Reviewer’s report

Title: Can Standardized Patients Replace Physicians As OSCE Examiners?

Version: 2 Date: 27 October 2005

Reviewer: Glen Bandiera

Reviewer’s report:

General
The manuscript is well-written, easy to read, and addresses some important issues in OSCE design. We all would like to see a more efficient OSCE that doesn't jeopardize it's high objectivity and reliability and thus this paper tackles an important question. The concerns below should be addressed, however, to ensure the paper adds to the literature on OSCEs that currently exists. I think the strongest outcome of this paper is the students' acceptance of SPs as evaluators. The other conclusions need more support (see below)

Major Compulsory Revisions (that the author must respond to before a decision on publication can be reached)

In the limitations sections, The authors address the issue of comparing observations of completely different interactions (different stations)by two groups of evaluators (SPs and MDs). This is an important weakness of the study. It would be much stronger if there were data to show, even historically on previous rotations, that scores on the different stations were similar (or by how much they differed). Alternatively, the authors could run the OSCE again with the same observers marking the different stations and estimate a difficulty level for each station. This would help readers of the current study to put the 8% difference in average mark between MD stations and SP stations in perspective. Alternatively, as alluded to by the authors, redoing the study with multiple observers of the same interaction would be even better.

Did the physicians know the students? If they had previous experience with the students, for example on the wards, they may be biased by their overall previously developed opinion of the students. The authors should state if this is the case and address any implications of this relationship.

Finally, the outcome measure is of concern. An OSCE has strength in its ability to evaluate objective, behaviour-based performance with discrete steps. A problem-solving examination is best used to test higher cognitive functioning, information processing and factual knowledge. It may well be that the two tests are assessing fundamentally different things, notwithstanding the weak predictive validity of the physicians’ scores. The authors mention the summative nature of the MCQ test as a limitation and I think this should be broadened to include this concern as well.

Minor Essential Revisions (such as missing labels on figures, or the wrong use of a term, which the author can be trusted to correct)

Abstract, results, sentence 1. Authors should clarify between which scores there is a correlation (one SP with another or SP with physicians)

Background, para 1, last sentence. Authors should acknowledge that competence does not lead to performance, but merely proves capacity to perform.

Methods, para 2. Clarification is required here. The students underwent 9 stations, but the authors report only 8: 6 evaluated by MDs and 2 evaluated by SPs. Who evaluated the final station. Also, if only seven involved actual examination of a SP, which of the 8 listed stations did not involve SP examination?

Methods, para 2, sentence 5, 'student' should be 'students'
Data analysis: The authors should state that the decision to combine likert scale responses was made a priori, before data collection and analysis.

The discussion should be more detailed and describe how this paper adds to current understanding of OSCEs. The student's opinions could be further emphasized, as I think this is an important conclusion of the paper. Include more references to current literature on OSCEs (showing marking differences between MDs and SPs, for example).

Discretionary Revisions (which the author can choose to ignore)

**What next?:** Unable to decide on acceptance or rejection until the authors have responded to the major compulsory revisions

**Level of interest:** An article of importance in its field

**Quality of written English:** Acceptable

**Statistical review:** No

**Declaration of competing interests:**

'I declare that I have no competing interests'