**Reviewer's report**

**Title:** Learning with case-based worked examples: Comparing the effects of additionally providing self-explanations and making or studying concept maps on physiotherapy intervention knowledge

**Version:** 1  
**Date:** 7 October 2014

**Reviewer:** Veronika Kopp

**Reviewer's report:**

Discretionary Revisions

The authors describe an empirical study that investigates the effects of three different scaffolds that are added to a digital, case-based learning environment designed to foster physiotherapy students' acquisition of conceptual knowledge and problem-solving skills (transfer). After having studied a worked example/completion example pair (overall, three such pairs were given), students in the first experimental condition were prompted to self-explain the solution steps of the examples (self-explanation condition). Students in the second experimental condition received an incomplete concept map and were instructed to complete it after each example pair (concept mapping condition). Students in the third condition received an already completed concept map that represented the main concepts and relation between concepts that were covered in the examples (model concept map study condition). Dependent variables were the mental effort students reported during the process and the pre- and posttest (i.e., when being exposed to the worked examples/completion problems), the conceptual knowledge students and the problem solving skills (i.e., transfer performance) they gained from pre to post test. Results indicated that conceptual knowledge acquisition was best fostered in the study concept maps condition (whereas the other two conditions were less effective, but did not differ from each other), whereas problem-solving skills were best acquired through self-explaining, whereas students from the two concept mapping conditions showed significantly lower performance (but did not differ from each other). Data from the mental effort instrument indicated that mental effort was lowest in the concept map study condition during the learning process, whereas during post test it was lowest in the self-explanation condition and generally negatively correlated with post test performance. Overall, the authors argue that being exposed to self-explanation prompts is likely to promote the construction and automatization of schemata (i.e. promote understanding) that could then easily be used during post test. Concept mapping, in contrast, might have provided learners with an opportunity to incorporate in-depth concepts into their schemas (i.e., promote information chunking), without necessarily understanding the relations between the different concepts that were presented/created in the concept maps.

Overall, the manuscript meets high scientific standards. The theoretical part
displays a thorough and concise line of argumentation that nicely prepares the research questions and hypotheses. The method of the empirical study is sound and follows the typical experimental paradigm (including the instruments that were selected) that is often used in research on the effectiveness of (different variants of) worked examples. The results of the empirical study are presented in a reasonable way and the discussion of the findings is compelling and well-grounded in theoretical arguments that were previously presented and developed in the theoretical part.

Having said this, my following criticisms are not at all meant as reasons for a rejection of the manuscript. Rather, they should be seen as constructive criticism that might strengthen some of the parts of the paper:

1. In the theoretical section, I felt that a bit more justification could be given for why case-based learning is supposed to be a helpful instructional approach in medical/physiotherapy education. Especially from the cognitive load theory perspective that the authors are taking, it seems a bit odd to promote an approach that proponents of cognitive load theory would probably criticize for evoking high levels of extraneous load (due to the fact that students would have to find solutions for the presented cases themselves – at least if case-based learning would be unstructured). Having said this, the authors might want to consider extending the arguments in favor of case-based learning that now only cover the bottom part of p. 4 and perhaps even devote a whole section to case-based learning that provides arguments for the approach. Arguments for case-based learning can be derived from work by Kolodner or Schank.

2. Although the presentation of cognitive load theory is very concise, it might be a bit more differentiated. Recently, Ton de Jong and Slava Kalyuga have expressed criticism regarding various aspects of CLT (e.g., regarding the question whether there are really three or perhaps only two types of CL). Also, the argument that “worked examples facilitate learning be decreasing cognitive load, and increasing germane load: all available working memory capacity can be dedicated to schema construction” might be complemented with explicitly stating that however it is not guaranteed that students will automatically invest working memory capacity in learning processes (germane load). Actually, making this open question explicit would nicely fit to why you added scaffolds to the worked examples. I.e. you could argue that to increase the likelihood that worked examples work (i.e. that students really use the opportunity of a decrease in extraneous load to engage in activities that “represent” germane load), further scaffolds/reflection prompts are necessary.

3. One aspect I really like about the theoretical part is that it presents quite differentiated arguments for why self-explanation prompts would work (and for what; see top of p. 8: “The key aspect of the self-explanation effect is that learners attempt to revise their understanding and make sense of the material, even if they are unsuccessful in articulating a correct explanation.”), and why concept mapping might work (and for what; see p. 9: “Concept map study may (...) foster deep learning by encouraging learners to judge the importance of the concepts and process their hierarchical organization and the relationships
between them."). Interestingly, these are not arguments that come from CLT, but rather from other cognitive approaches. Yet, in the empirical study, only a CLT-related measure is taken as a process measure. Thus, the lack of further process data that would for example more directly test the assumptions just mentioned (e.g., think-aloud data, eye-tracking data etc.) might be highlighted as a limitation of the study and a desiderate for future studies.

4. Another criticism that might be highlighted in the limitation section refers to the rather low external validity of the study. Based on Chi’s (2009) ICAP model, for example, one might hypothesize that having students create concept maps will be more effective (especially regarding transfer) than having them self-explain or study concept maps, because the former is a more constructive learning activity than the latter which – according to the data Chi has collected from other studies – is more conducive for learning than what she calls “active” activities (a category into which both self-explaining and studying concept maps might fall into).

However, this will probably only hold in settings where students do not have time pressure and can devote as much energy to the task as they want. In your study, however, the time for the single steps was limited, which may have been a major reason for why concept mapping has led to less positive results on transfer than self-explaining.

5. In the methods section I was a bit surprised by the fact that all students received training in concept mapping, but that there was no training in how to give self-explanations. Through providing learners with a concept mapping training, students from the two experimental conditions that creating resp. studied concept maps might have been advantaged over the students from the self-explanation condition, since the former were already familiar with the scaffold they would receive later, while the latter were not. The authors should justify this decision.

6. Also in the methods section, please explain whether students were forced to write something on the self-explanation sheets or whether they were free to do so. In both cases, I would be interested in how extensively the self-explanation sheets were used (i.e., how much students wrote).

7. In the concept mapping condition, I was surprised that students received incomplete maps instead of blank sheets of paper on which they would be asked to create concept maps from scratch. I understand that this probably has to do with the fact that students in the concept mapping condition would not be able to complete this task within the 15 minutes that were assigned to it (by the way: did students in all conditions use the 15 minutes in a comparable way, i.e. was actual time on task comparable across conditions?). Then, however, the question is whether the label “concept mapping condition” actually transports what happened correctly, or whether perhaps “completing concept maps condition” or “incomplete concept maps condition” or something like this would be more accurate.

8. One problem that is inherent to the comparison of prompting for self-explanations on the one hand and producing/studying concept maps on the other is that students in the two concept mapping conditions were simply presented with more learning material (the information that was presented on the
incomplete/complete concept maps) than students from the self-explanation condition. That way, higher post test performance of the study condition on conceptual knowledge may simply be caused by pure memorization effects. In other words, memorization might be an alternative explanation to the nice arguments you provided in the theoretical part on what cognitive processes may lead to the result that concept mapping should work (see point 3 in my review). Also from this point of view, it might be interesting to look for further process measures to find out what processes actually were responsible for the results. If no such process measures were used, the authors should at least speak to this problem in the limitations section.

9. As most research from the cognitive load paradigm, this study only uses a general measure of mental effort which only gives the opportunity to interpret post hoc on what kind of cognitive load has been promoted in the different experimental condition (a criticism that has also been expressed by de Jong). Although using this undifferentiated instrument still seems to fit to the zeitgeist of cognitive load research, I would suggest to critically discuss what other opportunities might be used to measure cognitive load in a way that would allow to differentiate between the three (or two) different load types.

10. Information on (intrarater) reliabilities might be moved from the section “data analysis” to the section on “test tasks”.

11. Minor issue: On p. 17, I would explicitly mention what you used as a covariate in your ANCOVAs (pretest performance).

12. Given the results, one might dare to hypothesize that a combination of concept mapping (esp. studying) and self-explanation prompts would be more effective than each of the two scaffolds alone. This might be stated as a hypothesis that might be investigated in future research.

Overall, I very much enjoyed reading this paper and - in light of my criticisms above – would argue for minor revisions before the manuscript might be accepted for publication in BMC Medical Education.

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