Reviewer’s report

Title: A model of the pre-assessment learning effects of assessment is operational in an undergraduate clinical context

Version: 1 Date: 6 October 2011

Reviewer: Larry Gruppen

Reviewer’s report:

This paper addresses a very real gap in our knowledge of and practice in medical education. As the authors note, assessment is often identified as a driver of learning, but we know little about the details of how this happens. Their effort to build a model to address this gap is valuable and these authors are perfect for the job. Thus I read this paper with great anticipation. Unfortunately, I found serious problems, which are listed below.

Major Compulsory Revisions

1. The major problem I had with the paper was readability. My first reading was in the evening and I considered my weariness to be a likely explanation. However, a re-reading the next morning showed the same problems. There are several things the authors might do to address this.

1a. Their model is presented in only the briefest and sketchiest of terms. Although they cite their other publications on the development of the model, these do not help the reader of this manuscript. I believe the model needs to be described with a couple of clear examples of what the authors mean by such components as "task type," "cues from lecturers," "persistence with learning," and the like. As a reader I was left to try to infer what behaviors and events might be covered by these terms and I wasn't at all sure I was right in my inferences.

1b. I think the model would be much more comprehensible if it were addressed more as part of the discussion than as the introduction. Having the concrete examples (quotations) from the interview results make the model components more concrete and will help the reader understand and evaluate the connections the authors are making to Prochaska’s framework.

1c. as is often the case, it seems that the authors have spent so much time with their model that it seems intuitive and obvious to them, which leads to assumptions in the writing of the manuscript that the reader will see things just as clearly. This is not the case, however, and I would recommend that the authors find a naïve colleague to give feedback on the assumptions that are being made unintentionally in the writing.

2. One of the areas I had to read and re-read before I think I understood it, was the nature of the data collection and sample. As far as I can tell, this study is essentially a re-analysis or a secondary analysis of the same dataset that gave
rise to the original model, except that the original mode focused on the "theory" modules in the curriculum and this paper focuses on the "clinical practice" modules. If so, this modifies my interpretation of the results considerably from what I would take if this study were done on a new set of students and interviews. Whether my inference is correct or not, this part of the paper needs to be clarified.

3. The authors spend much text on the question of validity evidence for their model. This is an important goal and I commend them for this, because it is too often neglected. One of the problems I had, however, was, again, the brevity of their linkages of their model to Prochaska's framework. For a reader who isn't familiar with either the model or Prochaska's framework, the claims made on page 6 are extremely difficult to verify. I am familiar with Prochaska, and I still had trouble. Again, if these were put at the end of the paper, it might be easier to link these abstract claims to some more concrete examples.

3a. Given the complexity of the validity argument, I still found some of the claims difficult to accept on face value. One was the claim that the constructs were potentially falsifiable, but given that I saw no predictions or hypotheses in this study, I cannot see the falsifiability claim. Indeed, a simplistic interpretation of the data in table 2 suggests that several of the model components were not observed in the clinical practice modules - isn't this a falsification? If so, why do the authors not remove that component from their model. What I see instead is an effort to apply the model to a (somewhat) new set of data to see how well it fits - a perfectly reasonable approach, but not really falsification, from my point of view.

3b. The claim that the model is generalizable (bottom of p. 18) is contradicted by the claim on the bottom of p. 20 that the "purpose of the study was not to yield generalizable results..." Please excuse me as a curmudgeon, but I've often heard this disclaimer applied to qualitative studies, but it always strikes me as disingenuous. If the results are truly not generalizable, why publish them - the purpose of which is to disseminate knowledge for use by others? It seems particularly anomalous in a paper that is trying to argue for the generalizability of a model of testing effects.

3c. The whole validity argument for this model should be a bit more restrained from my perspective. I am not compelled by the strength of evidence provided, given that the model is being re-applied to the same students and the same interviews from which it was derived (that is, if I am correctly understanding the design of the study). Much stronger evidence would need to come from an application of this model to a completely separate set of interviews - different students, at least.

Minor Essential Revisions

4. The paragraph at the bottom of page 6, beginning "Maxwell [17,18] argues..." is confusing and should be rewritten. It took multiple readings to understand that you were arguing AGAINST using experiments and that your context did NOT fit
the criteria Maxwell provides for experimental study.

5. Please provide a list of the questions used in the interviews. This is necessary for the reader both to understand the study and interpret the results.

Discretionary Revisions

6. a bit more detail on the contrasting characteristics of the 4 week clinical practice and 4 week theory modules in the South African system would be helpful for those not familiar with this curricular model. These contrasts could highlight the differences in educational and assessment activities and experiences between to the two settings.

All this negativity aside, I think this is important work and I look forward to catching up with the already-published articles not he model. It is something we need to guide the growing efforts to align assessment and teaching/learning. I expect that I'll be using it in some of my own work.

Level of interest: An article of importance in its field

Quality of written English: Acceptable

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

Declaration of competing interests:

I have no competing interests