| Article details: 2017-0110 | |
|---|---|
| Title | **Immersive and interactive virtual reality improves learning and retention of neuroanatomy in medical students: a randomized controlled study** |
| Authors | Chelsea Ekstrand BSc MA, Ali Jamal BSc, Ron Nguyen, Annalise Kudryk BSc, Jennifer Mann BSc, Ivar Mendez MD PhD |
| **Reviewer 1** | **Dr. David Pinzon** |
| Institution | University of Alberta, Edmonton, Alta. |
| General comments (author response in bold) | Overall, an excellent paper on the benefits of incorporating VR technology into medical education.<br><br>Methodology and statistical analysis were brilliantly performed, however, there are some minor queries I don't believe were addressed in the paper.<br><br>**Thank you for these positive comments of our work, we hope we can adequately address your concerns below.**<br><br>1. I feel is difficult to determine if it aids in measuring 3D structures. Perhaps a task oriented towards determining differences in distance within various brain structures will be easier to quantify.<br><br>**This is an interesting point and we have now discussed this in the Interpretation as an important future direction (lines 395-396). Our primary goal in this study was to examine whether or not students could learn the general spatial relationships of brain structures, which is a requirement of our curriculum, however it would be very interesting to examine the limits of specificity in their ability to understand these relationships.**<br><br>**"Further, it may be interesting to examine how VR may aid in quantification of the actual spatial distance between structures, rather than just relative positions."**<br><br>2. Not clear if the questions in the test inquired about the spatial distribution amongst themselves<br><br>**We have now included an appendix of the questions (Appendix B) that we hope will address this comment.**<br><br>3. Was there a "neurophobia" assessment pretest, or just subjective comparison after the VR exposure?<br><br>**Neurophobia was assessed after the study period and post-test in order to determine whether participants felt that their neurophobia was decreased after study. However, we acknowledge that it would have been an interesting measure to have performed prior to study and hope that future research will examine this further.**<br><br>4. In the appendix, it shows the 3D VR picture. Is this the similar view the subjects with the paper based test had?<br><br>**As stated in the methods, the paper-based group studied materials that have traditionally been used in our curriculum to teach neuroanatomy (i.e., 2D figures from a commonly used neuroanatomy textbook). There was a variety of views of each of the structures (i.e., axial, sagittal, coronal) to help them orient the structures in 3D space.** |
| **Reviewer 2** | **Dr. S.R. Lawson** |
| Institution | Virginia Commonwealth University, Continuing Professional Development, Richmond, Va. |
| General comments (author response in bold) | This paper addresses an important topic in academic medical education related to immersive and interactive virtual reality and the impact on learning retention of medical students in a neuroanatomy course.<br><br>I have these comments/suggestions/observations:<br><br>Methods:<br>1. Please clearly state your research questions along with the associated statistical test to answer them.<br><br>**We have now made this more explicit on lines 198-202:**<br><br>**"To address our primary research question (i.e., whether VR would lead to equivalent learning outcomes to paper-based study methods), two 4 (Test; Pre-test, Post-test, Retention pre-test, Retention post-test) x 2 (Group; VR, Paper-based) mixed-measures ANOVAs were conducted for the Test and Control Questions (Table 4). Independent samples t-tests were used to further examine these effects (Table 5)."**<br><br>2. A diagram of the research method would be helpful identifying the randomization, sample size for each group, and points of data collection.<br><br>**We have now included this as Figure 1.** |

3. Describe your sampling technique, ie convenience, stratified random, etc. From which population did you acquire the sample size of 64? Did participants volunteer?

**We have now made it more clear that recruitment was performed via email and posters to the first and second year medical students, whereby the participants volunteered to participate by contacting one of the experimenters (lines 91-100). They were then randomly assigned into either the VR of paper-based intervention groups.**

**"Participants consisted solely of 66 first or second year medical students from the University of Saskatchewan who were randomly assigned to either the VR or paper-based group (33 in each group) in blocks of four using an online randomization tool. Recruitment was performed via emails to all of the first and second year medical students and posters throughout the medical school, whereby interested participants contacted an experimenter to volunteer."**

4. How much time were the participants in the VR group given to become proficient in the VR navigation system?

**We have now clarified that they were given a five-minute tutorial by the experimenter to familiarize them with the VR environment and interface in the methods section on lines (137-138).**

Results:
5. Please provide a discussion of the practical significance of your findings.

**We have now made the practical significance of our findings more explicit in the Interpretation section (lines 348-352):**

**"Practically, integration of VR technology into medical education may be beneficial to 'self-taught' models of learning, whereby students may experience increased motivation to study and decreased neurophobia without compromising learning outcomes. This is supported by recent evidence from Stepan and colleagues (31) who that lecture style VR neuroanatomy training increased motivation without compromising learning outcomes."**

6. Please define the measure in VR and Paper columns. I assume this is the mean?

**We have now clarified that this is the mean**.

7. The interaction between Test and Question Type results appear missing in Table 3.

**We apologize for this mistake, it is now included.**

8. Typically multiple comparison tests are completed to determine differences between groups for study conditions, such as Scheffe, Tukey, Newman-Keuls, etc. It is inappropriate to use multiple t-tests when more than two comparisons are made within a single set of data. This is based on the desired protection against Type I error.

**Thank you for this point. We have now adjusted the alpha threshold for our multiple comparisons using a Bonferroni correction. Results significant at this more conservative threshold are shown in tables 5 and 7 with an asterisks.**

9. Provide a rationale for Not reporting the Sum of Squares.

**The sum of squares was not included because it is expressed in the mean square error term (MSE = SS/df), however we would be happy to include the sum of squares either in place or addition to the MSE if preferred.**

Discussion/Conclusion
10. How do your results compare to studies that compared 2D computer based instruction and paper-based study materials?

**Our results are consistent with other studies that have compared both computer and paper based neuroanatomy training with both 2D and 3D training methods, as highlighted in the Interpretation on lines 341-352:**

**"Based on both the quantitative and qualitative results of this study, there is evidence that VR technology can provide an effective supplemental learning tool for neuroanatomy by decreasing neurophobia and increasing knowledge retention. Our results are in concordance with studies using virtual 2D and 3D models (25-27), and suggest that learning complex, 3D relationships between neural structures is facilitated by training in a 3D VR environment. Our findings also corroborate those of Armstrong and colleagues (28) showing qualitative benefits of immersive and interactive VR compared to other methods, and extends them into the quantitative domain. Practically, integration of VR technology into medical education may be beneficial to 'self-taught' models of learning, whereby students may experience increased motivation to study and decreased neurophobia without compromising learning outcomes. This is supported by recent evidence from Stepan and colleagues (31) who that lecture style VR**

| | |
|---|---|
| | **neuroanatomy training increased motivation without compromising learning outcomes."**<br><br>12. Please elaborate on the limitations relevant to your statistical analysis: Factorial ANOVA, three way classification.<br><br>**We would appreciate clarification from the reviewer on this point.**<br><br>13. It appears that you have shown what has been shown over a number of years that learning outcomes using technology (computer-based instruction, VR) are comparable to learning outcomes using traditional tools (textbook, face to face instruction).<br><br>**A strength of this study is that it sought not only to examine performance outcomes, but also motivation and interest in learning neuroanatomy based on the study materials. VR led to enhanced motivation, decreased neurophobia, and outperformed traditional book learning on every qualitative measure that we polled. We believe that the qualitative benefits of including immersive VR technology in study, in conjunction with the evidence for efficacy of this technology, highlight the potential of integrating VR into the classroom setting. As curriculums continue to shift towards self-taught models of learning, VR modules such as this may become an invaluable resource for motivating students without compromising learning outcomes. We have now highlighted this in the Interpretation (lines 390-392):**<br><br>**"However, the novelty and potential for increased learning motivation is an asset for VR technology in the medical curriculum that may encourage student incentive to study."**<br><br>14. It may also be interesting to consider the level of learner (medical student). When compared with their peers, these students are typically at the top of the academic rankings and method of instruction may have little impact on learning in this population. In other words, despite the teaching method, the students will do well.<br><br>**This is an interesting point and would constitute an important future direction of exploration into the efficacy of VR training for learning outcomes (e.g., through inclusion in undergraduate curriculum). It is possible that more pronounced training differences might arise as a function of teaching method for less academically proficient populations, however our study was focused solely on establishing VR's efficacy with medical students. We have alluded to this future direction in the Interpretation (lines 393-395):**<br><br>**"Our results are currently only generalizable to a particular population and future studies should examine other student populations, including undergraduate students and neurology and neurosurgery trainees."** |
| **Reviewer 3** | **Dr. Bart Harvey** |
| Institution | University of Toronto, Dalla Lana School of Public Health, Toronto, Ont. |
| General comments (author response in bold) | Ekstrand and colleagues have conducted an interesting randomized controlled trial examining the effects of immersive and interactive virtual reality on the learning and retention of neuroanatomy in a group of medical students. While I believe the results of this study provide valuable insights, I do have several comments and suggestions for the authors and CMAJ Open editors to consider prior to the manuscript being accepted for publication.<br><br>1. To begin, I would suggest the presentation of the trial's results be expanded. While the ANOVA and t-test results are presented in Tables 3, 4 and 5, no complementary descriptive statistical results appear to be provided. For example, what was the average score (out of 23) for each of the two study groups on the pre- and post-test questions, and again on the delayed retention administration of these two tests on day 7? Perhaps that information has been implicitly provided in Figure 1, but even if it has, I would suggest this fundamental information be more clearly and explicitly presented by the authors, and accompanied by applicable measures of precision (e.g., 95% confidence intervals for the applicable average scores out of 23). This need is further highlighted by statements on lines 211 and 214 of page 12 where the imprecise phrase "significantly greater accuracy" is used instead of actually reporting the precise differences in average scores between the two study groups. I would also suggest that Figure 1 be re-drawn with the y-axis depicting "% Correct" rather than "% Error" which I believe will make that figure more understandable and intuitive to most readers (certainly this one).<br><br>**We have now included a breakdown of the average scores as an appendix (Appendix D), broken down by test and control questions. We chose to use % Error (now % Correct) as our primary measure due to the unequal number of test (12 questions) and control (9) questions, thus not allowing for a direct comparison between them. We have now further specified the 95 % confidence interval values in the figure captions for Figures 1 and 2, and have changed our primary analysis to focus on % Correct as you have suggested.**<br><br>2. For the "Qualitative results" on page 11, I would suggest that the percent of the paper-based group either strongly agreeing or agreeing be explicitly noted (i.e., it was reportedly 97% for the VR group, but what was it for paper-based?).<br><br>**We have now included an appendix (Appendix C) that contains the full list of qualitative questions as well as the results.** |

3. I’m not sure what the rationale is for including the phrase “and may even surpass” on line 205 of page 11. I don’t think it is justified by the presented results so would suggest it be deleted.

**We have now removed this problematic phrase.**

4. Given the relatively modest differences in results between the two modalities, I wondered if the discussion section wouldn’t be strengthened by the authors discussing the additional costs associated with the “immersive and interactive virtual reality” technology as compared to the traditional paper-based resources. While beyond the scope and intent of this trial, some form of cost-benefit analysis seems warranted to further inform whether the introduction of these technologies would be a justified expense for academic institutions and how much “bang for the buck” they might expect to realize. I also am uncomfortable with the authors using the word “efficient” (on line 226 of page 12) and the phrase “becomes more cost-effective” (on line 257 of page 14) to describe the VR technology given that no cost-benefit/cost-effectiveness analyses were carried out as part of this study.

**We have now removed both of the problematic phrases you identified. We have also addressed the necessity of a cost-benefit analysis prior to implementing immersive and interactive VR into the medical curriculum (lines 400-402):**

**“Finally, future studies should perform a cost-benefit analysis of immersive and interactive VR to further justify its integration into medical curriculum.”**

5. On line 216 of page 12 the authors refer to “an interesting finding” but I am not sure that this isn’t just as reasonably described as a “chance finding” requiring assessment in future study? Similarly, I would also suggest that the comment that the “paper-based group did not generalize as well...” (line 219 on page 12) might be more reasonably framed as a potential explanation that requires further assessment in future studies.

**Due to space constraints, we have now chosen to remove discussion on this point, as we agree with your comment that this finding may or may not be of significance in the current study.**

6. I also have concern about the authors mentioning “subspecialty practice” on line 225 of page 12 as this trial assessed effects no longer than 7 days following the VR learning experience in a group of medical students (i.e., many years before any of them would even enter relevant subspecialty residency training, let alone practice). As such, I would suggest any comments about “maintaining knowledge” should be clarified as being limited to the 7-day post-learning timeframe of the study.

**We have now changed ‘subspeciality’ to ‘medical’ to ameliorate any confusion. The short time frame of our study is discussed as a potential limitation on lines 382-386. We believe this technology has application, but we agree with the reviewer that we should limit our discussion to medical students.**

7. In the discussion section, I did not see any mention of what effect the lack of blinding/masking of study participants might have played in potentially explaining, at least in part, the observed quantitative and qualitative results. Is it possible that those who were allocated to the VR technology arm of the trial enjoyed a different level of motivation and appreciation, at least in part, because of the group they were randomized to (and the technology they experienced)? That is, how much of the observed effects, modest as they were, are attributable to subjective factors?

**Unfortunately based on the nature of the study materials in this experiment, full blinding was not possible. Participants were not aware of which group they were in until the study period itself and thus the pre-test scores should have been unaffected. We have included a discussion pertaining to your point in the Interpretation (lines 386-392):**

**“Further, there is an inherent learning curve when learning a new technology and the relatively short period of training and study time on the VR system may have decreased study time compared to the paper-based group, even with the training module. It is also possible that the VR group enjoyed a different level of motivation for the study materials from exposure to the novel environment that may have contributed to their performance. However, the novelty and potential for increased learning motivation is an asset for VR technology in the medical curriculum that may encourage student incentive to study.”**