Reviewer's report

Title: Rarely selected distractors in high stakes medical multiple-choice examinations and their recognition by item authors: A survey and simulation

Version: 1 Date: 28 July 2010

Reviewer: Danette McKinley

Reviewer's report:

Discretionary revisions

The authors’ methodology was appropriate given the study purpose. However, since this was a simulation using data from a high-stakes examination, it would have been interesting to see how pass-fail decisions would have changed, particularly for low-scoring examinees.

The authors’ recommendation to have item writers write as many distracters as possible is reasonable. However, they could have provided guidelines for those doing the analyses once the examination is administered. Should statisticians/psychometricians analyzing the data consider using one of the simulated processes? Typically, when the item doesn’t function well it is eliminated from scoring, and test score are recalculated based on the remaining scored items. Is it feasible to use these methods instead? Granted, the analyses conducted did not include how pass-fail decisions would change, but if these issues were included in the discussion, they could provide direction for future research. The limitations that the authors cited were appropriately stated.

The paper was well-written, generally clear, and their rationale for being exempted from ethical approval was acceptable.

Overall evaluation and general comments

This paper may be of some interest to those developing educational and high-stakes examinations. The authors reported on the results of a statistical analysis/simulation that examined what happens when the number of distracters is reduced in a single-best answer, multiple-choice question format. By simulating what examinees might do if non-functioning distracters were eliminated from the items, they examined changes in psychometric characteristics based on two models of examinee behavior. This research provides additional data on item writing and psychometric characteristics of multiple-choice items in high-stakes examinations. Item writers could be informed by research that assists them in developing MCQs that meet both statistical and content requirements.

While they concluded that the number of distracters could be reduced (based on their simulation) without much detriment to the psychometric characteristics of the examination, I wondered whether they considered other factors that could
affect examinee performance (and threaten the validity of the measure). First (as they acknowledged), for the most part, the non-functioning distracters attract some low-scoring examinees, they are serving their purpose. Second, when the number of distracters is reduced below three, random guessing is likely to provide the low scoring examinee with an advantage not related to the construct of interest – clinical competence or knowledge.

While expert reviewers may be able to identify distracters that are non-functioning (based on their plausibility) in the absence of item statistics, perhaps their time could be better spent rewriting the distracters to potentially improve item performance. This paper is interesting, but may be of interest to a narrow audience.

The results were consistent with the methods used, and, in general, there were no surprises. It was interesting to see that the discriminations were as high as they were given that the p-values were above 0.95.

**Level of interest:** An article whose findings are important to those with closely related research interests

**Quality of written English:** Needs some language corrections before being published

**Statistical review:** Yes, and I have assessed the statistics in my report.

**Declaration of competing interests:**

I declare I have no conflicts of interest to report.