Peer Review File

Article information: http://dx.doi.org/10.21037/atm-20-2785

Reviewer A

1. Q: Your search terms are lacking, and probably explain why you only used so few articles. Your simulation-related search term ("VR" or "AR" or "MR") is good, but "medicine education" is not. You leave out the more grammatically correct "medical education," as well as synonyms for medical (such as "surgical") and synonyms for education (such as "learning" or "training"). I do not believe you were able to capture the full scope of the literature with the term "medicine education."

A: We have modified our text as advised (see Page 10, line 9-10).

2. Q: In the statistical analysis section, you write about "The OR and 95% CI," but you should explain those for readers who are less familiar with the statistics required in meta-analyses.

A: We have modified our text as advised (see Page 11, line 19-Page 12, line 3).

3. Q: Also in the Statistical Analysis section, you write about using the random vs fixed-effects model. I am not sure if you have expressed yourself clearly because those models are for the meta-analysis as a whole, not for individual articles. Why are you writing that you used the random-effects model at times, and fixed effects at others? You cannot use different models to analyze studies in the same analysis. I suggest using the random-effects model only.

A: We have modified our text as advised (see Page 12, line 4).

4. Q: Your analytic strategy, in general, is hard to understand and would be impossible to replicate. You should go into more detail here

A: We have modified our text as advised (see Page 11, line 19-Page 12, line 6).

5. Q: You have too many moderator analyses for only 6 studies. How many articles were in the "freshman" group? The "postgraduate" group? How many were "small" or "large" groups? It seems that you can't possibly have many with only 6 articles, and you do not supply your number of studies.

A: The number of articles in each subgroup can be seen in *Table 2* (Page 25).

6. Q: Your discussion section is too repetitive.

A: We have revised it a lot (see Page 15, line 9-Page 17, line 10).

7. Q: You wrote in your abstract "Results: At last, six studies were included in our meta-analysis and the results indicated that there was a significant difference" But the correct phrasing here would be "results indicated that there WAS a significant difference

A: We have modified our text as advised (see Page 5, line 12).

8. Q: In your introduction, you write “Virtual Reality, known as VR, is a stimulation” but VR is a SIMULATION, not stimulation.

A: We have modified our text as advised (see Page 8, line 4).

9. Q: Additionally, the grammar and spelling need a lot of work.

A: We have revised it a lot.

Reviewer B

1. Q: Comments on the abstract: please describe the abstract in detail according to the PRISMA guidelines so that readers can clearly understand the research content. (especially in the conclusion part, please revise again), can the format of "virtual reality (VR)" be used in the keyword part?

A: We have modified our text as advised (see Page 5, line 2-21).

2. Q: Comments on the preface: please add previous studies to improve the necessity of this study and add the specific advantages of virtual reality to medical teaching so that readers can understand.

A: Several additional papers have been added to improve the necessity of this study (see Page 7, line 18-Page 8, line 18).

3. Q： As far as I know, although the three VR, AR, and MR technologies all belong to the same virtual reality technology, there are great differences in the different virtual environments. Please add such differences in the discussion section.

A： This analysis mainly involved VR, so AR and MR were deleted.

4. Q： please clearly indicate the references in the Statistical analysis section

A： We have modified our text as advised (see Page 12, line 4-5).

5. Q： there are deficiencies in the details of the article. For example, the author can delete the extra Spaces in the article

A: We have revised it a lot.

6. Q： the article should have a quality evaluation, but why didn't the author conduct a quality evaluation of the research?

A： We have modified our text as advised (see Page24, *Table 1*).

7. Q： Comments on the results section :(1) please give the reference "4 were conducted in North America countries (America and Canada), 1 in England and 1 in Korea" in this section.

A： We have modified our text as advised (see Page 13, line 6-7).

8. Q: please give the reference of the contents in table 1 appropriately (Maytin et al.).

A： We have modified our text as advised (see Page24, *Table 1*).

9. Q：revise the position of the figure

A：We have revised it a lot.

10. Q：Comments on the analysis and discussion section: please elaborate on the advantages of virtual reality in the discussion section.

A：We have modified our text as advised (see Page 15, line 13-22).

Reviewer C

1. Q: A major flaw is the literature search methodology. It is not clearly defined as it is, it leaves many uncertainties and doubts. It renders the method non-reproducible and therefore deviates the study from the correct scientific methodology.

A: We have modified our text as advised (see Page 10, line 7-16).

2. Q: The authors do not carry out a detailed analysis of the particularities of the only 6 studies considered for this meta-analysis. They leave many doubts to the reader that this should be solved by reviewing the original articles by themselves.

A:We compared the VR group and traditional education group by analyzing the passing rates of the students. However, Owing to different areas and cultures, the formats of the exam in these studies were different. Then the passing rate mentioned in the article was used directly. For articles that did not mention the pass rate, the number of successful people and the total number of people in the trial group that were mentioned in the article can be used to calculate the pass rate. And we add quality evaluation in *Table 1*.

3. Q: The discussion is short and practically dedicated to reproducing the results as presented in the study. Actors do not delve into the meaning of their findings. Possibly because there is no analysis of the different teaching methodologies in the 6 selected papers.

A：We have modified our text as advised (see Page 16, line 17-Page 17, line 10).

4. Q: Additionally, the manuscript needs a revision of the English wording and the expressions used. There is excessive use of the first person (we did, we selected ...) when the usual thing in scientific writing is to use the impersonal form (task was done, it was selected ...)

A: We have revised it a lot.

5. Q: Methods and results must be written in the past tense all along with the chapters.

A: We have revised it a lot.

6. Q: ABSTRACT

I completely disagree with the soundness of the expression of the conclusions in the abstract, when the authors say "it is essential for medical universities to apply virtual reality technology to medical education". Furthermore, it is not a conclusion drawn from this study and does not reflect the conclusions of the body of the manuscript.

A：We have modified our text as advised (see Page 5, line 17-21).

7. Q: BACKGROUND
The authors go directly from talking about a traditional lecture-centric and memorization-based medical education to the arrival of virtual reality as an educational solution. There is not a single word dedicated to the more than 20 years of experience in online teaching and medical simulation in 2D environments, which are undoubtedly part of the background of current educational resources.
A: Some of the benefits and drawbacks of 2D technology have been added (see Page 7, line 18-Page 8, line 3).

8. Q: The terms "VR", "AR" and "MR" were used in the search in the methodology section. The reader should infer that the last two acronyms refer to augmented reality and mixed reality, since it is not indicated in the text and, furthermore, there is no claim about augmented reality and mixed reality in this section and there should be.
A：This analysis mainly involved VR, so AR and MR were deleted.

9. Q: Page3. Line 12. The term "skyrockets" is a literary license that should be avoided.
A：We have modified our text as advised (see Page 7, line 15).

10. Q: Page3. Line 15. Perhaps the authors mean simulation instead of stimulation. Typing error?
A：We have modified our text as advised (see Page 8, line 4).

11. Q: Page 3. Last paragraph. The last two sentences before “For example…” need respective references.
A：We have modified our text as advised (see Page 8, line 11).

12. Q: Page 4. Line 3. This phrase about a better understanding of the nerves of the skull thanks to virtual reality (reference 4) is an example and should be so indicated. Better still would be to provide several examples of improved anatomical framework knowledge due to VR rather than just one
A: More examples have added (see Page 8, line 13-16).

13. Q: Page 4. Line 4. In a similar way, it should be said: ”For example, a study showed after receiving…”
A：We have modified our text as advised (see Page 8, line 20).

14. Q: Page 4. Line 9. References must be included to clarify which researches have been done to prove the validity of VR and, in which researchers still have concerns about its effectiveness (what studies support these statements rather than a statement of reference 6.
A：We have modified our text as advised (see Page 9, line 5-15).

15: Q: Page 4. Line 14. The expression that there is an urgent need for meta-analysis is a literary

license to be removed. It is a somewhat excessive self-assessment by the authors themselves.
A：We have modified our text as advised (see Page 9, line 20).

16. Q: The objectives should be specified more clearly. It is certainly unclear. For example, the authors say they have performed a meta-analysis of existing data from retrospective cohort studies but the article by Jung et al (2012), for example, is a randomized control trial. The objectives and methodology as reflected in the study are certainly not easy to understand.
A：We are not being rigorous enough and have modified our text as advised (see Page 9, line 22).

17. Q: The search terms were ("VR" or "AR" or "MR") and "Medicine Education".
This systematic search raises many doubts about its efficiency and its reproducibility.
The authors are supposed to have used the OR and AND Booleans as expressed, and to have used quotation marks to search for a character string, for example, "Education Medicine". The policy of using quotation marks in searches is different in PubMed and the other three databases. The authors should indicate in more detail what exactly the search methodology they have used in each database. If the search strategy is exactly as they say, it leaves important gaps in the study. For example, it would exclude any article that talks about "virtual reality" "augmented reality" or "mixed reality" without using the corresponding acronyms.

Another example. The use of the term (or chain characters) "Medicine Education" instead "Medical Education" lead to many potentially missed papers. A review in PubMed searching (only in the title) "Medicine Education" gave 374 results and with "Medical Education" gave 18,131.

Similar inconsistencies using the terms virtual reality augmented reality and mixed reality are found with fast reviews.
A: We have modified our text as advised (see Page 10, line 7-10). And we mainly used the site's advanced search function when searching literature, so search terms primarily presented the train of thought of searching literature to the readers.

18.Q: Page 4 last line. It must be said carried out (past tense).
A: We have modified our text as advised (see Page 10, line 5)

19. Q: For a better understanding of the work done, the authors should describe more clearly what data from the populations included in the studies found are compared and with which methods each one, instead of exclusively describing the statistical tools used. What exactly have they compared? What do they mean by efficacy analysis for case-control study?
A：We have modified our text as advised (see Page 11, line 19-Page 12, line 13).

20. Q: This section deserves a little more detail, beyond referring to Figure 1 (which is very clarifying). Finally, among the four databases they have found the sum of 16,623 papers. How many of them were redundant, not specified?
A: The selection process of eligible studies can be seen in *Figure 1* (Page 27)

21. Q: The 6 articles found to deserve a better description of them, explaining how the data has

been taken exactly to calculate the passing rates. Table 1 is somewhat confusing. and it is not finally understood what were the rates of the control (traditional education) and experimental (virtual reality) groups. How the passing rates were calculated were a base 10 rating, they were in percentage, how they were unified to compare them …

A： We have modified our text as advised (see Page 11, line 19-Page 12, line 2).

22. Q: The number of subjects in each study is unclear. The authors say a total of 633 (I suppose subjects instead of objects -page 7 line 3). For example, I have read the reference by Jung et al (2012) to find out that the subjects were first-year nursing students in a venipuncture practice. The study comprised 114 subjects distributed in three groups. A: with the use of mannequins (traditional education-TE). B with the use of virtual simulation (VR). C with the use of both VR and TE. How the mixed group (C) was considered in this study, for comparison. Authors should not leave so important uncertainties to the reader.

A: According to data extraction criterion, the number of the students enrolled and the passing rates of students from virtual reality and Traditional Courses were extracted. And the mixed group is outside our scope of study. We considered that comparing the mixed group with another group can't tell the difference between the VR and traditional education.

23. Q: Continuing with the stated previously, it is unclear whether the passing rates were homogeneous in the six studies, nor how they have been affected or homogenized, and how they were considered in mixed (TE-VR) formation groups. It is not enough to say that the 6 studies compare traditional education and virtual reality. The reader has to know what the educational context was in each study and what evaluation methodology was used. That is, you must know exactly what you are comparing, before proceeding to analyze differences by country, profession, or number of subjects.

A: We assess the teaching effect of the virtual reality technology by comparing the exam passing rate.

24. Q: The discussion has two paragraphs devoted to justifying the study (should be placed in the introduction-background).

A: We have modified our text as advised (see Page 15, line 9-Page 16, line 10).

25. Q: The following three paragraphs repeat what was found in the results indicating that there are significant differences between groups, sometimes without saying in what sense. It does not provide an in-depth analysis of the data found, perhaps because there is a lack of data to include in results (see previous comments).

A: We have modified our text as advised (see Page 16, line 17-Page 17, line 10).

26. Q: The first sentence of the conclusion is paradigmatic in relation to what was said previously.

A: We have modified the conclusion as advised (see Page 18, line 4-10).

27. Q: “The results in this meta-analysis showed that there is statistically significant difference between VR and traditional education in terms of the passing rates.” In what sense? What does

it means?
A: It means that in terms of the exam passing rates, there is statistically significant difference between VR and traditional education. In a sense, virtual reality technology might help students master medical knowledge better.

28. Q: I think 16 references are too few for an article of this type, on such a hot topic and of such relevance in the possible changes in current medical education.
A: We have added some new papers to the reference list (see Page 19, line 8-Page 22, line 16).

29. Q: Since Forest plots have been used as main (almost exclusive) data presentation elements, they should adequately fill in the meaning in the text and in the figure captions to facilitate a better understanding and interpretation of the data.
A: We have modified our text as advised (see Page 28-31).