Author’s response to reviews

Title: The impact of feedback during formative testing on study behaviour and performance of (bio)medical students: a randomised controlled study

Authors:

Dick Thijssen (dick.thijssen@radboudumc.nl)

Maria Hopman (maria.hopman@radboudumc.nl)

Marielle van Wijngaarden (m.vanwijngaarden@radboudumc.nl)

Joos Hoenderop (joost.hoenderop@radboudumc.nl)

rene bindels (rene.bindels@radboudumc.nl)

Thijs Eijsvogels (thijs.eijsvogels@radboudumc.nl)

Version: 2 Date: 06 Feb 2019

Author’s response to reviews:

Response to academic editor and reviewers of the paper

“The impact of feedback during formative testing on study behaviour and performance of (bio)medical students: a randomised controlled study”

(MEED-D-18-00599R1)

Dear Professor Messin,

Thank you for your letter from February 1st 2019, concerning our manuscript entitled "The impact of feedback during formative testing on study behaviour and performance of (bio)medical students: a randomised controlled study".

We really appreciate the opportunity to publish our work in the BMC Medical Education.
We have improved our manuscript, considering the requested amendments. The comments raised by the reviewers have been implemented in the revised manuscript and addressed point-by-point in the enclosed letter. Please find enclosed a red-lined copy of the revised manuscript, as well as a clean version in which we have made all final minor edits.

We thank the referees for their careful criticism and hope that in its present form, the manuscript is acceptable for publication.

Sincerely yours,

Dick Thijssen

---

Referee 1

The study was very well designed, although the Methods section in the manuscript is inexplicably located after the Conclusions section. The Methods section needs to immediately follow the Background section.

We have changed the order of the sections.

It is stated that the educational advisory board of the Radboud university medical centre provided approval for this study. However, no Ethics approval reference number is provided. The Ethics approval reference number needs to be included in the manuscript.
The project was discussed and evaluated by the educational advisory board of the Radboud university medical center. Subsequently, we received ethical approval for this study. However, no reference number was provided.

While it is unsurprising that heavy users of the app performed better in examinations than ‘none-users’ (this term is ungrammatical and should be changed to ‘non-users’ in the manuscript text, tables, figures and figure legends), the main findings of the study are interesting, in that no statistical difference was found in study behaviour or examination performance between the intervention (detailed feedback) and control (correct/incorrect feedback) groups. The authors ascribe the unexpected lack of difference between groups to a ceiling effect. However, if that was indeed the case, one might expect that low-moderate achieving students randomised to the intervention group might have outscored low-moderate achieving students in the control group. Although the authors don't show the analysis, they assert that previous study behaviours did not alter the results of the analysis. As a suggestion, perhaps segregation of the analysis to compare the effect of detailed feedback vs. correct/incorrect feedback on known low-moderate achieving students might help to confirm: a) whether detailed feedback is helpful for study behaviour and/or exam performance in that group; and b) whether a ceiling effect is indeed the most likely cause of the lack of difference in exam scores between the intervention and control groups.

We highly appreciate these useful comments from this reviewer. However, we feel uncomfortable performing such additional post-hoc analysis given the relatively small group size (~34% of the entire group) and, therefore, minimal power. Although these questions are interesting and important (and guided by our results), the study was not designed and appropriately powered to answer this question. Therefore, we prefer to refrain from further analysis regarding this matter. We have highlighted the importance to better understand the ceiling effect, as well as the potential that feedback may be useful in the moderate-users (P12).

The authors have otherwise drawn appropriate conclusions from the data obtained and have carefully stated the limitations of this study, including potential contamination of the control group by obtaining access to the detailed feedback intended for the intervention group.

Thank you for your constructive comments.
Minor comments

Page 4 Line 3: "during incorrect answers" should be "in response to incorrect answers";
We have changed this.

Page 5 Line 8" "Students scored a 6.6 +/- 1.1" should be "Students achieved a mean score of 6.6 +/- 1.1";
We have changed this.

Page 8 Line 5: "4% higher exam score compared to previous year" should be "4% higher mean exam score compared to the previous year";
We have changed this.

Page 8 Line 6: There is a redundant extra comma following "e.g.";
We have removed the redundant extra comma.

Page 8 Lines 11-12: "the correct answer has demonstrated" should be "the correct answer has been demonstrated".
We have changed this.
Referee 2

I was a little confused over study time - am i right in concluding this was self-assessed?

Study behaviour, including study time, indeed was self-assessed. We have clarified this in our revised manuscript (P6 and P20).

Your first result 'No differences in app-use was found between groups (P=0.15), whereas the intervention group more frequently reviewed information compared to controls (P=0.007)’ seems to refer to the time spent going back and forward in the app. in which case isn't this a change in app-use between the groups? If this is the case i would suggest your findings are important to the growing literature on learning analytics which you don't mention

This reviewer is correct in that the intervention group more frequently reviewed information than the control group. However, this did not impact the study results. Since this latter question was the primary research question, we did not want to overstate this observation. Nonetheless, we agree with the reviewer that this observation is relevant. Therefore, we have shortly described our observation in our discussion (P10-11).

Methods seems to be out of order - would expect before results

We have changed the order of the sections.

A couple of very small grammatical points -

'None-users' would more normally be 'non-users' (though none on it its own is indeed correct so e.g. usage: none, moderate, intensive)

We have changed this throughout our manuscript.

line 24 page 6 in 17% => by 71%

We have changed this.