomy software and applications, virtual dissection, augmented reality, online learning platforms, 3d printing and peer-to-peer learning are the new methods that are used in anatomy education. While VR can significantly enhance anatomy education by providing interactive and engaging learning experiences, its limitations compared to traditional methods emphasize the value of a blended approach. Combining both VR and traditional dissection methods can lead to a more comprehensive understanding of human anatomy, incorporating the strengths of each method. [Page 2,line 15].
Comment 4. Knowledge gap is not well explained in this study.
Response: Thank you for pointing this out we agree with this and have changed this point according below:
The challenge of the current research is that one of the major drawbacks of conventional methods is the limited access to corpse samples. Many universities are having trouble

preparing corpse samples, and virtual reality can be a solution to the gap[Page 2,line 18].
Comment 5. In study design, describe abbreviation for the first time in the manuscript. For instance: TL in page number 4.
Response: Thank you for pointing this out we agree with this and have incorporated your suggestion throughout the manuscript for example Page 2,line 37.

Comment 6. Although in this study, like most studies in the field of education, you had no choice but to use the students of one classroom, but still the sample size should be calculated so that the reader can decide on the power of the study and the use of its results in another setting. Describe the population: the level of students,
Response: Thank you for pointing this out we agree with this and have changed this point according below:
The participants were 70 students enrolled in anatomy course that we choose 62 according to Krejcie & Morgan 's table. [Page 3,line 3]
Comment 7. The procedure needs more clarification. 1. Describe how assign the students in the intervention and control group. 2. How was the instruction through the VR done? Describe what you summarized in figure 2. 3. What time sequence was the instruction in the intervention and control groups? 4. Were the content and teacher same in two groups? 5. What were the subjects taught in both groups? 6. How did you make sure that the input samples in the two groups are equivalent? 7. Did the instruction in each group last 8 weeks? And did you have 56 session of 60 minutes during the 8 weeks? 8. Did you use applications or VR glasses for intervention group? please clarify.
Response :Thank you for pointing this out we agree with this and have changed this point according below:
Then, we tested VR devices and lessons for effective performance in VRBT. At the same time, the group that used VR had to familiarize themselves with the VR device and the classroom structure, which reduced the effects of the new technology on student learning. During six weeks from the second to the seventh week, the professor taught anatomy in the VRBT group. The professor in the VRBT group introduced the content and prerequisites of the lesson every week in 10 minutes. Then, the students performed the activities specified by the professor to learn scientific material and experience virtual scenes in VR. Due to the lack of enough VR devices for students, we randomly divided them into 10 subgroups of three people and they used Headset VR Box 105 while learning. The mobile phone was placed inside VR Box 105, and wired headset earphone were connected to it. The students used the VR applications (Table 2) through the VR Box 105 remote controller. At the same time, they viewed images and listened to audio with wired headset earphone. Each student experienced approximately 7 to 8 minutes of virtual scenes. Next, each subgroup performed face-to-face classroom activities where the checklist including educational activities (printed materials) was reviewed and completed (Figure 2). TL group were taught through traditional learning method. The students of this group also used the Moodle cloud and educational materials online. This group, which was randomly divided into six subgroups of 5 students, who learned the same subjects with face-to-face teaching methods along with the Moodle cloud in a regular classroom. In each session, the teacher introduced the lesson content using PowerPoint and videos. After that, each subgroup learned the material through the Moodle cloud and anatomy textbooks and did learning activities and assignments together. In the Moodle cloud, there was a video, a PDF, an assignment that the student used in each session. After the students' discussion, the teacher summarized the answers (Figure 2). In order to ensure the same educational content in VRBT and TL groups, two anatomy professors outside the research reviewed the content of the application programs in Table 1, then determined the educational content for the TL according to that, and that content in TL was taught. Also, for the confounding factors in the TL group,

Moodle cloud was an online platform where educational materials were placed, and it was not a distinct method that could affect the performance of the TL. Therefore, this platform cannot be a confusing factor on the TL and the samples of both groups (TL and VRBT) were familiar with the Moodle Cloud environment and they were the same in terms of familiarity with this environment. [Page 3,line 9]

Comment 8. Both figure 1 and 2 can report in the tables not figures.
Response: Thank you for pointing this out we think that figures 1 and 2, if they are in the form of figures, will show the research process and method better.

Comment9. The titles of figures and tables need more description.
Response: Thank you for pointing this out we agree with this and have incorporated your suggestion throughout the manuscript.

Comment10. About Academic Achievement Test, please provide more information about the similarity of two pretest and posttest? Were the pretest and posttest the same in intervention and control groups? What is single choice question? We have Multiple Choice Questions (MCQs); maybe you meant best answer questions. Did you use Cronbach's alpha for vacancy fillers and short answer questions too? How did you ensure that these Academic Achievement Tests assess all objectives of the course?
Response: Thank you for pointing this out we agree with this and have changed this point according below:
Two anatomy professors assisted in developing a test, including a pre-test and a post-test, to evaluate student performance. The pre-test and post-test content were designed in a way to minimize the impact of test similarity on learners. The pre-test included 25 questions aimed at evaluating whether students had the same level of knowledge on pre-participation anatomy. This included 10 items for true or false and 15 items for multiple choice (total score = 40). The post-test included 24 items evaluating anatomy knowledge and students' progress in learning modules. This consisted of eight multiple-choice questions, eight true-false questions, five vacancy fillers, and three short response items. Cronbach's alpha was used to calculate the reliability of both the pre-test and post-test for the multiple-choice test, with values exceeding 0.80 indicating acceptable internal consistency. Additionally, we employed Kuder-Richardson Formula 21 for different types of inquiries, with all questions scoring higher than 0.80. A panel of experts, which included three educators, verified that the test has content validity. To guarantee that Academic Achievement Tests cover all course objectives, the most recent pre-test and post-test were examined by anatomy professors and a panel of experts. We eliminated every barrier to guarantee strong validity and reliability. The intervention and control groups had the same pre-test and post-test results. The pre-test and post-test utilized in this study were conducted in the Persian language.[Page 4,line 3].

Comment 11. How was the Self-reported Perceptions of Engagement designed? Was it a self-developed or a standard questionnaire? What was the validity of the questionnaire? What were the questions about? What was the minimum and maximum score of the questionnaire?
Response: Thank you for pointing this out we agree with this and have changed this point according below:
To measure students' engagement, the researchers designed a " Self-reported Perceptions of Engagement " based on Reeve's (33) research. To design the draft questionnaire, data was collected from 50 undergraduate students. The researchers analyzed the collected data and based on that they designed a draft questionnaire consisting of 45 questions. According to the

opinion of experts from several universities, unrelated and similar items were removed from the draft version. The revised draft included 37 questions. It was graded with a 5-point Likert scale in five categories including completely agree (5), agree (4), partially agree (3), disagree (2), completely disagree (1). Kaiser Meyer-Olkin (KMO) and Bartlett sphericity tests were used for factor analysis (Fraenkel & Wallen, 2006). The value of KMO equal to 0.94 was obtained, which was considered very good. Also, the result of Bartlett's test was significant and for factor analysis, the data were identified suitable ($\chi 2$=8530,19; p<,000). The internal consistency coefficient was calculated with Cronbach's alpha, which was 0.950, and this value was reliable. KMO value was obtained 0.93 and the researchers considered it very good. The result of Bartlett's test was significant ($\chi 2$=7719,09; p<,000). This questionnaire had four factors and explained 51.11% of the total variance. The factors had high internal consistency. (Behavioral (13 items) ,863; emotional (10 items),834; cognitive (7 items) ,825; agentic (6 items),711). [Page 5,line 16]

Comment12. How was the Attitude Toward VRBT Questionnaire designed? Was it a self-developed or a standard questionnaire? What was the validity of the questionnaire? What were the questions about? What was the minimum and maximum score of the questionnaire?
Response: Thank you for pointing this out we agree with this and have changed this point according below:
An attitude questionnaire was created by modifying an attitude questionnaire was created by modifying an existing scale(23). The attitude questionnaire with a five-point response scale (completely agree (5), agree (4) to completely disagree (0)) was developed in order to measure the attitude of the treatment group toward VRBT. The questionnaire included ten questions about system operation and learning activities in VRBT. The result of the KR20 attitude questionnaire was more than 0.81, which was higher than the reliability standard (0.7). [Page 5,line 33]
Comment 13. Interview: what were your questions in the interviews? what was the objectives of the interview? Where was the interviews done?
Response : Thank you for pointing this out We agree with this and have changed this point according below:
Semi-structured interviews were used to investigate participants' satisfaction with the process of using VR in the classroom. 11 students were randomly selected from the VRBT group. The semi-structured interview consisted of 11 open questions.

The participants in the research were briefly interviewed. Interview data was collected. To extract the views of students towards VRBT, two of the authors reviewed the interviews in a blind review and selected the most relevant sentences, and then the sentences that were common between both authors were selected. The interview questions included the following:

1- Have you worked with VR software before?
2- How did you feel about using VR glasses in class?
3- Do you learn more from this method or from the usual method?
4- What was the difference between this form of teaching and the classes you had so far?
5- Do you think you will learn more with this software?
6- Was there a technical problem while working with this software?
7- What was the difference between this type of training and regular training?
8- Does working with this software increase your motivation to learn?
9- Are you interested in using this software in other lessons?

10- Did you feel tired after using VR glasses several times?
11- Do you suggest using this method for other courses and other students? [Page 6,line 2]

Comment 14. Analysis: 1. How was the qualitative data analyzed? 2. How did the normality of quantitative data evaluate?
Response : Thank you for pointing this out we agree with this and have changed this point according below:
Thematic analysis was used to analyze the data. The students expressed that they felt excited, happy and valued by knowing about the use of VR in this course. Two students also stated that they were hesitant because they had no previous experience working with VR technology in the classroom. But interviews revealed that the motivation of students had increased:

Student1: “I was glad when I first heard. I hesitate a little bit. We were faced with things we didn’t know how to do them”.

Student3: “I felt nice things. I was excited and happy. I felt valuable in this course.”
Student4: “Access to the resources and samples any time kept me motivated.”
Student5: “It was very nice to me. I walked constantly around mobile. I continuously was mobile.”

Students said that VR had a positive effect on their learning and had effect on their motivation:

Student 2: “structures of anatomy was very clear and obvious which I got through VR device.”

Student 6: “I developed a strong interest in anatomy lessons after the VRBT experience.”

Student 7: “I filled parts with resources when I didn’t understand during course study and I tried to figure out the issues.”

Students had technical issues while researching internet connectivity, tablets, app and other technical issues. Also, in their opinion, VR can be used in other lessons by removing technical barriers. Students also gave advice on using VR in lesson:
Student2: “The information, which I encountered relevant and irrelevant, caused confusion. I think the persistence will be increased if it is used in practical courses.”
Student3: “I want to reach the course notes at the end of the course.”

Student8: “sometimes I was confused, because there many photos and information about topics”

Student 11: “After using the VR glasses several times, I felt tired” [Page 10,line 3]

Comment 15. Results: Was any difference between intervention and control groups in term of gender and age?
Response: there was not any difference between intervention and control groups in term of gender and age.
Comment 16. In table 3, please clarify with which test each significant value was obtained?

Response : Thank you for pointing this out we agree with this and have changed this point according below:
Students' previous knowledge of anatomy was examined with a pre-test before the learning activities. According to Table 3, the result of the independent t-tests (VRBT and TL groups: t =

0.67, $p = 0.82 > .05$) showed that there was no significant difference in academic achievement in the pre-test. This shows that the previous knowledge in the groups were probably equal. Also, the result of the independent t-tests (VRBT and TL groups: $t = 2.27$, $p = 0.05$) showed that in the post-test there was a significant difference between the two groups and the students in the VRBT group had higher scores than the TL group in the achievement test (Table 3 and Figure 4). In addition, as shown in Table 3 based on paired t-tests there was significant difference in academic achievement in the pre-test and post-test of VRBT group, while there was no significant difference in academic achievement in the pre-test and post-test of TL group. [Page 7,line 1]

Comment 17. Since, the views of students towards Virtual Reality-Based Teaching investigated through interviews, it is expected to report the results of these interviews in format of qualitative research.

Response : Thank you for pointing this out We agree with this. But we think that if we mention the statements of the students themselves without coding, the results will be more effective for the reader as follows:

Student1: “I was glad when I first heard. I hesitate a little bit. We were faced with things we didn’t know how to do them”.

Student3: “I felt nice things. I was excited and happy. I felt valuable in this course.”
Student4: “Access to the resources and samples any time kept me motivated.”
Student5: “It was very nice to me. I walked constantly around mobile. I continuously was mobile.” [Page 10,line 3]

Comment.18. The discussion section needs further elaboration and explanation on the results.

Response : Thank you for pointing this out We agree with this and have changed this point according below:

In this study, we used mobile VR software for anatomy lessons. In order to evaluate the effectiveness of the intervention method, an experiment was conducted in four units of anatomy course on students in a medical school. The results showed that students in VRBT group had significantly higher academic achievement compared to those who were taught in TL. These findings were in line with the results of past studies showed that VR had significant effect on increasing the learning rate(1, 3, 5, 7, 8, 24, 27-29). According to this study, three specific reasons illustrate the benefits of VR. First, the greater success of the VRBT group may be result of the positive characteristics of VR. Also, secondly, face-to-face training using VR is better for solving complex problems and creating knowledge, while computer environments alone are more suitable for creating connections between ideas and interpretations(4-6).Therefore, collaborative learning activities in VR-based environment can provide rich and deep learning experience (4,5,6). Also, VR may increase the conceptual knowledge of anatomy through the visual representation of the hidden structure of the content and quick feedback to the learner in the virtual space(8-10, 29, 30).

The results also showed that students who took VRBT lessons did not show more engagement than those who were taught through TL methods. These results are contrary to previous studies showed that learning with VR is effective on engagement in anatomy learning as well as traditional methods(12, 13, 22)). More studies are needed in this field. In addition, it is suggested to consider a seven-week interval between the intervention and the conflict

assessment to check the validity again and compare with longer time interval for the reliability test. To evaluate participation in the teaching method, there must be a certain period of time between the experiment and the test period. Researchers have suggested that a six-month interval is necessary to obtain reliable results(12). In the present study, instead of six months, a time interval of eight weeks was used, because firstly, the study program of the research participant was heavy, the study was conducted in two months. Also, the foundation block lasted eight weeks. For this reason, the evaluation was done at the end of the block.

Also, the results showed that students in the VRBT group had favorable tendency to use VR in classroom. This implies that *VRBT* had a good user interface and increases interest in learning. This result is consistent with previous studies showed that the integration of VR in anatomical lessons can have positive effects on the attitude of learners(28-30) . In addition, the interview showed that the VRBT lessons were more interesting for the learners and confirmed that they were interested in learning more from the HMD than the conventional methods of learning in the classroom. The VR model stimulated students' motivation, learning and sense of participation by creating an inclusive environment, encouraged them to reflect on the content and engage in learning experiences. The VR model stimulated students' motivation, learning and sense of participation by creating an inclusive environment, encouraged them to reflect on the content and engage in learning experiences. [Page 11,line 10]

Comment 19. What were the limitations of your study?
Response : Thank you for pointing this out we agree with this and have changed this point according below:
There were several limitations in this study. First, the results of the research showed that VR is effective on the learning of fourth-year nursing students, which means that it cannot be generalized to other students with different ages and semesters. Designing or implementing VR for anatomy learning activities in the classroom should consider more characteristics of learners (Learning style, mental imagery ability and passion for learning). Cheng & Tsai (2020) suggested, future studies could examine immersion as a factor for attention and enjoyment. Also, we can refer to a study (22) that in designing VR-based learning environments, VR capabilities such as direct control should be considered. Furthermore, although the present study showed that VR can provide immediate feedback for student learning, can this type of VR design that includes feedback elements be as effective as many other virtual learning spaces or traditional learning environments? Therefore, in the future, VR feedback elements should be educationally and technologically designed to improve anatomy learning. In addition, since groups of students use the HMD in the classroom for a short period of time, future studies should provide the students with the HMD in advance and then examine their activities in the VR environment. Finally, for the effectiveness of using VR in anatomy classes, long-term research should be conducted. [Page 12,line 7]
Comment 20. Has ethical permission been obtained to conduct this study? if so, please mention the ethical code.
Response : Thank you for pointing this out we did not have an ethical code in this study but Students acknowledged their willingness to use the VR tools and participate in the study voluntarily and provided consent, knowing their information would be used anonymously. Written informed consent was obtained from each participant prior to his or her inclusion in the study.

Comment 21: The Attitude Questionnaire is an adapted version of existing questionnaires. You need to specifically name the existing questionnaire

Response: Thank you for pointing this out we agree with this and have changed this point according below:

An attitude questionnaire was created by modifying an attitude questionnaire was created by modifying attitude scale in research of *Li, Mayer, Malmberg* (23). The attitude questionnaire with a five-point response scale (completely agree (5), agree (4) to completely disagree (0)) was developed in order to measure the attitude of the treatment group toward VRBT. The questionnaire included ten questions about system operation and learning activities in VRBT. The result of the KR20 attitude questionnaire was more than 0.81, which was higher than the reliability standard (0.7). [Page ,line 21].

4 Revision 3

1 comment · 22 Dec 2024

Referee 1 Reviewer

ACCEPT 22 Dec 2024

I agree with the acceptance of the manuscript once the final comments provided [Revision 3] have been addressed.