Reviewer’s report

Title: Voluntary vs. Compulsory Student Evaluation of Clerkships: Effect on Validity and Potential Bias

Version: 0 Date: 08 Aug 2017

Reviewer: Laura Hopson

Reviewer's report:

This is a clever little study which took advantage of what appears to have been a natural experiment (potentially as a change in accreditation environment inferring from the authors' structuring of their introduction) which occurred with a change from voluntary to mandatory student evaluations of clinical clerkships. It seems to fill a discrete knowledge gap and would have utility and applicability for individuals looking to design their own program assessment strategies. It also raises the question of whether we can do representative sampling rather than compelling all students to complete evaluations.

Study is designed as a pre/post intervention with the final group getting a sham question to assess engagement with the evaluation process. The authors looked at issues of validity and bias between the two groups in addition to assessment characteristics. They conclude that compulsory evaluations may not improve reliability and that a deeper dive into the representativeness and quality of responses is required to make decisions on clerkship assessments. Some of their conclusions (authority bias) get overstated based on their actual data and this issue will need to be addressed prior to acceptance.

* Methods well structured although the sham question to assess authority bias appear to have been a late addition to the study.

* Clearly defined characteristics
  
  o Respondents
  
  o Non response bias

* Standard approach of walking systematically through elements of validity evidence being considered.
  
  o Response Process Validity
  
  o Internal Structure
    
    • Consideration of authority bias
Wish the sham question had been included throughout and not just in the final cohort!

* Did a nice job finding comparison data between the groups and responders/nonresponders even though evaluations were anonymous. Within that systematic limitation they have optimized information.

* Why only half the year in the new (compulsory) system? Does the data or conclusions change if we look through the entire year?

* Compulsory system works to increase response rate from an average of 56% to 100% (tied to grade release). In voluntary system, the authors note significant variability of ratings across the clerkships which would imply different strengths/weaknesses and that students are providing meaningful evaluations. Clerkship ratings did not correlate with grades (group not individual) so it is not just the high performers who participate.

* Compulsory response generally did not change clerkship ratings except for OB/gyn clerkship but did result in higher overall scores. Similar presence of comments (this would be a goldmine to explore in a future qualitative study)

* Table 2 is useful and easy to read. Table 1 comes across as very busy and less useful (baseline characteristics from voluntary group). It could potentially be included as just a supplementary table if there are space limitations.

OB/gyn regresses to mean

* I am struggling to figure out Table 3 looking at nonresponse bias. This is also referenced in the abstract and requires clarification. R v. NR table clearly indicates a higher grade for the respondents but the authors minimize this saying that the difference disappears when whole class grade was considered. This will require rephrasing for clarification - are they saying only that throughout the clerkship R v. NR were the same even though individual clerkships looked different? I think there is a potentially significant finding here that is being underplayed that within each clerkship those who perform better are more likely to evaluate the experience even if in aggregate over the year there is no difference. ***This is a significant revision (p11 L8-12 and abstract p2 L17-19)

* I will need to defer to a statistical expert on p11 section 2 / Table 4 dealing with Internal structure validity evidence. Looks reasonable but I may be missing subtleties here in the factor analysis.

* Consequential validity evidence p12 starting on L4. Need to explicitly state that this is the final cohort with the sham question for clarity. This finding is probably underpowered to detect a significant difference and would have been very interesting to have a large sample size.
The intriguing albeit disturbing and probably not entirely unexpected finding is that the sham question in the final cohort demonstrated that students are not putting a great deal of effort into providing meaningful clerkship evaluations when they are compulsory (authority bias).

This comes from a Lebanese medical school following the US medical school training model. These specifications may raise a question about cultural bias and whether the findings can be abstracted to other training environments. However, this is a minor point that should be included in the limitations but overall does not degrade the quality of the work. Relatively small cohort size dictated by need to compare stable evaluations (2014-15 pre and 15-16 post since they indicate that the evaluation form was revised until 2014). Manuscript is well-written and readable.

Line by Line notes:

P2L3 Recommend changing "reliable" to "useful" or similar word in the background section of the abstract. A major aim of the paper is to assess evaluation reliability so would not use it here (see p12 1st paragraph of discussion where authors make this same argument).

P2L13 Change "bogus item" to Sham question

P2 L22-23 Too strong a statement given the limited data. "tendency toward" may make it more accurate

P4 - Introduction is well written and referenced in the first paragraph. Recognizes potential competing sources of bias between voluntary and mandatory methods of program/teaching evaluations.

P12 L 22-23 & P13 L1 Sentence reads awkwardly and could be rephrased for clarity.

P13 L12-13 awkward phrasing "non-response bias is frequently non-significant in a homogeneous target population."

**Are the methods appropriate and well described?**
If not, please specify what is required in your comments to the authors.

Yes

**Does the work include the necessary controls?**
If not, please specify which controls are required in your comments to the authors.

Yes
Are the conclusions drawn adequately supported by the data shown?
If not, please explain in your comments to the authors.

No

Are you able to assess any statistics in the manuscript or would you recommend an additional statistical review?
If an additional statistical review is recommended, please specify what aspects require further assessment in your comments to the editors.

I recommend additional statistical review

Quality of written English
Please indicate the quality of language in the manuscript:

Acceptable

Declaration of competing interests
Please complete a declaration of competing interests, considering the following questions:

1. Have you in the past five years received reimbursements, fees, funding, or salary from an organisation that may in any way gain or lose financially from the publication of this manuscript, either now or in the future?

2. Do you hold any stocks or shares in an organisation that may in any way gain or lose financially from the publication of this manuscript, either now or in the future?

3. Do you hold or are you currently applying for any patents relating to the content of the manuscript?

4. Have you received reimbursements, fees, funding, or salary from an organization that holds or has applied for patents relating to the content of the manuscript?

5. Do you have any other financial competing interests?

6. Do you have any non-financial competing interests in relation to this paper?

If you can answer no to all of the above, write 'I declare that I have no competing interests' below. If your reply is yes to any, please give details below.

I declare that I have no competing interests

I agree to the open peer review policy of the journal. I understand that my name will be included on my report to the authors and, if the manuscript is accepted for publication, my named report including any attachments I upload will be posted on the website along with the authors' responses. I agree for my report to be made available under an Open Access Creative Commons
CC-BY license (http://creativecommons.org/licenses/by/4.0/). I understand that any comments which I do not wish to be included in my named report can be included as confidential comments to the editors, which will not be published.

I agree to the open peer review policy of the journal