Reviewer’s report

Title: Adding to the debate on the numbers of options for MCQs: The case for not being limited to MCQs with three, four or five options

Version: 0 Date: 28 Jun 2019

Reviewer: Jacob Pearce

Reviewer’s report:

Thank you for the chance to review this interesting submission. While I think the topic is interesting and relevant, I have some major concerns about the paper in its current form.

Let me begin by stressing that I do not disagree with the author's overall contention. I think the existing literature supports the notion that content relevance should be paramount when deciding on how many MCQ distractors to use for a specific item. More generally, the content of items are much more important than their format.

However, there are many factors to consider when determining the number of options of MCQ distractors and the paper in its current form does not adequately deal with most of these. Finally, the overall position seems overblown - I think the case for having a variable longer list of options does exist, but only in certain circumstances (which require more nuanced articulation), and only when supported with specific content rationale and psychometric evidence.

No evidence is given throughout to substantiate the arguments. In fact, they are more assertions than arguments. The overall thesis that assessment organisers should consider having variable numbers of options is not controversial. Having an increased number of options for some specific items, if justified, is also not controversial. However, the justifications given are not without counterargument. If the author wishes to pursue the paper further, it would be good to at least anticipate the counterarguments to the position given, and to make some comment on them.

I think the biggest shortcoming is the failure to appreciate psychometric perspectives, especially based in Item-Response Theory or Rasch Measurement. No evidence based on validated models of MCQ assessment is provided.

The main thread running through the text involves a range of tautologies, such as there is no reason why the number of options has to be the same; and it is feasible to deliver tests via online platforms with different numbers of options. Neither of these statements necessarily mean that longer lists of options are best.

A few comments on each major point:

"Decisions related to broad clinical scenarios can are not (sic) limited to a small number of options"
This is probably the strongest suggestion in the paper as it relates to content relevance. Item quality is actually the most important element here, but, where relevant, more distractors may play a role. This does not mean, however, that more distractors should be used. Can does not imply ought. I think this assertion also falls over when considering the process of assessment through an item. The item is an indicator of competence, tied to the purpose of the exam, the articulation of a construct, and in the context of a blueprint to a curriculum. The item should be developed to ensure minimal false-positive and false-negative responses from candidates. In terms of measurement, more distractors add noise to the process, and compromise the ‘purity’ of the measurement process. They also run the risk of introducing a range of factors that are not content related, and thus are subjected to test-taking strategies of candidates, often resulting in a reduction in item discrimination.

"Options lists should include all possible combinations of factors"

I don't agree with the assertion that they 'should'. They certainly can, but again this can seriously compromise the functioning of the item. Also, 'all factors' is a subjective thing when it comes to medicine. When drafting an item I would start with the 'most relevant' factors and then cut out any that are superfluous and likely to add noise to the assessment data. Still, the position given simply strengthens the notion of having variable numbers of distractors. It does not necessitate long lists of them.

"Options that are rarely chosen can provide information regarding students and/or for students"

"Many of the analyses upon which the recommendations to reduce the number of options are based on the assumption that all incorrect responses are equivalent". Do they really? I would be very surprised if this was the case. Distractors work different for different questions. Anyone involved in the analysis of MCQ papers using IRT/Rasch appreciate this. But distractors do become irrelevant when they are a 'waste of space' - practically no candidates go for them. This just adds unnecessary reading loads onto time poor test-takers.

If one candidate selections a dangerous incorrect response, this may be useful information for students and teachers. But the outcome may also be because they ran out of time and guessed. It may also be because the question was poorly worded. While diagnostic feedback is incredibly important and distractor options should inform feedback processes, including long lists will not necessarily add value.

"Score reliability can increase with numbers of options"

I find this the weakest conjecture of all. It can, but that doesn't mean it will. The most important factors that influence reliability are item discrimination and item fit (when using IRT/Rasch models). Not to mention just running more items. I worry that this proposition is based on a simplistic, classical approach to MCQ item analysis.

"Longer lists of options are feasible for examinees to use, especially given computer based".
I have no issue with this. Yes it is feasible. That doesn't support the author's contention that is should be done. As a final point on feasibility though, this is fine.

I find the conclusion weak and simply another conjecture. Also, the final part of it is non-controversial. The paper require more detail, evidence and needs to be buttressed with a response to potential counterarguments.

A few minor typos:

- in the abstract, page 2 line 19 "can are not limited to"
- page 3 line 45.

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

No

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

No

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 am able to assess the statistics

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 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