Reviewer’s report

Title: Characterization of medical students recall of factual knowledge using learning objects and repeated testing in a novel e-learning system

Version: 2 Date: 19 September 2014

Reviewer: Stefan K. Schauber

Reviewer’s report:

Taviera-Gomes and colleagues report a study that illustrates the potential of technology-enhanced learning formats to integrate insights from research on learning in a flexible, dynamic way. More specifically, they integrate principles of spaced (vs. massed) practice and test-enhanced learning that have been found to be beneficial for learning. To this aim, they developed an online platform and conducted an experimental study contrasting the effect of self-assessment vs. self-assessment + studying. Their results suggest that „self-assessment + studying“ were associated with stronger increase in self-rated recall-accuracy than self-assessment alone.

The general topic – applying insights from the learning sciences to medical education using “educational technology” – clearly is of interest but there are several major shortcomings in this manuscript that need to be addressed.

INTRODUCTION

1. The Introduction/Background is too cursory. Indeed, key findings from the literature are mentioned but they are reported without theoretical background, framework, or context. The only justification for considering “spaced-repetition” and “repeated-testing” is that they “emerge” from the literature. The aim of the study (“to demonstrate”) is in a certain way misaligned with the design of the experiment (“self-assessment” vs. “self-assessment + study”).

2. The entire section on “ALERT STUDENT” should be – in my opinion – moved to the methods section because this platform is a tool, an instrument in their study.

METHODS:

3. I found it very hard to follow the METHODS section. It jumps between several pilot studies, study design, content design, and (descriptive) statistics. The dependent variables are sometimes ill-defined and terms that seem to have a special use (e.g. “flashcards”) are introduced without clarification. I can guess how the flashcards looked like but a supplement/figure would be really helpful here. Please consider moving the reports on pilot studies to a supplement. Was the study conducted in a laboratory setting? Is this meant by “in-person session”?

4. The “EVALUATION”-section reports details of the planned statistical analyses.
I would suggest re-naming it into “Statistical Analysis” or something similar – but this is a minor issue. The main issue here is that it lacks detail. I don’t understand why one would be interested in including “flashcard” as a fixed effect into the analyses. Is there a rationale for this? Otherwise it might be more appropriate to account for this source of variation by including it as a random effect in a mixed-effects/”multilevel” model (indeed it is included as a random effect in the G-Study). I doubt that it would change the results drastically but the rationale for the choice is not explicit. The rationale/background of the calculation of G-Coefficients is not explicated (absolute/relative error?). Who should agree on what? And why is this important? Is it really intended to “determine the reliability of the variance”?

5. When estimating variance components it is common to mention which software and procedures were used (ordinary least squares, maximum likelihood, Baysian procedures,….) as results may differ as a function of the estimation algorithms (cf. eg. Crossley et al. Med Educ. 2007 Oct;41(10):926-34.). I would also find it informative to know which software was used for the other calculations.

RESULTS:

6. From the METHODS section I had a mixed impression on the statistical procedure and this is also true for the RESULTS section: On the one hand the authors have conducted a set of rather elaborated analyses. On the other hand only a fraction of the results is reported which is a severe shortcoming. Indeed, there are – despite p-values – literally no results of the ANOVA reported (Df’S, Mean Squares/Sum of Squares, F-values, Eta-Squared, …). Thus, in it’s current form, the manuscript does not adhere to the relevant standards for reporting. Such details are needed to evaluate the soundness of both analysis and data.

7. There are other details that are not reported on a rational. Why are average course grades important in this context? Why are study resources relevant? Why is Figure 1 presented? Why is study duration reported for the experimental group?

DISCUSSION:

8. The DISCUSSION gives a summary of the findings and their relation to the introduction and literature. However, it also reports additional results and even introduces new expectations. Please consider moving such information to the results and the introduction, respectively.

9. Although being an interesting issue - the conclusions regarding affective/motivational influences on learning are, in my opinion, not supported by the data. As far as I understood the manuscript there is no information on such factors available in the current study.

Overall, the strength of the current manuscript is - in my opinion - the aim to design educational technologies based on findings from the literature on cognitive psychology, learning sciences, and educational psychology. However,
there are many “negligent” elements and inconsistencies in this manuscript that warrant consideration. Those shortcomings relate to an inadequate presentation of the relevant literature and a lack of detail/rationale in the statistical analyses.

**Level of interest:** An article whose findings are important to those with closely related research interests

**Quality of written English:** Acceptable

**Statistical review:** No, the manuscript does not need to be seen by a statistician.

**Declaration of competing interests:**

I declare that I have no competing interests