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: 1 September 2014

Reviewer: Ralf Schmidmaier

Reviewer's report:

I think the topics of repeated testing and spaced learning are very relevant and concrete application in curricula is strongly needed. E-learning platforms are a good way for implementation of these important aspects of learning. Novel randomized studies are urgently needed to optimize the use in the curricula and to gain more insights in the theories behind. Therefore I am happy to read this study addressing repeated and spaced testing in an e-learning environment.

Please number your comments and divide them into

- Major Compulsory Revisions

1. It was not easy to understand the research question and the concrete study. I suggest an illustration of the study design and examples of notebooks and flashcards and a screenshot of the study mode and the quiz mode – at least as supplement. If I understood it right participants had learned the learning objectives in the past (Golgi three years ago) during the normal curriculum (are there grades or achievements available?). They were randomized. Both group start in week 1 with a self-assessment session. This means that open ended questions are provided according to the Golgi notebook. I assume the participants had to compute their answer and were after that provided with the correct answer. As I understood only a part of the questions was displayed and at S1 and S2 different questions were displayed. Why was accuracy not assessed objectively by counting the correct answers? The students rated their accuracy by using a likert scale. Was a feedback given? So far I understood they did not have time and chance to restudy. Participants came back after one week and quizzed again in the standard group but studied in the experimental group and immediately after that quizzed as well and graded their achievement. S1 was repeated after another weeks (S2). I do not really understand the design. The design investigates whether self assessed achievement in a specific topic is better after studying the contents immediately before the assessment. What is the pedagogical theory behind? In my opinion this neither investigates retrieval of knowledge to foster learning nor spaced repetition. What is the concrete research question, which is the dependent and which is the independent variable, what is the intervention in a more abstract view (based on the scientific literature)?
recall accuracy is not well defined in the paper. What does the number 0.79 or 1.87 tell us? Were the questions the same in S0, S1 and S2? I suggest to provide the contents and the questions as supplement.

2. The experimental group has a longer time on task and additionally is provided with the answers of the assessment questions immediately before the assessment and therefore performs better than the group that was never confronted with the learning objectives in a controlled manner. Please explain more specifically what novel insights this result reveals and which evidence based conclusions we can draw for daily instruction from this result?

3. The authors again and again refer to the published research of test-enhanced learning. The authors should more precisely describe the influence of this theoretical background on the study design. Do they focus on knowledge acquisition or knowledge retention? As far as I understand the study investigates learning by repeated testing with feedback versus learning by repeated studying followed by repeated testing with feedback. The most published studies were focused on retrieval of knowledge versus acquisition of knowledge for knowledge retention. This does not compare to the presented study. Therefore I suggest to provide a more specific description of the theoretical background and a more precise description of the lack of knowledge. Based on this I would be happy about a operationalized research question (see 1) and an illustration of the study design with concrete examples (see 1).

4. Figure 4 needs to be improved. The legend should clearly explain what the graph shows.

Without the above mentioned information it is not possible to adequately comment on the results and the discussion section at this time point.

With the imbalance of time on task that is systematically induced by the design it seems almost impossible to interpret the results. This is really a major concern. The other named limitations are less interfering.

**Level of interest:** An article of limited interest

**Quality of written English:** Acceptable

**Statistical review:** Yes, but I do not feel adequately qualified to assess the statistics.

**Declaration of competing interests:**

'I declare that I have no competing interests'