Author's response to reviews

Title: Teaching of Evidence-Based Medicine to medical students in Mexico: A randomized controlled trial.

Authors:

  Melchor Sánchez-Mendiola (melchorsm@gmail.com)
  Luis F Kieffer-Escobar (drluiskieffer@gmail.com)
  Salvador Marín-Beltrán (salvador_marin@yahoo.com)
  Steve M Downing (sdowning@uic.edu)
  Alan Schwartz (alansz@uic.edu)

Version: 2 Date: 10 October 2012

Author's response to reviews: see over
Author's cover letter for revised manuscript submission

**MS:** 1067516138769915

**Title:** Teaching of Evidence-Based Medicine to medical students in Mexico: A randomized controlled trial.

**Authors:** Melchor Sánchez-Mendiola, Luis F. Kieffer-Escobar, Salvador Marín-Beltrán, Steve M. Downing and Alan Schwartz.

**Version:** 2

**Date:** October 10, 2012

**Editor, BMC Medical Education**

Dear Sirs,

We are submitting the revised version of the enclosed manuscript titled “Teaching of Evidence-Based Medicine to medical students in Mexico: A randomized controlled trial”, for consideration as an original research report for publication in your journal. In the following pages of this cover letter we provide a point-by-point response to the reviewers’ concerns.

We would like to express our heartfelt appreciation to the three reviewers, whose insightful comments and detailed recommendations motivated us to improve the quality of our paper. We believe that this revised version will satisfy the requirements of the reviewers and the journal, and look forward to the editorial decision.

Sincerely,

Melchor Sánchez-Mendiola MD, MHPE

Communications about the paper should be directed to me at the following address:

Melchor Sánchez-Mendiola MD, MHPE
Secretaría de Educación Médica, Facultad de Medicina de la UNAM
Ave. Universidad 3000, Ciudad Universitaria, México, D.F., México. 04510
Telephone number: (5255) 5623-2448
Fax number: (5255) 5550-0065
E-mail address: melchorsm@gmail.com
RESPONSE TO REVIEWERS

Reviewer 1: Carlos Estrada

MAIN COMMENTS

1. Main groups and analysis. Comparing all four groups simultaneously is problematic. Several aspects to address:
   a. For the main outcomes, the first and main comparison groups throughout the methods and results ought to be the randomized groups (intervention vs. control). This approach preserves randomization (balancing known and unknown factors). This section could be labeled as “Main Outcomes.”

   ✓ Agree with the reviewer. We rearranged the description of the Methods and Results, emphasizing the main outcomes of the randomized M5 groups.

   b. After the main outcomes are reported, then can include a section of “External Validation.” The comparisons of control with non-exposed group and intervention with previously exposed group greatly enhance the scientific validity of the study.

   ✓ Agree with the reviewer. We explain and justify the use of the M4 and M6 groups in the study design, and labeled the section ”Simultaneous validation”. We preferred this term instead of “External validation” to avoid confusion or ambiguity due to its technical meaning in clinical and educational research, which is more directed to applicability to groups and environments outside the research setting (which is not the case for the M4 and M6 groups).

   c. Analysis and reporting. The use of a Chi Square is problematic. The comparison between the two randomized and independent groups (intervention vs. control) could be done with a non-parametric test (Mann-Whitney U test) or a Chi-Square for trend. The information in Tables 2-4 can be summarized in 1-2 figures.

   ✓ Agree with the reviewer. We reanalyzed the data of the M5 EBM vs. M5 non-EBM groups in Tables 2 and 3 with the Mann-Whitney U test, removed those tables and replaced them with two figures to make the information more visually understandable.

   ✓ We eliminated Table 4, since those data are included in Figure 2 (critical appraisal confidence scores).

   d. How are the summary figures reporting the Taylor’s questionnaire different from Tables 1-4?

   ✓ Table 1 should remain since it outlines the content of the EBM course, we think this information is important for content validity of the test. Tables 2 and 3 are removed and replaced with Figures 2 and 3, presenting the information in
graphical form. These two figures present data from the first portion of Taylor's questionnaire, which asks two questions related to the type of resources one uses to keep up to date and to solve a specific healthcare problem. These two questions are separate from the other portions of Taylor's questionnaire that are described in the other figures (critical appraisal skills, attitudes and knowledge about EBM). Table 4 is eliminated as mentioned above in the response to "c".

e. Taylor's questionnaire figures could be summarized in a single graph. Is there a summary overall measure for this questionnaire?

✓ Taylor's instrument doesn't have a summary overall measure, since it is designed to measure different constructs: use of evidence, critical appraisal skills, attitudes to EBM and EBM knowledge. Therefore, we think each figure should stand on its own.

2. Main outcomes. The main outcomes seem to be: a) EBM knowledge, attitudes, skills (using Taylor's questionnaire), b) knowledge test (100 items), and c) instructional sensitivity.

   a. The Methods section should parallel the main objective and the main outcomes.

✓ Agree with the reviewer. We revised and corrected the Methods section following this advice.

   b. If instructional sensitivity is a worthy outcome (this reviewer was unaware of this metric), the reader would benefit from a brief explanation (the current explanation was unclear).

✓ Instructional sensitivity is a family of indices sometimes used by educational measurement specialists. The one we used is the Pre-to-Post Difference Index, which: "provides the simple difference in item difficulty based in two samples of test takers known to differ with respect to instruction" (Haladyna T. Developing and Validating Multiple-Choice Items, 3rd Ed, 2004). Nonetheless it is an index rarely used in mainstream medical education assessment literature, and probably does not have enough added value beyond the measures of difference reported in our paper, so we decided to exclude it in the revised manuscript.

3. Methods, lack of clarity and order. The report would benefit from significant reordering of the information; it does not do justice to the quality of the work. Some, but not all, examples:

   a. Reorder all aspects that are related and remove duplication. Refer to randomization and groups Figure very early in the Study Design section.
We reordered and rewrote the Methods section striving for more clarity. We removed duplication as much as possible and referred to the flow diagram of participants figure early in Study Design.

b. Description of the groups should parallel comment 1a (above).

de. Taylor Questionnaire. Did the authors use the full 152-item test? http://onlinelibrary.wiley.com/doi/10.1046/j.13652923.2001.00916.x/abstract;jsessionid=6D5FC6B37CD42EE871F41556B4F58F1D.d04t03 The description could be shortened.

done.

e. Move item analysis to the analysis section, or remove completely from the manuscript. The purpose of the present report does not seem to be the study of the psychometric characteristics of the questionnaire.

We removed the detailed data of the EBM MCQ test psychometric analysis, and added a brief comment about the method of analysis and the instrument's reliability.

f. The detailed description of the setting of the test (pg 10) is unnecessary.

We removed the detailed description as recommended.

We reordered and rewrote the Methods section striving for more clarity. We removed duplication as much as possible and referred to the flow diagram of participants figure early in Study Design.

c. Split setting and randomization, the latter should go with study participants. Setting should be the first section in the Methods.

done.

e. Move item analysis to the analysis section, or remove completely from the manuscript. The purpose of the present report does not seem to be the study of the psychometric characteristics of the questionnaire.

We removed the detailed data of the EBM MCQ test psychometric analysis, and added a brief comment about the method of analysis and the instrument's reliability.

f. The detailed description of the setting of the test (pg 10) is unnecessary.

We removed the detailed description as recommended.

g. Methods description is included in the Discussion. Remove any and all description of the Methods included in the Discussion section.

done.

4. Conclusion. The statement that the intervention persists after one year cannot be concluded from the study design – it might be inferred, but not concluded. Please
tone down this conclusion. The conclusion can be reach with 1 year follow-up of the INTERVENTION group that was randomized, not a 1 year follow-up of a non-randomized group.

✓ Agree with the reviewer. We rewrote the different areas of the manuscript that had this type of statements, and decreased the categorical tone of the discussion and conclusions regarding the M6 group data.

OTHER COMMENTS
1. Abstract. Remove/reword the statement of the first study. Could reword to evidence from rigorously conducted studies is limited.

✓ We reworded the mentioned statement, as suggested.

2. Clearly state the objective of the study (remove delayed as mentioned above) that matches the Abstract, Introduction, and Discussion section.

✓ We rewrote the objective as suggested, and modified the relevant areas of the manuscript to be aligned with the stated objective.

3. Strengthen the Introduction/Background section with some of the literature review included in the Discussion.

✓ Done.

4. Box plots. Figures require being self-explanatory. Add comparison p values: intervention vs. control, control vs. concurrent non-EBM (M4), intervention vs. concurrent post-EBM (M6).

✓ Done.

5. The first paragraph of the Discussion is unnecessary.

✓ Discussion was rewritten and the first paragraph was removed.

6. The Discussion section could be shortened by 30%. May move some aspects to the Introduction.

✓ The Discussion was shortened and some relevant portions were moved to the Introduction.

7. Table 5 is unnecessary; the same information could be summarized in a 2-3 sentences in the text (or in a single figure).

✓ Table 5 was removed and the information was summarized in the text.
Reviewer 2. Nai Ming Lai

Reviewer’s report:
Researchers in evidence based medicine (EBM) education today face a double challenge: the expectations on them to demonstrate the effectiveness of EBM training that they champion in accordance with the standards that they preach, and a common difficulty for medical education researchers in demonstrating “educationally important” learning gains, i.e. learning gains that are likely to impact on the student’s future practice. The authors of this paper addressed one of these challenges by designing a randomised controlled trial comparing a block EBM curriculum versus a controlled, non-EBM curriculum. I believe the main value of this study is that it presents a higher-level evidence in confirming what previous studies have consistently shown that EBM training improves knowledge and confidence at the completion of the training. I am less convinced, however, that the degree of learning gain demonstrated in the study was educationally important as stated by the authors, as the significant loss of knowledge in the students who received EBM training six months to one year earlier left one to wonder whether there would be any material amount of knowledge retained that could influence clinical decision-making in the future. In fact, the over-optimistic and sometimes misleading interpretation of their findings, which is found consistently in the abstract and discussion, is to me a major criticism of the paper. I would also suggest that the authors tighten up their discussion to focus on what is directly substantiated by the findings and not elaborate too much on general issues in teaching and application of EBM.

✔ We considered all the reviewer’s concerns, and modified the different sections of the manuscript accordingly. Please see the point-by-point responses below.

Following are specific comments pertaining to different parts of the paper:

• Major Compulsory Revisions (which the author must respond to before a decision on publication can be reached)

Abstract, Background: “The objective was to assess the learning of EBM by undergraduate medical students, and retention one year later.”
Results
“Confidence in critical appraisal skills was higher in the intervention groups and was maintained one year later.”
Conclusions
“The educational intervention has a positive effect that may persist one year later.”
In my opinion, there is misleading use of wording in the sentences quoted above. I think words such as “retention”, “maintained” and “persist” should not be used, as retention was only implied, and not directly assessed in the study, because the study evaluated EBM knowledge and confidence among different groups of students, and not the same group over time. I suggest that the authors simply describe the results without making any assumption, for example, “confidence in critical appraisal skills
was higher in the intervention group and the group of students who received EBM instruction in the previous year”.

✓ We eliminated the words that could be construed as misleading, and described the results without making inappropriate inferences.

Methods, study design
“The outcomes were measured at the end of the first semester of the academic year, after the EBM course ended, comparing the fifth year half-group that took the course with the 5th year half without the course, the 4th year students without the course, and the 6th year students that had the course one year previously”. If there were no changes in the curriculum in the authors’ institution, I figure that there would be two distinct groups of 6th-year students in relation to EBM exposure: one who received EBM in the first half of their fifth year and one in the second half. It seems more appropriate to separate these two groups to more usefully evaluate the amount of knowledge and confidence in students who were six months and one year post-intervention.

✓ We agree with the reviewer in the sense that half of the 6th year student population had the EBM course six months before, and the other half one year previously. The situation is that the medical school changes the groups' makeup every year, so the 6th year group was mixed again at the beginning of the academic year. We don't have the information about which student belonged originally to which group when they were in 5th year, since the results were anonymized to comply with the ethical aspects of IRB approval. Therefore, we cannot separate the M6 group in two groups to evaluate them separately, and we can't speculate on the amount of knowledge or change in attitudes that could be found in the M6 subpopulations.

It was also stated that students in year four were exposed to some elements of EBM (statistics, medical informatics, research methodology and epidemiology) and the authors distinguished this from their EBM course but made no further description of the difference between the delivery of these subjects in year four versus year five. This is important in light of the authors’ claims in their discussion that EBM intervention in year five appeared more effective than teaching of similar subjects such as clinical epidemiology and research methodology in year four, which should lead to an exploration on possible reasons for the difference. A description of the methods of delivery in the subjects of epidemiology and research methodology in year four should be included here.

✓ We acknowledge the reviewer’s concerns and added some information regarding these courses in the text. We are not certain of the specific teaching methodology and assessment instruments used in those courses, so we can’t make specific inferences regarding their differential effectiveness with our EBM course. It is interesting to note that the M4 students had a slightly higher score than the M5 non-EBM group in the critical appraisal skills portion of Taylor’s instrument (Figure 4 of the revised manuscript). The items in this
score explore research methodology and statistics concepts, which M4 students could have fresh in their minds compared to the M5 non-EBM group, at the time the instruments were applied.

“The assignment of students to groups was done by the medical school using Microsoft Excel for block randomization (two blocks with 48 students each in this case, for the 5th year class)"  
I am unclear how block randomization was actually achieved in this study. It appears to me that the methods of randomization as described by the authors, using “two blocks each of 48 participants” was an unlikely way of randomization. It seems more appropriate in this case to randomize using blocks of two (instead of blocks of 48). Otherwise, using two blocks of 48 would be akin to cluster randomization, with different implications on analysis and reporting of results. The authors may wish to clarify this and describe their methods of randomization more accurately.

✔ We agree with the reviewer, the text was inaccurate. We modified the randomization description, clarifying that the use of blocks of two was done to ensure equal sample sizes.

Methods, Statistics, data analysis and ethical aspects.  
“The smallest meaningful difference (SMD) for the EBM Test instrument was estimated to be a difference of 10 questions between the intervention group and the control group.”
There should be more description here on how the “smallest meaningful difference” was determined. In medical education research, it is a challenge to define a meaningful effect size, mainly because there is a lack of literature on how great a learning gain would translate into an impact in students’ future practice. As such, any threshold would seem arbitrary unless it is supported by a process which involves careful consideration from a panel of teachers who were closely involved in student teaching and had reasonable expectation on student performance. Was this the case? If not, this should be acknowledged as a limitation of the study.

✔ We agree with the reviewer that sample size estimation should use as much solid information as possible, in order for the estimation to be accurate and scientifically based. We reviewed the available papers that have used a controlled trial design in EBM education and didn’t find a consistent approach or gold standard for this measure. For example Argimon-Pallas used an instrument with a maximum possible score of 212 points, and described it succinctly like this: "The sample size was based upon the ability to detect a minimal important difference before and after of 10 points on the test (standard deviation = 20) with 90% statistical power and two side alpha level of 0.05", which means a 5% percent difference in their instrument score (Argimon-Pallàs JM et al. Effectiveness of a short-course in improving knowledge and skills on evidence-based practice. BMC Fam Pract 2011;12:64. www.biomedcentral.com/1471-2296/12/64). In several EBM education papers the differences in knowledge have ranged from almost none to about 60% increase. For our a priori sample size estimation we reviewed carefully the literature, analyzed our EBM test pilot data (mean difference, standard
deviation), and the data from previous applications of Taylor's instrument in our educational environment. The course Faculty involved in this analysis were very closely involved with the educational process and had experience in the area, and we agreed that it would be reasonable to find a 10% difference in the EBM knowledge test, which would imply a moderate to large effect size with Cohen's $d$ (>0.5) based on our pilot data and educational experience.

Results, EBM summative test-Knowledge score
“This scores were similar between the non-EBM groups, the M4 group had a mean score of 30.6 +/- 5.6, and the M5 non-EBM group had 32.6 +/-6.6 (p=0.18). The M5 EBM group had a test score of 58.5 +/-7.9, higher than the M4 and M5 non-EBM groups.”
Converting to percentages, there was a 26% difference between the intervention group and the control group, which convinced me that the EBM training produced a substantial learning gain immediately post-intervention. However, year-six students only gained less than 9% compared to the pre-intervention group, and that was measured midway through their year-six training. It was worrying whether there would be any material learning gain at the end of their six-year course. I would like to see the authors putting forward their thoughts on this in discussion.

✓ As discussed previously, comparing the “gain” in knowledge of M6 students with students in the control and pre-intervention groups and speculating on the difference between the M5-EBM group and M6 students, can be open to several interpretations. It is difficult to state with certainty that the knowledge level of M6 students is somehow inadequate or that a “forgetting curve” slope could be inferred from the data. We added some thoughts about the limitations of our data and the possible explanations of the differences in the Discussion. It is important to note however, that the 9% difference in knowledge when translated to effect sizes with Cohen's d is still large (1.2 for the M6 vs. M4 and 0.93 for the M6 vs. M5 non-EBM groups).

Discussion
Overall, I think while the section is very comprehensive, it may flow better if the authors focus on discussing the findings of this study and not elaborate too much on general issues in teaching and application of EBM. The authors may wish to consider removing paragraphs 1, 3 and the last paragraph all together and trimming the other paragraphs to omit excessive commentaries on EBM in general.

✓ We removed the indicated paragraphs and modified the Discussion following the Reviewer’s recommendations.

Paragraph 2, “The educational effect of the course appears to decrease with time, but lasts at least a year after the educational intervention compared to the control groups.”
This statement is misleading for two reasons. First, from my understanding, not all year-six students received EBM training a year previously, as some received the training six months previously (assuming the curriculum remained unchanged). Next, “lasting at least a year” was an over-statement as the performance of year-six students did not suggest that they had kept a substantial portion of the knowledge learnt in year five. I think the second half of the sentence should be replaced by something more neutral, such as “…and year-six students appeared to still retain some EBM knowledge learnt in year five”.

Also, what did the findings mean in relation to the expected outcomes of EBM training in the authors' institutions?

We agree with the reviewer’s observations and modified the quoted paragraph to clarify the statements. We also added comments in the Discussion about the implications of the study for our institution.

Paragraph 9, “The effect size immediately after the course in the critical appraisal skills score was higher than 1.0, which reveals a large effect that is maintained one year later (Table 6).”
Paragraph 13, “These increases are educationally significant, with effect sizes ranging from 0.88 to 4.6, which reveal an important effect of the educational intervention.”

I think these statements should be omitted as they are difficult to interpret and contain misleading statement on knowledge retention and educational significance, as commented earlier.

We modified the paragraphs to be more objective regarding the data presented, and eliminated the implied temporal cause-and-effect relationships between the main comparison groups (M5 vs. M5 non-EBM) and the separate simultaneous groups (M4 and M6). We maintain however, that the use of an effect size index measured with Cohen’s d to define the magnitude of the difference between groups is reasonable in our manuscript. There are several papers by recognized experts in the field that recommend the use of an effect size index like Cohen’s d when reporting educational research studies, for numerous reasons described in detail in the following references:


Fraenkel et al go as far as recommending that an effect size of 0.5 or larger should be considered as important, taking into account the whole picture of the educational intervention and its context. We agree with the reviewer that educational significance and educational importance are terms that are ambiguous and could be misleading, so we used more neutral adjectives to describe our findings.
Paragraph 15, “This apparent decrease in knowledge needs to be moderated by the fact that our one year post-course measure was done in a different group of students, not the randomized 5th year class, so it may not represent a true measure of knowledge decay. On the other hand, the knowledge and attitude scores were significantly higher than the pre-EBM course scores in M4 and M5 students, which suggest that the course had a long-term effect in this setting.”

In the statements, the authors are not consistent in their arguments. They chose to attribute the positive learning gain to their EBM course and lower scores in year-six students to possible differences in student characteristics. The truth is, one would not know whether the differences in scores among the groups assessed were due to the effects of EBM training or to the groups themselves, as there were no follow-up assessments performed in this study. Rather than making postulation either way, I would suggest that the authors acknowledge the uncertainties and include them as a limitation of the study.

✓ We modified the statements to be consistent in our arguments, and discussed the uncertainties of the study in the limitations section of the Discussion.

• Minor Essential Revisions (such as missing labels on figures, or the wrong use of a term, which the author can be trusted to correct)

Abstract, RESULTS: “M5 EBM=48, M5 non-EBM=47, M4=87, and M6=107.”

Expand each term in full when used for the first time.

✓ Done.

Methods, outcomes and instrumentation
“The knowledge portion of the questionnaire includes six multiple-choice questions”

Should it be “multiple true-false”?

✓ We agree and changed the description.

“incorrect responses are negatively scored (-1) to try to persuade the students from guessing”

Should it be “try to prevent” or “dissuade” instead of "persuade"?

✓ Changed to “prevent”.

Results

In general, the report contains excessive technical details and the authors should omit this in favour of a more descriptive presentation. For instance, statistical tests employed need not be mentioned in the results.

✓ We decreased the amount of technical details in the Methods and Results sections, using more descriptive narrative.

Table 6: comparison between cohorts
It would be good to also include “M6 vs M5 EBM” and “M6 student who received EBM training in the first semester versus second semester”.

✓ We added the M6 vs. M5 EBM comparison, and for reasons stated above we were not able to compare M6 first semester vs. M6 second semester (the students are mixed each year).

- Discretionary Revisions (which are recommendations for improvement but which the author can choose to ignore)

Abstract, methods: “EBM attitudes, knowledge and self-reported skills were measured using Taylor’s questionnaire and a summative objective test.”

The authors may wish to add the following description after the term “summative objective test”: “, which comprised of a 100-item multiple-choice test”.

✓ Done.

Methods, intervention

“The course faculty were six professors with training in EBM teaching”

What was the teachers’ background? Were they involved in clinical service, and were they involved in the training of year four and year six students as well? I think how the EBM training programme was clinically-linked should be described in greater detail in the methods, including specific methods of delivery for each domain. A convenient way of including this is by adding an extra column in Table 1 on how teaching of each subject was delivered, i.e. whether via lecture, small-group tutorial, bedside session or journal club etc.

✓ The course faculty were six professors with training in EBM teaching, all of them board-certified physicians with clinical practice, one of them with postgraduate degree in health professions education. All of them had more than six years of experience teaching EBM to undergraduate medical students, residents of several specialties, and providing faculty development EBM workshops to teachers of several specialties. The lead author had specific training in EBM-teaching at McMaster Faculty of the Health Sciences in Canada. They were not involved in the training of the M4 group, but they collaborated in the training of the M6 group.

✓ The EBM program was clinically linked with the internship program and the residency programs in the hospital, through the medical school curricular committee and the University Postgraduate Studies Division. We used several teaching methods, as outlined in the paper, but each course session used varying combinations of them (for example: one session had a brief interactive lecture, followed by pair-discussion and them small group problem solving), so adding a column to Table 1 to describe in detail the different teaching methods for each of the 14 sessions would probably add unneeded complexity to the table.
Results, Instructional sensitivity and reliability.
I don’t think these needs to be described in full in the results. These technical details can be appended, and only a brief mention is sufficient at the beginning of the results, with a link to the appendix.

✓ We modified and limited the description of these technical details in the Methods and Results sections.

Discussion
Paragraph 16, “As with any implementation of a new course in a medical school, there was an intense interest from the course director and instructors to develop and implement as effective an educational intervention as possible, so there could be a tendency for confirmation bias. This bias would be expected in this type of educational intervention experimental study, where it is impossible to blind either the instructors or the students to the educational intervention.”
One other possible reason that could lead to bias would be the Hawthorne effect, in which students in the intervention group were aware that they were being evaluated on the effectiveness of their current training programme, unlike M6 students who were not “watched” as closely when the EBM training was administered in the previous year. The authors may wish to incorporate this in their discussion.

✓ We agree with the reviewer and added the Hawthorne effect comment in the Discussion.

Paragraph 17, “...it provides evidence that the course can impact knowledge, attitudes, critical appraisal confidence, and self-reported skills and behaviors about EBM and its related concepts, in a statistically and educationally significant manner.”
As in my previous comments, I think describing the degree of learning gain from year five to year six as “educationally significant” is an overstatement. A more cautionary note on possible degree of learning gain that would remain at the end of the students training and in their clinical practice would be more appropriate here.

✓ We modified the text to follow the Reviewer’s recommendations.
**Reviewer 3. Dragan Ilic**

**Reviewer’s report:**
The manuscript offers new information, in the form of a randomised controlled trial, on the effectiveness of a formal EBM teaching program on competency in knowledge, skills and attitudes. The following suggestions may further improve the manuscript:

**Major Compulsory Revisions**

**Abstract - Background**
- In stating your objectives of the studies, please specify what is meant by assessing the 'learning' of EBM (e.g. knowledge, skills etc...)

✓ Done.

**Manuscript Introduction**
- Your introduction is very brief. Please use this heading to further explain to readers what the principles of EBM are, the findings of previous research, the current information gaps and how your proposed RCT will add to the current body of literature.

✓ We modified the Introduction following the Reviewer’s recommendations.

**Methods**
- Why have you included a group of 4th year students. I understand that this group has not been exposed to the EBM teaching, but is this group not the same as the 5th year non-exposed EBM group?

✓ We decided to add quasi-experimental static-groups comparisons besides the randomized trial, which included a more junior group of 4th year students (M4 non-EBM) and a more senior group in 6th year that had the EBM course in the previous year (M6 EBM). This was done to acquire more information from our population in concurrent groups and increase the validity of the study. The M4 is similar to the M5 non-EBM regarding the non-EBM intervention status, but the 4th year students have recently completed courses on Medical Informatics, Statistics, Research Methodology and Epidemiology. We wanted to address the issues of history, maturation and contamination threats to validity that may occur in educational studies, and decided to assess this group as a static group in the overall study design.

- Why was the Taylor questionnaire used, when there are instruments such as the Fresno, Berlin, EBBS tools that also assess knowledge, skills and attitudes?

✓ Taylor's questionnaire is a published instrument designed to evaluate the effectiveness of evidence-based medicine teaching, which has validity and reliability evidence, and includes items to assess self-reported skills and behaviors, knowledge and attitudes regarding evidence-based clinical practice, which were our study selected outcomes. A systematic review assessed the
currently available instruments to assess EBM learning, with detailed information about their development, format, learner levels, assessment domains, feasibility, reliability and validity (Shaneyfelt T, Baum KD, Bell D, Feldstein D, Houston TK, Kaatz S, Whelan C, Green M: Instruments for evaluating education in evidence-based practice. A systematic review. JAMA 2006, 296:1116-1127). Taylor's instrument was categorized in this review as a level 1 instrument, which means that it has reasonable psychometric properties, has been evaluated for validity from at least three sources of evidence, and is recommended for use in the summative evaluation of individual trainees. We have experience with Taylor's instrument in our educational practice and have used it in previous papers, developing evidence of validity for its use in our environment (including the translation to Spanish and back-translation process), so we preferred this tool (the Fresno test is more complex to score and it would have been unfeasible for our almost three hundred students, and the Berlin test measures mostly knowledge and doesn't assess EBM attitudes).

**Results**
- The information under the 'subjects' heading can be removed, as it repeats much of the information presented in Figure 1. The remaining demographic information may be condensed and presented in a table.

✓ We removed the information under the subjects' heading, following the Reviewer's recommendation. The demographic information was summarized.

- The second paragraph under 'Taylor's questionnaire' page 14 seems like it could be transferred to the methods section.

✓ The mentioned paragraph was moved to the Methods section.

- Much of the descriptive information in the results is repeated in the various tables and figures presented.

✓ We decreased the amount of descriptive information, to avoid repetition of the data.

**Discussion**
- Much of the previous literature cited centres on studies published from 1998-2004. I would suggest that there are more recent studies that could be referenced.

✓ We added more recent and relevant references to the Discussion and references list.

- How do your results relate to your objectives? For example, I would expect that if you teach EBM to one group of students and not the other, that there would be a difference in their knowledge, skills and attitudes. So, is there a need for such high quality evidence?
✓ We modified the sections of the paper to be clearer on the objectives of the study. We agree that comparing an educational intervention to no intervention usually results in positive findings, but we argue that it is important to document the effectiveness of our educational interventions as well as its magnitude, as part of the justification research needed in medical education. Some previous reviews have found mixed results (nil or very small effects) with EBM educational interventions (Norman GR, Shannon SI: Effectiveness of instruction in critical appraisal (evidence-based medicine) skills: a critical appraisal. *CMAJ* 1998, 158:177-81), so it is important to document with high-quality evidence that EBM courses can be effective, and add to the body of literature the characteristics of successful interventions.

✓ There has been discussion in the medical education literature about the need for randomized controlled trials and meta-analysis in the field, for example Cook et al identified more than 600 studies comparing simulation training with no intervention that reached the same conclusion ("if you teach people using simulation, they will learn"), and that certainly is "research overkill". This is not the case in EBM educational research, as far as we are aware (Cook DA. Randomized controlled trials and meta-analysis in medical education: what role do they play? *Med Teach* 2012;34(6):468-73.)

- The discussion is excessive in its current form (9 pages). Can you please condense and highlight how your study's results differ/support previous studies (be it RCT or non-RCT evidence), what are the implications of your results to the broader medical education community and impact on future research in this area.

✓ We decreased the size of the Discussion section, and focused it following the Reviewer's recommendations.