

PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form (<http://bmjopen.bmj.com/site/about/resources/checklist.pdf>) and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below.

ARTICLE DETAILS

| | |
|----------------------------|---|
| TITLE (PROVISIONAL) | Effects of high-fidelity simulation based on life-threatening clinical condition scenarios on learning outcomes of undergraduate and postgraduate nursing students: a systematic review and meta-analysis |
| AUTHORS | La Cerra, Carmen; Dante, Angelo; Caponnetto, Valeria; Franconi, Ilaria; Gaxhja, Elona; Petrucci, Cristina; Alfes, Celeste; Lancia, Loreto; |

VERSION 1 – REVIEW

| | |
|------------------------|--|
| REVIEWER | Dr Robyn Cant Nursing and Midwifery, Monash University, Australia and Federation University Australia |
| REVIEW RETURNED | 18-Jul-2018 |

| | |
|-------------------------|---|
| GENERAL COMMENTS | <p>Thank you for this volume of scholarly work. It is well constructed and informative; it present new information; however I have reservations about your presentation and assertions regarding the study scope.</p> <p>The title and abstract: There is discord between the title and the content which raises issues of lack of clarity. Perhaps these are just anomalies in your use of English language? However, I suggest you overstate the boundaries of your title and aim as the study clearly comprises research in broad areas of nursing education and is not limited to <u>critical care scenarios</u>. The title and aim should be revised to accurately state the content. Did you mean to say 'clinical care? Rather than critical care? The boundaries need to be accurate. As you also include professional nurse studies, it is a study not only of nursing students? Please revise. In addition, I would just like to draw your attention to a further few issues that I outline below. These relate to use of English language in the main.</p> <p>Page 2 line 33 abstract- please add in a qualifier that there was significant heterogeneity among studies. Revise/reduce anecdotal language in Strengths and Limitations</p> <p>Page 3 line 42-50 introduction: please revise language to omit your claims of no prior 'wide overview' in critical care research (which is erroneous). Refocus- your study includes generalist topics in SBE, caring for children with bronchiolitis; Simulator effects on cognitive skills and confidence levels; maternal- newborn nursing; undergraduate nursing health assessment course. – and certainly,</p> |
|-------------------------|---|

| | |
|--|--|
| | <p>critical care topics as well? Your study is one of a number of meta-analyses of SBE in nursing in recent years.</p> <p>P3 Line 52-55 Suggest revise the objective to suit above.</p> <p>P3 line 3-4 : incorrect and unclear impact- please revise statement</p> <p>P3 line 25- 30 you might mention simulation standards? Eg, INACSL Standards of Best Practice</p> <p>https://www.nursingsimulation.org/article/S1876-1399(17)30294-3/abstract</p> <p>P4 line 13: I suppose you can say that any trained nurse who is undertaking simulation education is a student however we would not regard professional nurses as 'students'? they have a differing basic knowledge.</p> <p>P4 line 28: "performance is referred to the demonstration of clinical skills"- revise grammar for clarity</p> <p>P4 line 45: 'The consistency of raters' judgments" -unclear what this refers to in study selection- is it the quality assessment? Or is it a yes/no inclusion criteria And what did it need to be to be approved? please advise. Was not really needed?</p> <p>P5 line 3: the journal title is redundant and is not required in the included data/the table?</p> <p>P5 line 12: were any studies excluded?</p> <p>P5 line 31-46 your explanation reads well</p> <p>P5 line 52: increase in studies over time is well known and the figure 1 should only be presented in the supplementary file?</p> <p>P6 line 3: inter-rater reliability- please explain the scale? Is it on article selection?</p> <p>P6 line 5, line 9: there is a need to label each supplementary file – eg, Table A - as this is unclear</p> <p>P6 line 15-20: suggest decimals to one decimal place for better clarity?- M age = 25.72 is not necessary. Also in line 27-35. How can you justify reporting "low-fidelity manikin (11.00%)?" Think critically.</p> <p>Line 37-41: there is a requirement to make a distinction between self-reported data and objective measurements in sim studies (validity and reliability). Some journals do not now accept self-reported data designs. Can you explain further in this section? And add detail to your later commentary?? (please round the decimals!)</p> <p>Page 6 line 652: the table in supp file make no sense to me as the</p> |
|--|--|

| | |
|----------|--|
| | <p>headers are blacked out and we do not know what they report?</p> <p>Figure 3: very clear</p> <p>Page 7 line 8-9 What was the cut-off for these to describe effect or no effect? I'm not clear perhaps you should shade in the variables charts that were improved?</p> <p>P7 Line 10: yes "significant heterogeneity:- name the range? And say what is acceptable/expected?</p> <p>P7 Line 52- page 8 line 40: suggest revise this section too many ideas all jumbled together, not well justified. Present each issue you advise and then cite how it links to the current literature?</p> <p>P8 line 16: remove the first sentence it is erroneous. Actually, all the numeric citations should have a comma between each? Suggest Add: Distinction between objective and subjective evaluation measures and this recent literature.</p> <p>P8 Line 24 -34: add citations to justify what you assert?</p> <p>P8 Line 40-42 The included study citations are not needed here- remove? As they were presented earlier in Figure 3. The sentence: "Probably ..." suggest alter to read 'may be'- more tentative is correct. What about objective measures here??/& sample size? Sample sizes of 10 to 20 were included? The page would be better/clearer divided into paragraphs?</p> <p>P 9 Line 12-24 many studies have been conducted in these fields and also there are recommendations for such, so suggest be more aware that your suggestions concur with current research – cite some.</p> <p>P9 Line 30-32: revise grammar, include the word 'experimental'?</p> <p>P9 Line 34-35: disagree that further self-efficacy studies are necessary we have moved beyond that, and this is not a limitation- remove - journals do not want subjective studies ...objective and experimental studies are needed see: <i>Clinical Simulation in Nursing</i> recently.</p> <p>P9: Line 46-49: remove sentence "Students trained ..." as this data is not seen in the results. The final sentence is clear and correctly states the field.</p> <p>Abstract: revise in line with changes to conclusion</p> <p>Enable headers to supplementary table to be seen, explain each variable in the table as a footnote for the reader to understand the table.</p> <p>Well done.</p> |
| REVIEWER | Bryan Boling |

| | |
|-------------------------|---|
| | University of Kentucky United States of America |
| REVIEW RETURNED | 12-Sep-2018 |
| GENERAL COMMENTS | This is a well-written piece and a very relevant topic. I found it especially nice that the authors had the linguistic expertise to include papers written in English, Spanish, French, and Italian. It not often that I see such linguistic diversity. |
| REVIEWER | Sang Suk Kim Chung-Ang University, Republic of Korea |
| | Education, Simulation |
| REVIEW RETURNED | 26-Sep-2018 |
| GENERAL COMMENTS | <p>The study was well delivered and the outcomes are clearly explained and explored in the discussion.</p> <p>Use the moderator terms in subgroup consistently in the table and in the description section. (ex. p5 line 40 : the 'scenario', 'manikin brand', 'control intervention', and 'randomization' as moderators. p7 line12: subgroup analyses : scenario topic, type of manikin, control group treatment, and method of selecting groups and Table 1)</p> <p>Please Update the references in the introduction and discussion section.</p> <p>p5. line 55 Clarify the duration of sample studies selected for SR analysis without confusion with the HFPS studies for the past 30 years described in the previous sentence.</p> <p>p8, line 21 Rewrite the sentence "When compared with us testing methods." Or, you can delete the sentence. It is also shown in Table 1 as being effective in comparison with the no interaction group.</p> |
| REVIEWER | John Stephenson University of Huddersfield United Kingdom |
| REVIEW RETURNED | 01-Nov-2018 |
| GENERAL COMMENTS | <p>This review is limited to data analysis issues. This is a strong study: I have only a few comments, relating to presentation.</p> <ol style="list-style-type: none"> 1. There is no legend on the axes on the forest plots and the figures on the axis are too small in relation to other text on the plots. 2. On "Performance", one study (Alnier) is out of order. 3. A Z-test should be conducted for each meta analysis and the result presented on each forest plot. 4. Resolution in Figure 1 is poor and R squared statistic is quoted with a comma as decimal separator instead of a dot, and with too many decimal places. Is this figure really necessary? I don't think it adds much. 4. "Publication bias was examined by the funnel plot" (page 5) - where are the funnel plots? 5. Study characteristics (page 6) quote frequencies instead of or as well as percentages. |

| | |
|------------------------|---|
| REVIEWER | Yu-Lun Liu University of Texas Southwestern Department of Clinical Sciences, USA |
| REVIEW RETURNED | 14-Nov-2018 |

| | |
|-------------------------|--|
| GENERAL COMMENTS | <p>To the authors:</p> <p>In this manuscript, Lancia and his colleagues presented a systematic review and meta-analysis, which examined the effectiveness of HFPS critical care scenarios improving the learning outcomes, compared with the other teaching methods. Overall, this is a very interesting manuscript with balanced discussion, but I have the following suggestions for the study's data analysis, results and conclusions.</p> <ul style="list-style-type: none"> - Methods: Regarding outcome measures, different types of outcome measurement tools, including Likert-type scales, multiple-choice questions, dichotomous scales, checklists, open questions, and others were considered in this systematic review. Taking "self-rated knowledge" outcome as an example, one study adopted the 20-item dichotomous tool (with IG=0.86), but another study adopted the ACLS Written Examination tool (with IG=90). The same learning outcome seemed to be measured from multiple measurement instruments. Such outcome variable selection may lead to bias in meta-analysis, as well as increase the impact of heterogeneity among studies. Please provide the details and evidence for choosing/combining your specific learning outcome measures. - Results: I would recommend conducting the sensitivity analysis (e.g., leave-one-out analysis) to evaluate whether any study removed from the included studies would alter the findings for each learning outcome, i.e., to assess whether the conclusion could be driven by any single study. - Results/Limitations: Regarding the risk of bias, the authors should provide the results of funnel plot, p-values for Egger's test, etc. in the main text or the supplementary. When the publication bias was detected for "self-efficacy" outcome, I would recommend comparing the results with/without adjustment for small-study effects. Also, I would recommend discussing the possible reasons for the presence of publication bias in the Limitations section, e.g., it is possible that negative studies for this "self-efficacy" outcome were less likely to be published or reported. - Limitations: The authors need to discuss the type of study designs for this meta-analysis. Among the included 33 studies, some studies actually were non-RCT data. Such non-RCTs or observational studies examining these learning outcomes would be limited by potentially unmeasurable confounding and selection bias. This could be one limitation of this study. |
|-------------------------|--|

| | |
|------------------------|--|
| REVIEWER | Chakra Budhathoki Johns Hopkins School of Nursing USA |
| REVIEW RETURNED | 17-Nov-2018 |

| | |
|-------------------------|---|
| GENERAL COMMENTS | <p>Overall this is a very interesting manuscript, and it is a nicely done systematic review and meta-analysis!</p> <p>I suggest that percentage reporting in the text and tables be limited</p> |
|-------------------------|---|

| | |
|--|---|
| | <p>to one decimal place.</p> <p>Would it be better to say larger effect sizes instead of higher effect sizes in abstract?</p> <p>Is any other teaching method a comparison group? How similar or different were they?</p> <p>Page 5 2nd line: Would it be better to say control conditions rather than control interventions?</p> |
|--|---|

VERSION 1 – AUTHOR RESPONSE

REVIEWER 1 - Robyn Cant

| Reviewer Comment | Response to Comment |
|--|--|
| <p>The title and abstract: There is discord between the title and the content which raises issues of lack of clarity. Perhaps these are just anomalies in your use of English language? However, I suggest you overstate the boundaries of your title and aim as the study clearly comprises research in broad areas of nursing education and is not limited to <u>critical care scenarios</u>. The title and aim should be revised to accurately state the content. Did you mean to say 'clinical care? Rather than critical care? The boundaries need to be accurate. As you also include professional nurse studies, it is a study not only of nursing students? Please revise.</p> | <p>In accordance with your suggestions, we revised title and abstract in order to make them consistent with the content of the review. Our aim was, in fact, to evaluate the effectiveness of the HFPS on the learning outcomes of nursing students (undergraduate and postgraduate) when the scenarios were based on life-threatening clinical conditions, that are not faced only in critical care settings. Thank you for your precious suggestion.</p> <p>We appreciate your advice. Regarding to the issue about the type of students, we are aware that postgraduate students have a different basic knowledge when compared to undergraduate. Since the basic knowledge of postgraduate students is hypothetically higher than undergraduate students, this could potentially affect the effect size of the considered outcomes. Furthermore, it was not possible to detect whether postgraduate students were certainly employed in clinical practice before or during the conduction of the studies. For these reasons, we always referred to the sample as 'nursing students'. In this regard, we added a specific section in Limitations.</p> |
| <p>Page 2 line 33 abstract- please add in a qualifier that there was significant heterogeneity among studies.</p> | <p>In the Abstract we added a qualifier about the presence of a significant heterogeneity among</p> |

| | |
|--|--|
| | studies. |
| Revise/reduce anecdotal language in Strengths and Limitations | We reduced anecdotal language in Strengths and Limitations. |
| Page 3 line 42-50 introduction: please revise language to omit your claims of no prior 'wide overview' in critical care research (which is erroneous). Refocus- your study includes generalist topics in SBE, caring for children with bronchiolitis; Simulator effects on cognitive skills and confidence levels; maternal- newborn nursing; undergraduate nursing health assessment course. – and certainly, critical care topics as well? Your study is one of a number of meta-analyses of SBE in nursing in recent years. | We revised the sentence containing 'wide overview' and clarified that our review focused on the effectiveness of HFPS based on life-threatening clinical scenarios referred to different clinical settings. |
| P3 Line 52-55 Suggest revise the objective to suit above | Done. Thanks. |
| P3 line 3-4: incorrect and unclear impact- please revise statement | In Strengths and limitations, the statement 'Data heterogeneity and the limited availability of high-level evidence limits the generalizability of results in current nursing education practice' was revised to express the concept more clearly. Data heterogeneity and limited amount of high-quality research had been referred to primary studies. |
| P3 line 25- 30 you might mention simulation standards? Eg, INACSL Standards of Best Practice https://www.nursingsimulation.org/article/S1876-1399(17)30294-3/abstract | In the Introduction INACSL simulation standards were mentioned. |
| P4 line 13: I suppose you can say that any trained nurse who is undertaking simulation education is a student however we would not regard professional nurses as 'students'? they have a differing basic knowledge. | We appreciate your comment and we are aware that post-graduate students could have a different basic knowledge when compared to undergraduate students. However, our intention was to consider all the academic students, independently from their educational and professional background which was not possible to ascertain. For this reason, all participants are intended to be 'nursing students'. However, considering your suggestion, we referred to such threat in Limitations section. Thank you. |

| | |
|--|---|
| P4 line 28: "performance is referred to the demonstration of clinical skills"- revise grammar for clarity | In Methods, we revised the definition of performance as the student's ability to demonstrate clinical skills. |
| P4 line 45: 'The consistency of raters' judgments" -unclear what this refers to in study selection- is it the quality assessment? Or is it a yes/no inclusion criteria. And what did it need to be to be approved? please advise. Was not really needed? | In Methods, 'The consistency of raters' judgments' was referred to the agreement among the authors about the eligibility and inclusion of titles and abstracts and full-texts, respectively. We added the range of the values of Krippendorff's alpha coefficient, which is an index of agreement among the raters. |
| P5 line 3: the journal title is redundant and is not required in the included data/the table? | According to your suggestion, data referred to the journal titles of the included studies were deleted. |
| P5 line 12: were any studies excluded? | Thanks to your suggestion, we reported in the paper that the quality of the studies was not deemed to be an exclusion criterion. |
| P5 line 52: increase in studies over time is well known and the figure 1 should only be presented in the supplementary file? | As you suggested, Figure 1 has been moved in the supplementary file. |
| P6 line 3: inter-rater reliability- please explain the scale? Is it on article selection? | As explained in Methods in accordance with your advice, the "inter-rater reliability" was referred to the agreement among the authors about the eligibility and inclusion of titles and abstracts and full-texts, respectively. In this regard, we also added the range of the values of Krippendorff's alpha coefficient in Methods. |
| P6 line 5, line 9: there is a need to label each supplementary file – eg, Table A - as this is unclear | We added the label for each Box, Figure, and Table in the Supplementary file. Also, we inserted labels for each visual in the manuscript, adjusting the order of appearance. |
| P6 line 15-20: suggest decimals to one decimal place for better clarity? M age = 25.72 is not necessary. Also in line 27-35. How can you justify reporting "low-fidelity manikin (11.00%)?" Think critically. | We appreciated your suggestion. We rounded decimals to one, when needed. |

| | |
|---|--|
| Line 37-41: there is a requirement to make a distinction between self-reported data and objective measurements in sim studies (validity and reliability). Some journals do not now accept self-reported data designs. Can you explain further in this section? And add detail to your later commentary?? (please round the decimals!) | In this regard, we clarified the nature and type of the instruments utilized for the outcomes' measurement. We also modified the labels in Table C of the supplementary file in order to make them consistent with the changes performed in the manuscript. Thank you for your precious comment. |
| Page 6 line 652: the table in supp file make no sense to me as the headers are blacked out and we do not know what they report? | We removed the black background from the header in the Table C of the supplementary file. It was a mistake. Thank you. |
| Page 7 line 8-9 What was the cut-off for these to describe effect or no effect? I'm not clear perhaps you should shade in the variables charts that were improved? | In the Method section, we clarified the condition of the significant effect of HFPS. We also specified the interpretation modality for the effect size. Thank you. |
| P7 Line 10: yes "significant heterogeneity: name the range? And say what is acceptable/expected? | We added the range of the heterogeneity. Furthermore, we recognized as acceptable any level of heterogeneity according to Higgins (2008) as stated in Methods. |
| P7 Line 52- page 8 line 40: suggest revise this section too many ideas all jumbled together, not well justified. Present each issue you advise and then cite how it links to the current literature? | Thank for your suggestions. We revised Discussions in order to clarify and justify the expressed concepts. |
| P8 line 16: remove the first sentence it is erroneous. Actually, all the numeric citations should have a comma between each? Suggest Add: Distinction between objective and subjective evaluation measures and this recent literature. | In this regard, we modified the first sentence to make it closer to the current literature. Moreover, according to the editorial guidelines, no commas were added between the numeric citations. Finally, we provided some integrations in order to clarify the distinction of subjective and objective outcomes. Thank you. |
| P8 Line 24 -34: add citations to justify what you assert? | We appreciated your suggestion, but these critical considerations were written to support readers in interpreting the results and their implications in clinical and educational practice. Moreover, given the lack of studies about the effectiveness of the HFPS in |

| | |
|--|---|
| | modifying patients' outcomes especially for life-threatening clinical conditions, no citations have been possible to add. |
| P8 Line 40-42 The included study citations are not needed here-remove? As they were presented earlier in Figure 3. The sentence: "Probably ..." suggest alter to read 'may be'- more tentative is correct. What about objective measures here??/& sample size? Sample sizes of 10 to 20 were included? The page would be better/clearer divided into paragraphs? | In this regard, the citation style of BMJ Open requires numbers instead of first author and year. Therefore, it was not possible to utilize the plots to cite the included studies neither to remove the citation of the included studies. Instead, your useful suggestion gave us the opportunity to improve discussions related to not significant results about some learning outcomes. Thank you. Finally, we solved the issue related to the distinction between the type of outcome measurements (objective vs. subjective), thank you. |
| P 9 Line 12-24 many studies have been conducted in these fields and also there are recommendations for such, so suggest be more aware that your suggestions concur with current research – cite some. | Thanks for your advice. In this regard, we added some citations and specifications to the expressed concepts in order to clarify that no concurrence exists between our statements and current research. |
| P9 Line 30-32: revise grammar, include the word 'experimental'? | We revised the sentence and added the word 'experimental'. Thanks. |
| P9 Line 34-35: disagree that further self-efficacy studies are necessary we have moved beyond that, and this is not a limitation- remove - journals do not want subjective studies ...objective and experimental studies are needed see: Clinical Simulation in Nursing recently. | In this regard, we revised the sentence in order to better describe the implications derived from the publication bias detected for self-efficacy. |
| P9: Line 46-49: remove sentence "Students trained ..." as this data is not seen in the results. | We removed the sentence according to your suggestion. Thank you. |
| Abstract: revise in line with changes to conclusion | Abstract is in line with the conclusion contents. Thank you. |
| Enable headers to supplementary table to be seen, explain each variable in the table as a footnote for the reader to understand the table. | We removed the black background from the header in the Table C of the supplementary file and we added the footnotes for some variables. |

REVIEWER 3 - Sang Suk Kim

| Reviewer Comment | Response to Comment |
|--|--|
| Use the moderator terms in subgroup consistently in the table and in the description section. (ex. p5 line 40: the 'scenario', 'manikin brand', 'control intervention', and 'randomization' as moderators. p7 line12: subgroup analyses: scenario topic, type of manikin, control group treatment, and method of selecting groups and Table 1) | We revised moderator terms consistently in the table and description section, thank you for your advice. |
| Please Update the references in the introduction and discussion section. | In accordance with your suggestion, we added some citations in the Introduction and Discussion sections. |
| p5. line 55 Clarify the duration of sample studies selected for SR analysis without confusion with the HFPS studies for the past 30 years described in the previous sentence. | We moved the sentence related to HFPS studies over the last 30 years to the end of the section, not only to make the passage more readable, but also to give an additional information that is independent from the temporal inclusion criteria. It was a mistake. |
| p8, line 21 Rewrite the sentence "When compared with us testing methods." Or, you can delete the sentence. It is also shown in Table 1 as being effective in comparison with the no interaction group. | In accordance with your suggestion we rewrote this sentence and integrated it with the observations of another reviewer. Thank you. |

REVIEWER 4 - John Stephenson

| Reviewer Comment | Response to Comment |
|--|---|
| There is no legend on the axes on the forest plots and the figures on the axis are too small in relation to other text on the plots. | The forest plots of meta-analyses were performed once again through ProMeta 3.0 software and the legends were added on the axes, as suggested. Moreover, a Z-test was conducted and presented for each meta-analysis in the text. |
| On "Performance", one study (Alinier) is out of order. | |
| A Z-test should be conducted for each meta-analysis and the result presented on each forest plot. | Now, R squared is quoted with a dot, and decimals are rounded to two places. We consider this figure to have just only an informative purpose, so that we moved it in the supplementary file. We appreciated your advice. |
| Resolution in Figure 1 is poor, and R squared statistic is quoted with a comma as decimal separator instead of a dot, and with too many decimal places. Is this figure really necessary? I don't think it adds much. | |
| "Publication bias was examined by the funnel plot" (page 5) - where are the funnel plots? | We added the funnel plot for self-efficacy in the supplementary file (Figure 2). Thank you. |
| Study characteristics (page 6) quote frequencies instead of or as well as percentages. | Thanks to your suggestions and to other reviews, we revised the text using both frequencies and percentages. |

REVIEWER 5 - Yu-Lun Liu

| Reviewer Comment | Response to Comment |
|--|---|
| Methods: Regarding outcome measures, different types of outcome measurement tools, including Likert-type scales, multiple-choice questions, dichotomous scales, checklists, open questions, and others were considered in this systematic review. Taking "self-rated knowledge" outcome as an example, one study adopted the 20-item dichotomous tool (with IG=0.86), but another study adopted the ACLS Written Examination tool (with IG=90). The same learning outcome seemed to be measured from multiple measurement instruments. Such outcome variable selection may lead to bias in meta-analysis, as well as increase the impact of heterogeneity among studies. Please provide the details and evidence for choosing/combining your specific learning outcome measures. | Thanks for your comment. We stated the reason by which we adopted the Cohen's d effect size and sustained it by evidence. |

| | |
|--|--|
| Results: I would recommend conducting the sensitivity analysis (e.g., leave-one-out analysis) to evaluate whether any study removed from the included studies would alter the findings for each learning outcome, i.e., to assess whether the conclusion could be driven by any single study. | We conducted sensitivity analysis with the 'leave-one-out' approach, as you suggested. Thank for this precious comment. |
| Results/Limitations: Regarding the risk of bias, the authors should provide the results of funnel plot, p-values for Egger's test, etc. in the main text or the supplementary. When the publication bias was detected for "self-efficacy" outcome, I would recommend comparing the results with/without adjustment for small-study effects. Also, I would recommend discussing the possible reasons for the presence of publication bias in the Limitations section, e.g., it is possible that negative studies for this "self-efficacy" outcome were less likely to be published or reported. | Thank you for your suggestion. In Results, we added the funnel plot as Figure 2 of the supplementary file, values of Egger test (intercept and p-value), Trim and Fill (d, 95% CI), and Fail-safe number. In addition, we discussed the possible reasons for the presence of publication bias in the Limitations section. |
| Limitations: The authors need to discuss the type of study designs for this meta-analysis. Among the included 33 studies, some studies actually were non-RCT data. Such non-RCTs or observational studies examining these learning outcomes would be limited by potentially unmeasurable confounding and selection bias. This could be one limitation of this study. | We discussed the limitations regarding the design of the included studies and their possible implications for meta-analyses. Thank you. |

REVIEWER 6 - Chakra Budhathoki

| Reviewer Comment | Response to Comment |
|--|---|
| I suggest that percentage reporting in the text and tables be limited to one decimal place. | Thank you for your suggestion. We limited the decimal place to one in the text and tables only for variables other than outcomes, since the meta-analytic software provides data and output figures with two decimals. In this regard, each decimal rounding affects the real effect size. |
| Would it be better to say larger effect sizes instead of higher effect sizes in abstract? | We changed the word 'higher' with 'larger' both in the abstract and text. |
| Is any other teaching method a comparison group? How similar or different were they? | Since we performed all the meta-analyses comparing HFPS with other teaching methods, the answer is yes; therefore, as highlighted in Table 1, control groups varied along with the outcome investigated (e.g. lecture and standardized patient were present as a control group in all the outcomes, as opposed to the other control interventions). However, as highlighted by the subgroup analysis, control group characteristics seemed to have an influence on effect size only for self-efficacy, as reported in the text. |
| Page 5 2nd line: Would it be better to say control conditions rather than control interventions? | As suggested, we changed 'interventions' into 'conditions'. |

VERSION 2 – REVIEW

| | |
|-----------------|---|
| REVIEWER | Sang Suk Kim Chung-Ang University, Republic of Korea |
| REVIEW RETURNED | 26-Dec-2018 |

| | |
|------------------|--|
| GENERAL COMMENTS | It was modified according to the comments of the review. Thank you for your efforts. |
|------------------|--|

| | |
|-----------------|--|
| REVIEWER | John Stephenson University of Huddersfield UK |
| REVIEW RETURNED | 19-Dec-2018 |

| | |
|-------------------------|---|
| GENERAL COMMENTS | All my comments have been addressed. |
| REVIEWER | Yulun Liu University of Texas Southwestern Medical Center |
| REVIEW RETURNED | 26-Dec-2018 |
| GENERAL COMMENTS | The authors have addressed my comments well. I do not have more comments to add. |
| REVIEWER | Chakra Budhathoki Johns Hopkins School of Nursing USA |
| REVIEW RETURNED | 24-Dec-2018 |
| GENERAL COMMENTS | All of my previous comments have been addressed, and the manuscript appears to be much improved. Thank you! |