Author's response to reviews

Title: Evaluation of a Biomedical Informatics Course for Medical Students: A Pre-Posttest Study at UNAM Faculty of Medicine in Mexico.

Authors:

Melchor Sánchez-Mendiola (melchorsm@gmail.com)  
Adrián I Martínez-Franco (adrianfile@gmail.com)  
Marlette Lobato-Valverde (marlettaeva@gmail.com)  
Fabián Fernández-Saldivar (fab@ieee.org)  
Tania Vives-Varela (taniavives@yahoo.com.mx)  
Adrián Martínez-González (adrianmartinez38@gmail.com)

Version: 2  
Date: 3 February 2015

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

**MS:** 1330209611435812

**Title:** Evaluation of a Biomedical Informatics Course for Medical Students: A Pre-Posttest Study at UNAM Faculty of Medicine in Mexico.

**Authors:** Melchor Sánchez-Mendiola, Adrián I Martínez-Franco, Marlette Lobato-Valverde, Fabián Fernández-Saldivar, Tania Vives-Varela and Adrián Martínez-González.

**Version:** 2

**Date:** February 2, 2015

**Editor, BMC Medical Education**

Dear Sirs,

We are submitting the revised version of the manuscript “Evaluation of a Biomedical Informatics Course for Medical Students: A Pre-Posttest Study at UNAM Faculty of Medicine in Mexico”, 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 profound appreciation to the reviewers, whose insightful comments and recommendations contributed to improving 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 (Reviewers’ comments in italics)

Associate Editor:

Major revisions:

Methods: The authors need to show more explicitly the psychometric characteristics and the blueprint of the pre/post test is equivalent to those of the knowledge component of the summative test. Otherwise, one could argue that the pre/post test was pitched at a lower level than required, and hence getting higher marks at the post test does not really mean much.

✓ The blueprint and test specifications of the pre-post test were the same as those of the BMI courses’ summative examinations, with the difference that the pre-post test had a lower number of items (the pre-post test had 36 items, and the summative end-of-course exams had 60 items). Our study assessment instrument and the summative exams done by the BMI Department covered the same themes and subthemes in equivalent content proportions, although we decided to include fewer items in our study test so we didn’t end up with an instrument that was too long and occupied more students’ time than strictly necessary.

✓ Regarding the psychometric aspects of our instruments, we posit the following considerations: we used the software Iteman for psychometric analysis, which uses the Classical Measurement Theory (CMT)/Classical Test Theory (CTT) conceptual framework, for both pre-post test and summative examinations, so the same psychometric analysis program was utilized in all tests. The reliability of our pre-post tests measured with Cronbach’s alpha ranged from 0.5 to 0.7, and the reliability of the summative exams from 0.80 to 0.90. The lower reliability of the pre-post test is acceptable and reasonable for a non-summative diagnostic test.

✓ On the other hand, the pre-post test was not targeted to a lower knowledge level than the summative exams, since the p-value (difficulty index in psychometric parlance, which represents the percentage of items answered correctly in the test; the higher the p-value the easier the test, the lower the value the harder the test) of the diagnostic post-test was about 0.58 in the BMI-1 courses and 0.42-0.44 in the BMI-2 courses, whereas the difficulty index of the summative tests fluctuated around 0.67. As mentioned in the Discussion, students usually do not put the same intensity of effort in answering a diagnostic/formative test vs. a summative exam, which could contribute to explain the higher scores in the summative exams. Nonetheless, the pre to post increase in knowledge with the same instrument suggests that learning occurred, with the caveats mentioned in the Discussion.

✓ We added more data in the Methods and Results section to address these issues.
As suggested by one of the reviewers, the authors need to state whether the pre-test and post-tests were the same.

✓ The pre and post-test instruments were the same. We added some clarification regarding this issue and, as the reviewers suggest, some comments about its implications in the Discussion.

Results: Student survey results need to be appropriately tabulated as a separate table. Otherwise readers may not know the questions/items within this survey.

✓ Agree. We added the separate tables so readers are aware of the survey components and its specific proportions of answers.

Discussion: Delete the section starting from the beginning of “Discussion” (page 11, line 5) up to the paragraph starting with “Silverman” (page 12, end of line 17). The purpose of “Discussion” is to discuss the results of the study in the light of the results of other studies. This section, hence, is not appropriate to be included under “Discussion”.

✓ We modified the Discussion section addressing this concern.

Similarly, the results of the Silverman study need to be compared with those of this study. Instead I could find only the methodology of the Silverman study.

✓ Agree. We added a comparison of Silverman’s results with our study and included it in the Discussion.

In the absence of studies on BMI courses you could compare your results with those of studies that have used the same methods to evaluate educational courses.

✓ We added some comments about this issue in the Discussion.

Conclusions: Omit the section starting from the sentence “There is a clear need for medical schools to teach core BMI competencies”; i.e. omit the section starting from line 22 on page 17 downwards until the end of this paragraph on page 18. This section does not address any of the research objectives, and hence these “conclusions” (if they can be called conclusions) are not supported by the study results.

✓ Done.

Minor revisions: Page 7, Line 20: should read as: “.a 70-item”

✓ Corrected.

Page 16, line 1: should be: “.external validity of the findings to...”

✓ Corrected
Reviewer 1: Alvaro Margolis

Reviewer’s report:
The article is of interest in the field of biomedical informatics education, and is well-written and thoroughly referenced. In general terms, I would recommend to put this evaluation research more into context, as described below.
- Minor Essential Revisions
- In Background, since the importance of the discipline in the medical curriculum is closely linked to the Country's current and future needs, please expand about the Mexican context regarding - for example - the electronic health record and telemedicine adoption, and what teaching the discipline would add.

✓ Done. We added some information about the Mexican context in the Background section.

- In Background, please note that the IMIA recommendations on BMI education in health professionals’ careers is intended at the level of users, as opposed to the graduate training in the field, seeking for experts. In this context, please expand on what can be obtained through formal BMI education at the user level, as opposed to daily use of computer tools by young medical students.

✓ Done. We added some comments on this issue in the Background section.

- In Methods, Setting. Even though the design of the course has been published elsewhere, there needs to be a minimal description of the program (e.g., topics addressed, methodology used in the context of large audiences, why so early in the career, and so on).

✓ Done. We added a brief description of the programs in the Methods section.

- In Methods, last paragraph of "Main outcomes and instrumentation": please explain if the anonymous survey was implemented through the online platform, and how was it made anonymous.

✓ Done. We added information regarding the survey implementation and anonymization process in Methods. We modified the paragraph to describe more accurately the method. Briefly, the students answered the course evaluation questionnaire after the course last session, at the same time that they filled the teachers’ evaluation questionnaire forms. The filled out the forms in the Biomedical Informatics’ Department web page in the Moodle platform, but after the information was collected it was summarized without identifying data by personnel not involved with the study. The investigators had no way of knowing which student produced which course or teacher evaluation data.

- In Discussion, consider discussing the pros and cons of teaching this set of knowledge and skills early in the medical career, rather than later and in an integrated fashion with the clinical disciplines.
Done. We added some comments about this topic in the Discussion.

- In Discussion, consider analyzing the fact that the satisfaction survey may have been biased, since the survey seems to have been implemented through the same online platform where students took the course. Therefore, fear of expressing criticisms may have been present. This may also explain the fact of having such a high response rate in a voluntary satisfaction survey.

As mentioned above, the course evaluation questionnaire was applied after the last course session, together with the teachers’ evaluation questionnaire. With surveys of this type it’s difficult to be absolutely sure there are no biases in the responses. We based the format and implementation of our courses’ and teachers’ evaluation questionnaires on a long experience in these areas in our University. We have developed and published these evaluation instruments in a rigorous academic manner and have shown that there is reasonable evidence of their validity and reliability (e.g.: Flores-Hernández F, Martínez-González A, Sánchez-Mendiola M, García-Cabrero B, Reidl LM. 2011. Modelo de competencia docente del profesor de medicina en la UNAM. RELIEVE 2011; 17(2):3.http://www.uv.es/RELIEVE/v17n2/RELIEVEv17n2_3.htm). Furthermore, students’ in our setting are used to filling out these instruments and the general perception is that the information collected is for their benefit and our Institution’s. We have used several course evaluation formats (paper and online) and have had response rates above 80%. The individuals in charge of collecting the course and teacher evaluations anonymize the resulting data, so the risk of retaliation to the students is minimal. On the other hand, we have an open-ended portion of the questionnaire where the students can write comments and suggestions in free-text, and have had some very strong and brutal criticism from several students, which we wouldn’t have if they were afraid of reprisal. We recognize, however, that we haven’t done qualitative or mixed-method studies to explore the reasons and motivations for the students’ high response rates and feelings toward this type of questionnaires. We included some comments in the paper about this issue, so the readers are aware of the potential of bias.
**Reviewer 2.** Guillermo de la Calle

**Reviewer’s report:**
This research article presents the evaluation of a Biomedical Informatics (BMI) Course carried out at UNAM Faculty of Medicine in Mexico, for improving the knowledge in BMI of students. The focus of the article is well defined by the authors, and the methods detailed for the evaluation of the new curricular reform to assess knowledge and satisfaction of students are appropriate. The article is well-organized and clear. It includes an extensive and great discussion section, providing clearly the achievements and limitations of the study. Both title and abstract are also appropriate, summarizing the content of the article.

**Discretionary Revisions**
In the discussion section, authors remark a limitation in the approach followed in the study. This limitation is about the absence of a control group to compare fairly the results. Authors argue that because of ethical and logistical issues, this kind of strict experimental design was not possible but different measures to prevent results from threats were considered. In my opinion, authors have made a good job. I would like to make just a recommendation in order to improve the assessment of results. Maybe it would be a good idea to use as control group the last (or last two) group of students of the old curriculum, that did not receive any training in BMI. Or even pass the same evaluation questionnaires to students from other faculties. In both cases, the results of the evaluation would be fairer. Did authors consider these possibilities?

✓ Agree with the reviewer. Indeed we considered these possibilities, but for ethical and logistical reasons we could not include the students that took the previous curriculum, and to attempt to use students from other universities would have been extremely complicated due to the complexities of multi-school Institutional Review Board approvals, plus the political and symbolic implications. At the end of the day, we decided to use only our local population, we feel however that the large simple sizes and the consistency of results in different subsequent student cohorts adds validity to our results. We expect to widen our simple in a different subsequent study.

**Minor Essential Revisions**
*Page 2, line 4: this paper was --> this paper is*
✓ Corrected.

*Page 2, line 11: was --> were*
✓ Corrected.

*Page 15, line 17: ...among others), we... --> ...among other). We...*
✓ Corrected.
**Major Compulsory Revisions**

Authors made a nice statistical analysis of the results, but I miss in the article the statistical justification of the sample size to demonstrate that this is representative enough.

- Since we included the totality of our available population, and due to the large size of our student cohorts, we deemed not necessary to do an *a priori* sample size calculation. We performed *post-hoc* power calculations and for our comparisons the power was >95% (in other words, the possibility of a type 2 or beta error is extremely unlikely), which is reasonable since our simple sizes are large and the differences are considerable.

Several times are mentioned the different tests (satisfaction and knowledge) completed by the students, but no examples of the questions are presented. I think that authors should provide some examples of representative questions used in the tests.

- Agree. We included some examples of the knowledge items, and added two tables of the course evaluation questionnaires’ results, which include the items.

After reading the article, it is not completely clear to me whether the pre/post tests are the same, I mean, whether the questions are the same. If so, probably the results could be a bit biased because the second time, students would be aware about questions and if they are “proactive”, they could look for the solution between the first and the second try, and the results would be better even if they do not receive extra training. Please, could authors explain, clarify and discuss a bit more about it.

- Agree. The pre and the post-test instruments are the same, as commented above in the response to the Associate Editor observations. We clarified this in our revision in Methods and reiterate it in the Discussion. In one-group pre-posttest quasieperimental research there is always this “catch-22” situation: if you use a different test for the pre and post, you have to demonstrate unequivocally that the tests are similar in difficulty, otherwise the inference of difference in achievement between groups cannot be done; and if you use the same test, the “testing threat” to internal validity can introduce bias in the measurement, as the Reviewer mentions. This risk of bias cannot be completely excluded in quasieperimental design. We argue, however, that our sample sizes, the magnitude of the differences, the diagnostic nature of our tests, and the time interval between pre and post-test (one semester) can contribute to attenuate this potential bias.