# PEER REVIEW HISTORY

## ARTICLE DETAILS

| | |
|---|---|
| **TITLE (PROVISIONAL)** | Efficacy of adaptive e-learning for health professionals and students: a systematic review and meta-analysis |
| **AUTHORS** | Fontaine, Guillaume; Cossette, Sylvie; Maheu-Cadotte, Marc-André; Mailhot, Tanya; Deschênes, Marie-France; Mathieu-Dupuis, Gabrielle; Côté, José; Gagnon, Marie-Pierre; Dubé, Veronique |

## VERSION 1 - REVIEW

| | |
|---|---|
| **REVIEWER** | Kimberly Stowers<br>Assistant Professor<br>University of Alabama<br>United States |
| **REVIEW RETURNED** | 06-Aug-2018 |

| | |
|---|---|
| **GENERAL COMMENTS** | Overall major revision, also relating to line 175 and line 499:<br><br>I don't believe this paper provides an adequate explanation for constraining the article search to the years 2005 onward. After reviewing the source being cited (Kulik & Fletcher, 2015), there is no evidence that bias of earlier studies is a key issue in their findings. Overall, the relationship between publication year and effect size wasn't significant. Furthermore, in looking at the mean effect sizes for the different year categories, there wasn't much difference between 2001 - 2005 and 2005 onward. The larger difference appeared to occur in studies up to the year 2000. If anything, Kulik & Fletcher suggest that the relevant issue to consider would be implementation adequacy, something that doesn't appear to correlate with year of implementation, but rather the number of times any given person has implemented an ITS (i.e., prior experience).<br><br>Unless you can provide additional support for why you should constrain to a specific year, please open your search to include all years.<br><br>Minor revisions:<br><br>Line 59: Typo--"more beneficial more learning" should say "more beneficial for learning"<br><br>Lines 161-162: Regarding your definitions of each type of adaptation, given the age of the paper you are citing (Knutov & |

colleagues), did you account for finding adaptation processes that would fall outside the scope of the definitions proposed by Knutov et al? Did you find any additional systems falling outside these definitions that might be relevant? Please clarify in text.

Line 168: Please provide operational definitions of your outcomes of interest so it is clear what you were trying to capture.

Line 338 (rather, the section under this heading): Please add a statement clarifying that some studies used multiple adaptation techniques (with a reference to table 2).

Line 362: For the study that only implemented adaptation at the beginning, what was the adaptation based on? Survey responses? Please include a clarifying statement.

Line 406: Word correction-- "same knowledge scores than...." should say "same knowledge scores as..."

| REVIEWER | Fahad Alam<br>Sunnybrook Health Sciences Centre, Canada |
|---|---|
| REVIEW RETURNED | 23-Oct-2018 |

| GENERAL COMMENTS | Overall, the authors have taken a very detailed systematic approach to this systematic review and meta analysis. Before publication I would like further clarification on how they defined knowledge and competence as outcomes of their meta-analysis. These terms have been used in a variety of different contexts in medical education and have also been used interchangeably (incorrectly in my opinion). Thus it would be prudent for the authors to outline how they defined knowledge and competence as it applied to their search and analysis. I'm sure this was also quite heterogenous in the studies that were included which might make it difficult to make conclusions from the analysis. This would be my major concern. Educators especially in this competence by design (CBD) era in medicine are trying to precisely predict success in training by looking at both knowledge and competence with one not meaning attaining the other.

One good reference is Epstein RM, Hundert EM. Defining and Assessing Professional Competence. JAMA. 2002;287(2):226–235. doi:10.1001/jama.287.2.226

After discussing knowledge/competence and carrying this through their analysis (which could change their conclusions), I would then ask the authors to include a section on further recommendations as to how an educator should structure their own AEE tool if they were creating one. This has been done in pieces in the paper but as a reader, this is what will have an impact on me and lead to knowledge translation. As of right now, it is reading very much like a descriptive paper only.

In the background, the authors make the claim that "…they do not consider users' characteristics to provide a personalized training. They are generally considered to be as effective as non–e-learning educational interventions, such as large- group classroom instruction and printed text, in improving learning outcomes." (page 7 line 99-101) I do not think this is true. In fact, the Cook et al |

paper they referenced does not come to this conclusion. Cook et al emphasize that design of online learning modalities can influence effectiveness and can make them better that traditional learning. I would encourage the authors to re-read this reference as well as the reviews Cook et al completed prior to and after the one they cited. The claim made by the authors might not be correct and as it seems to be an integral piece of their 'story' as to why this review is being done - re-writing this piece in the background will be necessary in my opinion.

Minor points:

Page 3 line 59 "AEEs may be more beneficial more learning" I am not sure what this means and whether it is true?

Page 6 line 84 "However, designers of e-learning environments and educators rarely make use of this data to optimize learning effectiveness and efficiency." This might be true but this conclusion definitely needs a reference.

"Two review authors (T.M., M.-F.D.) validated the data extraction forms." – Please describe how they were validated.

| REVIEWER | Mari Lahti |
| | Turku University of Applied Scince and Turku University |
| REVIEW RETURNED | 18-Nov-2018 |

| GENERAL COMMENTS | Reviewer's comments: |
| | |
| | This is most certainly a timely piece of work addressing an issue that is of upmost important. |
| | Concept of adaptive learning is relatively new and we need more knowledge about its effects. |
| | |
| | I have provided some comments that hopefully will help you to improve your manuscript: |
| | |
| | Introduction and background: |
| | 1. I would like to have a bit more extended definition about AEE. Now this is not clear by all means. |
| | 2. How AEE is more effective and what makes it so effective? |
| | |
| | Methods: |
| | 1. Research questions are well formulated! |
| | 2. Clear inclusion and exclusion criteria's could be written here. |
| | 3. Why you choose to search between 2005 to 2017? What justifies to start on year 2005? |
| | 4. Heterogeneity %-values seems unclear, how it is estimated if the I2 % is 80%? |
| | |
| | Results: |
| | 1. Well written result part. |
| | 2. It would have been nice to read a bit more about the elearning interventions. |
| | |
| | Conclusions: |
| | 1. It would be good to have a bit deeper discussion about the effects of this review to science, clinical practice etc. |

| | In general, this review is really well written and conducted. Its rare to be able to read such a fine piece of work. I feel privileged to review this and learn myself too. Best of luck with this paper! |

| **REVIEWER** | Alexandra Ellis<br>Institute for Clinical and Economic Review, USA<br>Brown University, USA |
| **REVIEW RETURNED** | 22-Dec-2018 |

| **GENERAL COMMENTS** | The authors conducted a systematic review a meta-analysis to assess learning environments on educational outcomes in health professionals and students. My review focuses on the statistical elements; however, I did find the introduction difficult to follow and suggest that a content expert provide feedback to the authors. As noted below, I am not convinced that the meta-analytic approaches were appropriate based on what the authors reported. Specific concerns related to the validity of the meta-analysis using Standardized Mean Differences (SMD) include:<br>(1) Outcomes: It is not clear what "knowledge" or "competence" means or how it is measured. Please provide a description of how the included studies measured these outcomes and also any similarities or differences across the studies. The assumptions for using SMD require that the studies are measuring the same concept albeit with perhaps different scales; without such detail reported, it is not possible for the reader to determine the validity of the SMD analysis.<br>(2) Populations: As the authors note in the discussion section, SMD assumes the differences in SD's should not be attributed to different variance across the studies' populations. The authors continue by stating they judged SMD to be the best option. Yet, in the subsequent paragraph the authors state that they attribute the differences in study results mainly to the differences in populations. As such (and in addition to the likely differences in outcomes noted above), the SMD analysis does not seem appropriate.<br><br>The authors might consider removing the meta-analyses and focus on the qualitative aspects of their literature review, for which there is much to discuss. I do believe there is value in this study's descriptive and qualitative components.<br><br>Minor comments:<br>(3) Publication year: There are some places were the start year is 2005 and others where the year is 2006 (e.g., page 14 line 269. Please check for consistency. Please also provide more rationale for the 2005 cut-off. While it may be true that "older" studies are different than more recent studies, is there something that occurred in 2005? e.g., why not 2006 or 2004? Any reason why focusing on publication date rather than study enrollment or completion?<br>(4) The "Types" of interventions: As I understand, Type B2 is a subset of B1. However, some text implies that the two are disjoint (e.g., page 9, line 163). Please clarify and perhaps update the labels.<br>(5) Study selection: In addition to the mechanisms of screening studies, please provide details regarding the study selection criteria. E.g., were there specific studies that you excluded based |

| | on a design or intervention characteristics (e.g., an intervention with both e-learning and non-e-learning attributes?) |
| | (6) Provide context for details in Table 2. What is cognitive tutoring, slide tutor, etc.? Please provide more details in the text. |
| | |
| | I hope the authors find these comments and suggestions helpful in their work. |

**VERSION 1 – AUTHOR RESPONSE**

Response to the Comments of Reviewer #1

Major Comments

1. I don't believe this paper provides an adequate explanation for constraining the article search to the years 2005 onward. After reviewing the source being cited (Kulik & Fletcher, 2015), there is no evidence that bias of earlier studies is a key issue in their findings. Overall, the relationship between publication year and effect size wasn't significant. Furthermore, in looking at the mean effect sizes for the different year categories, there wasn't much difference between 2001 - 2005 and 2005 onward. The larger difference appeared to occur in studies up to the year 2000. If anything, Kulik & Fletcher suggest that the relevant issue to consider would be implementation adequacy, something that doesn't appear to correlate with year of implementation, but rather the number of times any given person has implemented an ITS (i.e., prior experience). Unless you can provide additional support for why you should constrain to a specific year, please open your search to include all years.

Reply: We have updated our search strategy to include all years, from the inception of each database up to February 2019. This led to the inclusion of four additional studies in the systematic review (Casebeer-2003, de Ruijter-2018, Lee-2017, Micheel-2017).

Minor Comments

2. Line 59: Typo--"more beneficial more learning" should say "more beneficial for learning"

Reply: The correction has been made.

3. Lines 161-162: Regarding your definitions of each type of adaptation, given the age of the paper you are citing (Knutov & colleagues), did you account for finding adaptation processes that would fall outside the scope of the definitions proposed by Knutov et al? Did you find any additional systems falling outside these definitions that might be relevant? Please clarify in text.

Reply: We thank the reviewer for this comment. We did not find any type of adaptivity different from the 5 types proposed by Knutov and his colleagues (content, navigation, presentation, multimedia and tools). However, as indicated in the manuscript, we characterized the process of adaptivity by 5 subdomains: adaptivity method, adaptivity goals, adaptivity timing, adaptivity factors, and adaptivity types (this last subdomain being the focus of Knutov and colleagues)9. Thus, the 5 subdomains we propose in this review go beyond the work of Knutov and his colleagues.

4. Line 168: Please provide operational definitions of your outcomes of interest so it is clear what you were trying to capture.

Reply: We defined the outcomes of interest (knowledge, skills, behavior) based on the work of Cook et al. (2008) in the section "Study Eligibility" on page 11 in the revised manuscript.

5. Line 338 (rather, the section under this heading): Please add a statement clarifying that some studies used multiple adaptation techniques (with a reference to table 2).

Reply: We added the following sentence to the section: "Overall, 17 out of 21 (81%) AEEs examined integrated more than one type of adaptivity."

6. Line 362: For the study that only implemented adaptation at the beginning, what was the adaptation based on? Survey responses? Please include a clarifying statement.

Reply: An additional study, found after updating the search strategy, implemented adaptivity at the beginning of the training. We added a clarifying statement to the sentence: "In two studies, adaptivity was only implemented at the beginning of the training with the AEE following survey response".

7. Line 406: Word correction-- "same knowledge scores than...." should say "same knowledge scores as..."

Reply: The correction has been made.

Response to the Comments of Reviewer #2

Major Comments

1. Overall, the authors have taken a very detailed systematic approach to this systematic review and meta-analysis. Before publication I would like further clarification on how they defined knowledge and competence as outcomes of their meta-analysis. These terms have been used in a variety of different

contexts in medical education and have also been used interchangeably (incorrectly in my opinion). Thus it would be prudent for the authors to outline how they defined knowledge and competence as it applied to their search and analysis. I'm sure this was also quite heterogenous in the studies that were included which might make it difficult to make conclusions from the analysis. This would be my major concern. Educators especially in this competence by design (CBD) era in medicine are trying to precisely predict success in training by looking at both knowledge and competence with one not meaning attaining the other. One good reference is Epstein RM, Hundert EM. Defining and Assessing Professional Competence. JAMA. 2002;287(2):226–235. doi:10.1001/jama.287.2.226

Reply: We reviewed thoroughly the scientific literature on the concept of professional competence, including the paper suggested by Reviewer 2 (Epstein and Hundert, 2002), and we believe our review focused on procedural and cognitive skills rather than competence. Thus, we made some changes to the manuscript to reflect this important distinction by replacing the term "competence/competencies" by "skill/skills" where appropriate. Moreover, we defined the outcomes of interest (knowledge, skills, behavior) based on the work of Cook et al. (2008) in the section "Study Eligibility" on page 11 in the revised manuscript.

2. After discussing knowledge/competence and carrying this through their analysis (which could change their conclusions), I would then ask the authors to include a section on further recommendations as to how an educator should structure their own AEE tool if they were creating one. This has been done in pieces in the paper but as a reader, this is what will have an impact on me and lead to knowledge translation. As of right now, it is reading very much like a descriptive paper only.

Reply: We thank Reviewer #2 for this suggestion. We have added an additional table to our paper (Table 3, Pages 37-8) to highlight 8 practical considerations for the design and development of adaptive e-learning environments for educators and educational researchers.

3. In the background, the authors make the claim that "…they do not consider users' characteristics to provide a personalized training. They are generally considered to be as effective as non–e-learning educational interventions, such as large- group classroom instruction and printed text, in improving learning outcomes." (page 7 line 99-101) I do not think this is true. In fact, the Cook et al paper they referenced does not come to this conclusion. Cook et al emphasize that design of online learning modalities can influence effectiveness and can make them better that traditional learning. I would encourage the authors to re-read this reference as well as the reviews Cook et al completed prior to and after the one they cited. The claim made by the authors might not be correct and as it seems to be an integral piece of their 'story' as to why this review is being done - re-writing this piece in the background will be necessary in my opinion.

Reply: We reviewed the Cook et al. paper and we agree with Reviewer #2. Thus, we removed this statement and we have rewritten the section of the introduction on the different types of e-learning environments in order to be accurate (Page 8).

Minor Comments

4. Page 3 line 59 "AEEs may be more beneficial more learning" I am not sure what this means and whether it is true?

Reply: This sentence has been corrected.

5. Page 6 line 84 "However, designers of e-learning environments and educators rarely make use of this data to optimize learning effectiveness and efficiency." This might be true but this conclusion definitely needs a reference.

Reply: We have added a reference to support this statement.

6. "Two review authors (T.M., M.-F.D.) validated the data extraction forms." – Please describe how they were validated.

Reply: A statement has been added to clarify the validation process. "Two review authors (T.M., M.-F.D.) validated the data extraction forms by reviewed the contents of each form against the data in the original article, adding comments when changes were needed."

Response to the Comments of Reviewer #3

Comments

1. This is most certainly a timely piece of work addressing an issue that is of upmost important. Concept of adaptive learning is relatively new and we need more knowledge about its effects. I have provided some comments that hopefully will help you to improve your manuscript.

Reply: We thank Reviewer #3 for the feedback.

2. I would like to have a bit more extended definition about AEE. Now this is not clear by all means.

3. How AEE is more effective and what makes it so effective?

Reply: We have rewritten the section of the introduction on AEEs in order to better define them and to specify which methods can potentially increase their efficiency and effectiveness in comparison with other training methods (Page 8).

4. Research questions are well formulated!

Reply: We thank Reviewer #3 for this comment.

5. Clear inclusion and exclusion criteria's could be written here.

Reply: We added additional details regarding the inclusion and exclusion criteria of this review in the section "Study Eligibility" (Pages 10-1).

6. Why you choose to search between 2005 to 2017? What justifies to start on year 2005?

Reply: We have updated our search strategy to include all years (from the inception of each database) up to February 22 2019. The time span searched is now 1971-2019. This led to the inclusion of four additional studies in the review.

7. Heterogeneity %-values seems unclear, how it is estimated if the I2 % is 80%?

Reply: As we specified on Page 14 in the manuscript, we used the I2 statistic computed by the RevMan software to quantify how much the results varied across individual studies (i.e., between-study inconsistency, or statistical heterogeneity). The I2 statistic "describes the percentage of the variability in effect estimates that is due to heterogeneity rather than sampling error (chance)" (source: Cochrane, https://bit.ly/2FTNdvL).

8. Well written result part.

Reply: We thank Reviewer #3 for this comment.

9. It would have been nice to read a bit more about the elearning interventions.

Reply: We have added a section to describe the different types of adaptive e-learning platforms (Page 22).

10. It would be good to have a bit deeper discussion about the effects of this review to science, clinical practice etc.

Reply: We have reworked the section on implications for practice and research (Pages 36-8). We have added an additional table to our paper (Table 3, Pages 37-8) to highlight 8 practical considerations for the design and development of adaptive e-learning environments for educators and educational researchers.

11. In general, this review is really well written and conducted. Its rare to be able to read such a fine piece of work. I feel privileged to review this and learn myself too. Best of luck with this paper!

Reply: We thank Reviewer #3 for this comment.

Response to the Comments of Reviewer #4

Major Comments

1. The authors conducted a systematic review a meta-analysis to assess learning environments on educational outcomes in health professionals and students. My review focuses on the statistical elements; however, I did find the introduction difficult to follow and suggest that a content expert provide feedback to the authors. As noted below, I am not convinced that the meta-analytic approaches were appropriate based on what the authors reported. Specific concerns related to the validity of the meta-analysis using Standardized Mean Differences (SMD) include: (1) Outcomes: It is not clear what "knowledge" or "competence" means or how it is measured. Please provide a description of how the included studies measured these outcomes and also any similarities or differences across the studies. The assumptions for using SMD require that the studies are measuring the same concept albeit with perhaps different scales; without such detail reported, it is not possible for the reader to determine the validity of the SMD analysis. (2) Populations: As the authors note in the discussion section, SMD assumes the differences in SD's should not be attributed to different variance across the studies' populations. The authors continue by stating they judged SMD to be the best option. Yet, in the subsequent paragraph the authors state that they attribute the differences in study results mainly to the differences in populations. As such (and in addition to the likely differences in outcomes noted above), the SMD analysis does not seem appropriate.

The authors might consider removing the meta-analyses and focus on the qualitative aspects of their literature review, for which there is much to discuss. I do believe there is value in this study's descriptive and qualitative components.

Reply: We thank Reviewer 4 for these relevant comments on review outcomes and the use of the Standard Mean Difference (SMD) in this meta-analysis. First, to address the reviewer's comment, we defined the outcomes of interest (knowledge, skills, behavior) based on the work of Cook et al. (2008) in the section "Study Eligibility" on page 11 in the revised manuscript. Second, we would like to specify that our use of the SMD is in accordance with the Cochrane guidelines: "The standardized mean difference is used as a summary statistic in meta-analysis when the studies all assess the same outcome but measure it in a variety of ways (for example, all studies measure depression but they use different psychometric scales)" (source: Cochrane, https://bit.ly/2JmRlrv). Indeed, in this review, the three outcomes of interest (knowledge, skills, behavior) were the same conceptually in examined studies, but were measured with different scales. Moreover, there were no significant differences in study populations since all participants were health professionals and students (mostly physicians and medical residents). Thus, we believe this to be an appropriate use of the SMD. Multiple meta-analyses in the field of medical education have also used the SMD for the same outcomes of interest:

• Cook, Levinson, A. J., Garside, S., Dupras, D. M., Erwin, P. J., & Montori, V. M. (2008). Internet-based learning in the health professions: a meta-analysis. Journal of the American Medical Association, 300(10), 1181-1196. doi:10.1001/jama.300.10.1181

• Cheng, A., Lockey, A., Bhanji, F., Lin, Y., Hunt, E. A., & Lang, E. (2015). The use of high-fidelity manikins for advanced life support training—a systematic review and meta-analysis. Resuscitation, 93, 142-149. doi: 10.1016/j.resuscitation.2015.04.004

• Johnson, J., & Panagioti, M. (2018). Interventions to improve the breaking of bad or difficult news by physicians, medical students, and interns/residents: a systematic review and meta-analysis. Academic Medicine, 93(9), 1400-1412. doi: 10.1097/ACM.0000000000002308

In addition, since we updated our search strategy to include all years up to 2019, we included 4 additional studies in the review, 3 of which could be integrated in the meta-analysis. Thus, we strongly believe in the value of the quantitative synthesis of evidence in this review.

Minor Comments

2. Publication year: There are some places were the start year is 2005 and others where the year is 2006 (e.g., page 14 line 269. Please check for consistency. Please also provide more rationale for the 2005 cut-off. While it may be true that "older" studies are different than more recent studies, is there something that occurred in 2005? e.g., why not 2006 or 2004? Any reason why focusing on publication date rather than study enrollment or completion?

Reply: We have updated our search strategy to include all years (from the inception of each database) up to February 22 2019. The time span searched is now 1971-2019. This led to the inclusion of four additional studies in the review.

3. The "Types" of interventions: As I understand, Type B2 is a subset of B1. However, some text implies that the two are disjoint (e.g., page 9, line 163). Please clarify and perhaps update the labels.

Reply: We have rewritten the section of the introduction on adaptive e-learning environments in order to better define this type of intervention, and to specify which methods can potentially increase their efficiency and effectiveness in comparison with nonadaptive e-learning environments and conventional training methods (Page 8).

4. Study selection: In addition to the mechanisms of screening studies, please provide details regarding the study selection criteria. E.g., were there specific studies that you excluded based on a design or intervention characteristics (e.g., an intervention with both e-learning and non-e-learning attributes?)

Reply: We added additional details regarding the inclusion and exclusion criteria of this review in the section "Study Eligibility" (Pages 10-1). We specified that we considered for inclusion studies in which AEEs had designed or algorithmic adaptivity, and studies including a co-intervention in addition to adaptive e-learning (e.g. paper-based instruction).

5. Provide context for details in Table 2. What is cognitive tutoring, slide tutor, etc.? Please provide more details in the text.

Reply: The theoretical frameworks used in examined studies, such as cognitive tutoring and perceptual learning, are described on page 21. In addition, we have included a section regarding the adaptive e-learning platforms, such as Slide Tutor, that were used in examined studies (Page 22).

### VERSION 2 – REVIEW

| REVIEWER | Mari Lahti |
| | Turku University of Applied Science, Finland |
| REVIEW RETURNED | 02-Apr-2019 |

| GENERAL COMMENTS | Thank you for authors of following closely the reviewer's earlier comments. This paper is timely piece presenting interesting results of AEE. I feel that this paper has improved a lot and I am willing to recommend this paper to publication. |
| | |
| | I appreciate authors hard work of conducting such a fine paper. Best of luck with getting this published. |

| REVIEWER | Alexandra Ellis |
| | Institute for Clinical and Economic Review, US |
| REVIEW RETURNED | 14-Apr-2019 |

| GENERAL COMMENTS | I thank the authors for updating their literature search and revising content from the first round of review. The strengths of this work are in its qualitative components, which (rightly) comprises the |

| | most text in the results section. Indeed, there are sufficient results on the qualitative components for a manuscript - I have yet to be convinced that the SMD meta-analysis is appropriate for this dataset, or that it contributes meaningfully to the broad purpose of this work. If the authors decide to continue this approach, please address the following:<br><br>(1) More details on the outcomes and results reported by each study are needed. In the forest plots, the study-specific results are available. There is clearly a wide range in scales, but as a reader, I do not know what metrics were used in each. In addition, do any of the studies use the same metric or all the each unique? In the methods section, the authors note that the study-specific definitions were abstracted - please present them as well as your assessment of how similar they are.<br><br>(2) Regardless of the study-specific outcomes, more justification for quantitatively synthesizing the results are needed in light of the differences in designs/patients across the studies. Do you anticipate any effect modification by the range of baseline characteristics? For example, years of education or years of experience - the authors note that some studies focused on medical studies while others on physicians in practice. In addition, the "knowledge" and "skills" necessary for microscopy are different than for diagnostic imaging or for behavior change counseling.<br><br>Minor comments:<br>-The authors state that 4 studies had missing data - please be specific in what data were missing<br>-Lines 309-311 list 6 outcomes, yet only 3 of them have been defined in the methods section. Please provide the other 3 definitions, or otherwise remove if they are not pertinent to the reminder of the paper.<br>-Typos: Line 90 "ITC" should be "ICT"; line 137 contains "in in"; line 190 "accordaSsnce" should be "accordance" |
|---|---|

**VERSION 2 – AUTHOR RESPONSE**

Response to the Comments of Reviewer #3

Comment #1

Thank you for authors of following closely the reviewer's earlier comments. This paper is timely piece presenting interesting results of AEE. I feel that this paper has improved a lot and I am willing to recommend this paper to publication. I appreciate authors hard work of conducting such a fine paper. Best of luck with getting this published.

Reply to Comment #1:

We thank reviewer #3 for her comment.

Response to the Comments of Reviewer #4

Comment #1

I thank the authors for updating their literature search and revising content from the first round of review. The strengths of this work are in its qualitative components, which (rightly) comprises the most text in the results section. Indeed, there are sufficient results on the qualitative components for a manuscript - I have yet to be convinced that the SMD meta-analysis is appropriate for this dataset, or that it contributes meaningfully to the broad purpose of this work.

Reply to Comment #1:

We would like to thank reviewer #4 for her insights and comments. We have reviewed the literature in our field and examined the appropriateness of the SMD meta-analysis. After careful consideration, we have decided to continue with this approach. However, as suggested, we have made specific changes to our manuscript to ensure readers have enough details to interpret the results (see below).

Comment #2

If the authors decide to continue this approach, please address the following: More details on the outcomes and results reported by each study are needed. In the forest plots, the study-specific results are available. There is clearly a wide range in scales, but as a reader, I do not know what metrics were used in each. In addition, do any of the studies use the same metric or all the each unique? In the methods section, the authors note that the study-specific definitions were abstracted - please present them as well as your assessment of how similar they are.

Reply to Comment #2:

To provide additional details regarding the outcome measures of each study for the outcomes of interest in the review (i.e., knowledge, skills, clinical behaviour), we have modified Table 1 "Characteristics of Included Studies" on page 18. We added a column titled "Outcome Measures", which reports how each outcome was measured (study-specific metrics), and the score range (if reported). In addition, we now discuss the similarity of outcome measures used across studies on page 17, lines 314-326:

• Outcome measures for knowledge were similar across studies: in 9 out of 14 studies measuring knowledge, investigators employed multiple-choice questionnaires developed by the research team with input from content experts that were tailored to training content to ensure specificity. Knowledge was also assessed using true-false questions in two studies, and the type of questionnaire was not specified in three studies. Outcome measures for skills were also similar across the 9 studies reporting this outcome, since in all studies investigators measured cognitive skills rather than procedural skills. Indeed, all outcomes measures for skills were related to clinical reasoning. In 6 studies, skills were measured through tests that included a series of diagnostic tests (eg electrocardiograms, x-rays, miscroscopy images) that learners had to interpret. In 3 studies, skills wre measured through questions based on clinical situations in which learners had to specify how they would react in these particular situations. We were not able to describe the similarity between the outcome measures for clinical behaviour no details were provided in one of the two studies reporting this outcome.

Comment #3

Regardless of the study-specific outcomes, more justification for quantitatively synthesizing the results are needed in light of the differences in designs/patients across the studies. Do you anticipate any effect modification by the range of baseline characteristics? For example, years of education or years of experience - the authors note that some studies focused on medical studies while others on physicians in practice. In addition, the "knowledge" and "skills" necessary for microscopy are different than for diagnostic imaging or for behavior change counseling.

Reply to Comment #3:

To investigate the potential effect modifications suggested by Reviewer #4, we conducted multiple subgroup analyses through the RevMan software. More specifically, for each individual outcome, we conducted subgroup analyses according to population (healthcare professionals versus healthcare students) and comparator (adaptive e-learning versus nonadaptive e-learning, adaptive e-learning versus paper-based instruction, adaptive e-learning versus classroom-based instruction). No statistically significant differences between subgroups were found regarding the effect sizes for both knowledge and skills. We have provided these additional details regarding subgroup analyses on page 31, lines 464-468.

Comment #4

The authors state that 4 studies had missing data - please be specific in what data were missing.

Reply to Comment #4:

We added additional details on page 16, lines 285-287: "The 4 studies with missing data did not report data regarding the results, i.e. mean scores and standard deviations in both study groups at post-test, regarding the outcomes of interest in this review (i.e., knowledge, skills or clinical behavior)." These 4 studies were thus excluded from the meta-analysis.

Comment #5

Lines 309-311 list 6 outcomes, yet only 3 of them have been defined in the methods section. Please provide the other 3 definitions, or otherwise remove if they are not pertinent to the reminder of the paper.

Reply to Comment #5:

The three outcomes not relevant to the review were removed.

Comment #6

Typos: Line 90 "ITC" should be "ICT"; line 137 contains "in in"; line 190 "accordaSsnce" should be "accordance".

Reply to Comment #6:

The corrections have been made.

## VERSION 3 - REVIEW

| REVIEWER | Alexandra Ellis<br>Brown University, USA |
|---|---|
| **REVIEW RETURNED** | 27-Jun-2019 |

| **GENERAL COMMENTS** | Thank you for revising the manuscript based on the prior feedback.<br>In the response to reviewers, the authors state they have updated the few typos highlighted, but they remain in this version. I've copied them below and trust the authors will make the revisions without further review.<br><br>Typos: Line 90 "ITC" should be "ICT"; line 137 contains "in in"; line 190 "accordaSsnce" should be "accordance". |