PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Cross-sectional study comparing Single Best Answer and Very Short Answer questions for the assessment of applied medical knowledge in 20 UK medical schools
AUTHORS	Sam, Amir; Westacott, R; Gurnell, Mark; Wilson, Rebecca; Meeran, Karim; Brown, Celia

VERSION 1 - REVIEW

REVIEWER	Jan de Laffolie University Giessen, Germany
REVIEW RETURNED	06-Jul-2019

GENERAL COMMENTS	The authors address a very important issue im current medical education. First they demonstrate that SBA questions can overestimate true knowledge of the candidates, second a cost analysis for marking VSA questions is presented. Design is in general appropriate, some minor concerns arise: Questions 1. How large is the group of participants in comparison to all UK candidates? Can you provide some level of estimate if the group is representative and comparable? 2. Will you provide eSupp restricted read-only access to the VSA platform or parts of it for interested readers? The implementation of such a multicenter platform may be key to success. 3. Example Box 1 – the N, % of all students answering this question correctly in VSA seems very low? Any explanation? Was this an extreme example? 4. P8, line 55 – was a blanc answer in VSA recorded as incorrect accordingly? What happened if students did not complete one of the exams or both? Please explain. 5. It would be interesting to include negative cue rates, where the presented SBA would mislead students who answered VSA correctly? 6. The electronic platform should be described in more detail. concerns 1. The mean positive cue rate may be overestimated. It could be corrected by eliminating questions from the analysis that were answered correctly in the VSA part (therefore representing true knowledge). 2. Of all 32 medical schools only 20 participated. Is this group different from the others in terms of location, size, average score etc? Could participation and associated bias be a concern beyond the limitation section on level of medical school?

	3. If analyzing cost for marking, the authors might add some information or estimate how the costs may change with better/worse knowledge, anwering behavior etc.

REVIEWER	Phil Smith Cardiff University, UK I am on the Board of the MSCAA and so knew of the planning of this study, although did not participate in its planning or execution.
REVIEW RETURNED	12-Jul-2019

GENERAL COMMENTS	This is a useful paper that will take forward the improvement in knowledge assessment that is linked to the use of very short answer questions. It might have been useful/interesting to give an example of a question with negative SBA cueing, i.e. a question where VSA was easier for the candidate than SBA. I am not keen on putting the conclusion of the paper into the title, and feel it would be better for the reader to decide if the conclusion that the authors draw from their results is correct.

VERSION 1 – AUTHOR RESPONSE

REVIEWER 1

We thank the reviewer for their extremely helpful and constructive comments. We have addressed these in turn below.

1. How large is the group of participants in comparison to all UK candidates? Can you provide some level of estimate if the group is representative and comparable?

We apologise that this was not clear in the original manuscript. There are approximately 7,500 final year medical students in the UK, therefore the participants represented 19-20% of this cohort. We have edited the abstract to include all UK medical students rather than those at the participating schools (please see page 2, line 38) and included this figure in the results (please see page 9, lines 222-3). Our ethics approval did not include collection of data on the characteristics of the participants and therefore we are unable to assess whether or not the sample was representative. We have acknowledged this as a limitation of the study (please see page 13, lines 323-8).

2. Will you provide eSupp restricted read-only access to the VSA platform or parts of it for interested readers? The implementation of such a multicenter platform may be key to success.

Unfortunately, due to intellectual property restrictions and question bank security reasons, we are unable to provide access to the platform. Expressions of interest sent to the corresponding author will be forwarded to the UK Medical Schools Council.

3. Example Box 1 – the N, % of all students answering this question correctly in VSA seems very low? Any explanation? Was this an extreme example?

Thank you to the reviewer for highlighting this pertinent point. As the reviewer has quite rightly pointed out, we have chosen the question with the lowest facility to highlight the potential extent of positive cueing from an SBA, and we have now noted it as an extreme example in the discussion of this question (please see page 12, lines 292-3). We are currently exploring the potential underlying reasons for the low facility in a further study.

4. P8, line 55 – was a blank answer in VSA recorded as incorrect accordingly? What happened if students did not complete one of the exams or both? Please explain.

We apologise for not clarifying this in the original manuscript. A blank answer in the VSA was recorded as incorrect and therefore scored 0 (please see page 8, lines 182-3). Incompletion was rare (1,411 students completed all 50 SBAs, which we have now highlighted in the discussion, as well as comparable information on VSAs), so we did not exclude students without a full set of responses (please see page 9, lines 224-5 and page 13, lines 316-20).

5. It would be interesting to include negative cue rates, where the presented SBA would mislead students who answered VSA correctly?

We apologise for not including this data in the original manuscript. The negative cue rate was generally very low and we have included this in the discussion (please see page 13, lines 377-85).

6. The electronic platform should be described in more detail.

We thank the reviewer for highlighting this omission. We have now added further details in the methods section (please see page 8, lines 178-81).

7. The mean positive cue rate may be overestimated. It could be corrected by eliminating questions from the analysis that were answered correctly in the VSA part (therefore representing true knowledge).

We apologise that this was not adequately explained in the original manuscript. The mean positive cue rate for each VSA question (as a percentage) was calculated by:

$$\frac{\text{Number of participants answering VSA incorrectly AND SBA correctly} \times 100}{\text{Number of participants answering VSA incorrectly}}$$

(please see page 9, lines 202-4). Therefore the questions answered correctly as a VSA were not included in this calculation.

8. Of all 32 medical schools only 20 participated. Is this group different from the others in terms of location, size, average score etc? Could participation and associated bias be a concern beyond the limitation section on level of medical school?

We apologise for not including sufficient detail in the original manuscript. The participating medical schools were representative in terms of size and location; however there is no official comparison of UK medical schools by student ability therefore we are unable to assess whether the sample is representative in terms of ability (please see page 13, lines 312-3 and 323-8). We have also included the median and inter-quartile range of the number of participants per school (please see page 10, line 227).

9. If analysing cost for marking, the authors might add some information or estimate how the costs may change with better/worse knowledge, answering behaviour etc.

Many thanks for highlighting this point. We have expanded on this point in the discussion section (please see page 12-13, lines 304-9).

REVIEWER 2

We thank the reviewer for their supportive comments. We have addressed these in turn below.

1. It might have been useful/interesting to give an example of a question with negative SBA cueing, i.e. a question where VSA was easier for the candidate than SBA.

We apologise for omitting this information in the original manuscript and have added in details about the negative cue rate as also suggested by Reviewer 1, and an example question (please see page 13, lines 376-85).

2. I am not keen on putting the conclusion of the paper into the title, and feel it would be better for the reader to decide if the conclusion that the authors draw from their results is correct.

We apologise for this and have revised the title accordingly. Please see page 1, lines 1-2.