Author’s response to reviews

Title: How to set the bar in Competency-Based Medical Education: Standard setting after an Objective Structured Clinical Examination (OSCE)

Authors:

Tim Dwyer (drtimdwyer@yahoo.com.au)
Sarah Wright (sarah.wright@utoronto.ca)
K Kulasegaram (mahan777@gmail.com)
John Theodoropoulos (john.theodoropoulos@gmail.com)
Jaskarndip Chahal (jchahal@utosm.com)
David Wasserstein (david.wasserstein@utoronto.ca)
Charlotte Ringsted (charlotte.ringsted@au.dk)
Brian Hodges (brian.hodges@uhn.ca)
Darrell Ogilvie-Harris (leafdoc1@rogers.com)

Version: 1 Date: 03 Nov 2015

Author’s response to reviews:

MEED-D-15-00108

How to set the bar in Competency-Based Medical Education: Standard setting after an Objective Structured Clinical Examination (OSCE)

Tim Dwyer; Sarah Wright; K Mahan Kulasegaram; John Theodoropoulos; Jaskarndip Chahal; David Wasserstein; Charlotte Ringsted; Brian Hodges; Darrell Ogilvie-Harris

BMC Medical Education
Dear Dr Dwyer,

Your manuscript "How to set the bar in Competency-Based Medical Education: Standard setting after an Objective Structured Clinical Examination (OSCE)" (MEED-D-15-00108) has been assessed by our reviewers. They have raised a number of points which we believe would improve the manuscript and may allow a revised version to be published in BMC Medical Education.

Their reports, together with any other comments, are below. Please also take a moment to check our website at

http://meed.edmgr.com/l.asp?i=3639&l=GEK0V0AN for any additional comments that were saved as attachments. Please note that as BMC Medical Education has a policy of open peer review, you will be able to see the names of the reviewers.

If you are able to fully address these points, we would encourage you to submit a revised manuscript to BMC Medical Education. Once you have made the necessary corrections, please submit online at:

http://meed.edmgr.com/

If you have forgotten your username or password please use the "Send Username/Password" link to get your login information. For security reasons, your password will be reset.

Please include a cover letter with a point-by-point response to the comments, describing any additional experiments that were carried out and including a detailed rebuttal of any criticisms or requested revisions that you disagreed with. Please also ensure that all changes to the manuscript are indicated in the text by highlighting or using track changes.

Please also ensure that your revised manuscript conforms to the journal style, which can be found at the Instructions for Authors on the journal homepage.
A decision will be made once we have received your revised manuscript, which we expect by 11 Oct 2015.

I look forward to receiving your revised manuscript and please do not hesitate to contact us if you have any questions.

Best wishes,

Hanan M.F Al-Kadri, MD., Ph.D.
BMC Medical Education
http://www.biomedcentral.com/bmcmemeduc/

Reviewer reports:

Reviewer #1: This is a well written and interesting paper. Although the investigation of the validity of the modified Angoff method is not particularly original (and the authors have not strongly established that the validity is likely to significantly differ in the setting of postgraduate training) this paper does take it further and analyses whether or not the modified Angoff method can credibly be used to establish two pass marks for a single OSCE assessing two cohorts of different experience levels, as part of a competency-based medical programme. It provides some indication that indeed, the modified Angoff method is valid for such use.

However, what the study fails to elucidate is whether or not the modified Angoff method is truly credible/valid for junior students. Whilst the authors have shown that using the modified Angoff method for senior students is valid (through comparison to the established BLR/BG methods), they have not shown the same for junior students. The results of this study reveal that for these junior students, there is a significant decrease in the pass mark, resulting in a significant reduction in the number of fails. Whether or not this new pass mark is veritably valid/credible has not been established, which is an important omission in the study design.
See page 12, line 2: One of the main limitations of this study was an inability to demonstrate the credibility of using the modified Angoff to establish a pass mark for the junior residents. Credibility for the use of the modified Angoff to set pass marks for senior residents was established by a comparison with the pass marks established using the BLR / BG methods. However, in the setting of the junior residents, the pass mark created was significantly lower than that set by the BLR / BG methods. While some credibility for this standard setting method was demonstrated by the finding that all three groups of judges set similar cut scores for junior residents, there was no alternative standard setting method that could be used for comparison.

Downing wrote that there is no single correct answer when comparing standard-setting methods [21], with Norcini and Shea stating that issues of student fairness are the most important – the passing scores must be acceptable to students, faculty and administrators [16]. Having nearly 30% of junior residents fail using the BG / BLR methods certainly appears unacceptable. The purpose of the end of rotation OSCE was to identify those residents not performing as well as their peer group, which the modified Angoff method appears to effectively do. Credibility can also be established by using a systematic approach, produced by qualified judges with a clear purpose [16], which was the case in the study. Methods should also be supported by a body of published research, be transparent, easy to implement and easy to explain – such methods justify the final result [15].

See page 7, line 21: Credibility of the modified Angoff method for the senior residents was judged by the number of PGY4&5 residents / orthopedic fellows passing the OSCE determined by each of the methods [20].

In addition, whilst on the whole this is a well written and readable manuscript, there are several occasions where the authors have made statements which have not been adequately cited (see minor revisions below). The authors should review the whole manuscript for such errors.

MAJOR REVISIONS:
Results: exact p-values and confidence intervals have not been provided.

Thank you. This has been done for Table 1. This has also been done for Table 2, which has been split into two tables for easier reading and understanding. See page 15-17.

Discussion: there is very little awareness/discussion of the limitations of this study.

Thank you, paragraphs on limitations have been added.

See page 12, line 22: Other limitations of this study include the use of only six judges in each group to perform the modified Angoff method, despite evidence that increased number of judges improve the reliability of the modified Angoff – however there was a high correlation between judges for the modified Angoff, and the pass marks created for the senior residents matched that set by the BG / BLR method. In this study, the credibility as opposed to the validity of standard setting methods was studied, with credibility established by comparing the pass/fail rates of different methods with a reference group that is expected to have a high pass rate [20]. While these two terms could be used interchangeable, credibility is typically used in the standard setting literature and was thus used in this study [20]. Finally, this study also uses the OSCE results of relatively few residents, especially in comparison to other studies that have used the results of medical students - however it would be difficult to increase these numbers in the setting of postgraduate orthopedic training without performing a multicentred study.

MINOR REVISIONS:

Page 5, Line 9: "judges review the each question"

Thank you, this has been edited.
See page 5, line 2: In the modified Angoff methods, judges reviewed each question after defining a borderline candidate, and decided whether the borderline examinee will respond correctly [13, 2, 6, 14].

Page 6, Lines 24-29: "the lead author discussed the modified Angoff method individually with each judge..." it would be useful to give more detail about the exact nature of the training that each judge received. Did the judges have previous experience of standard setting?

No judge had experience with standard setting – this has been added to the manuscript.

See page 5, line 23: All judges, with the exception of the fellows, were members of faculty, experienced at both teaching and examining orthopedic residents – no judge had previous experience with standard setting methods.

See page 6, line 10: The lead author discussed the modified Angoff method individually with each judge, and participated in the evaluation of the first 10 checklist questions on random stations.

Page 7, Lines 4-12: This section requires citations.

Thank you, citations for this section have been provided.

See page 6, line 6: For each checklist item for each station, the judges answered yes or no for “will the borderline examinee respond correctly to this item?” [14, 2]. Items were then assigned as yes=1 and no=0, and the pass point calculated by averaging the scores [2].

See page 6, line 15: After the conclusion of the study, the residents who were rated advanced beginner for each station were called the borderline group for that station. The mean checklist score of this group was calculated and used as the station pass mark [6-10].
See page 6, line 21: The Borderline Regression (BLR) method was used to calculate pass marks based on the results of a regression analysis, using the global station mark as the independent variable and the station checklist mark as the dependent variable [2,12,20].

Page 7, Line 48: ‘validity’ would perhaps be a better term than ‘credibility’

In the standard setting literature, credibility is the term that is typically used – this is established by comparing the pass/fail rates of different methods with a reference group that is expected to have a high pass rate [20]. However, we agree that in this setting that both terms could be used interchangeably.

This has been added to the limitations as above.

See page 13, line 3: In this study, the credibility as opposed to the validity of standard setting methods was studied, with credibility established by comparing the pass/fail rates of different methods with a reference group that is expected to have a high pass rate [20]. While these two terms could be used interchangeably, credibility is typically used in the standard setting literature and was thus used in this study [20].

Page 9, Lines 38+: This entire paragraph is lacking citations.

Thank you. Citations have been added regarding the challenges facing CBME, as well as the CBC program at the University of Toronto.

The information regarding the difference in performance between senior and junior residents is currently in press.

See page 9, line 15: One of the challenges of the CBME model is the organization of frequent, objective assessments, requiring considerable faculty involvement and resources [22-24]. The curriculum at our institution has 21 different rotations – each resident is required to demonstrate
a minimal level of competence in one rotation before progression to the next [25]. The sports medicine rotation is one of the major rotations, which residents undertake both as a junior and as a senior, with the curriculum map detailing different objectives for each group. While previous research had demonstrated that the junior residents could not master clinical skills as well as the seniors, we would prefer to continue using a single OSCE in the interest of feasibility (research will be published in December). Using the modified Angoff method to set different pass marks for junior and senior residents allows us to do so. In this manner, we can also identify the one or two junior residents who are performing poorly compared to their peers.

Page 10, Line 14-21: "However, there had been some concern that…” this sentence requires a citation.

Thank you. This concern was from the authors of this manuscript – this sentence has been reworded to clarify this issue.

See page 10, line 6: Initially, the authors of this paper had been concerned that sports medicine specialists, who were involved in content review and exam writing for the sports rotation, would set the cut-score too high, especially for the junior residents.

Reviewer #3:

Thank you for the opportunity to read this manuscript which provides new knowledge about standard setting for OSCEs.

The description of the problem is clear i.e. can you use the same OSCE station for trainees of different standards and generate acceptable pass rates for both groups, appropriate to their level of training. The description of the three different methods is clear.

There are a number of interesting findings - including the proposal that the content experts may generate higher cut scores.

The study raised the question of restriction of range for me. This OSCE is looking a little bit like Progress Testing. For the junior trainees, the range of scores accessible to them would be restricted, in comparison to the range of scores for an OSCE where they could be expected to
score 100%. Could this be a problem in reliability of the OSCE or ability to discriminate between trainees?

Thank for you this interesting point. The OSCE in this study was not designed to be progress testing, but rather to determine the competence of residents after a CBME rotation. However, in the CBME program at the University of Toronto, residents undertake this rotation twice, due to the perceived importance of the rotation. The results of the OSCE study demonstrated that junior residents were not able to achieve a minimal level of competence, while senior residents were – certainly this finding supports that typically demonstrated by progress testing, which uses regular assessments throughout an academic program to provide longitudinal evidence of the growth of student knowledge (Wrigley 2012). However, no resident undertook the OSCE twice, once as a junior and once as a senior. It is important to note that, as you highlight, continued iterations of the OSCE will be used in a manner similar to progress testing, to ensure that residents are performing as expected in comparison with their peer group.

With regards to restriction of range, the range of scores available to them was not restricted. The original paper demonstrated excellent reliability for the OSCE, and was able to identify significant differences between junior and senior residents. As a result we don’t believe that this affected either the reliability or the ability to discriminate between trainees.

This has been discussed in the manuscript.

See page 11, line 15: The results of the OSCE study demonstrated that while senior residents were able to achieve a minimal level of competence, junior residents were not (research to be published in December 2015). The results of this study are not dissimilar to studies that use progress testing, whereby regular assessment throughout an academic program are used to provide longitudinal evidence of the growth of student knowledge [26]. The sports medicine OSCE was not designed as a progress test, but rather to determine whether both senior and junior residents could achieve a minimal level of competence. However, continued iterations of the OSCE will be used in a manner similar to progress testing, to ensure that all residents are performing as expected in comparison with their peer group.

Minor point for clarification
Page 6 line 34 - after conclusion of the study - does the "study" refer to the OSCE examination?

Thank you, this has been clarified.

See page 6, line 15: After the conclusion of the study using an OSCE to examine the performance of residents after the sports rotation, the residents who were rated advanced beginner for each station were called the borderline group for that station.

There is a bit of inconsistency in the use of terms e.g. setting bar / pass mark / cut score. Or are they intended to mean something different? Consistent use of terminology, where possible, could reduce confusion

Thank you. Pass mark and cut score do mean the same thing, and the manuscript has been edited to use pass mark consistently.

The term setting the bar has been removed.

Dear Sir;

Looking into the two reviewers comments, you will see a clear agreement that this is a well-written manuscript that discusses important issue. However, the decisions of the two reviewers are widely varied.

Therefore the authors need to modify / justify their work through doing the needed major revision

1. Elucidate whether or not the modified Angoff method is truly credible/valid for junior students.
2. The results of this study reveal that for junior students, there is a significant decrease in the pass mark, resulting in a significant reduction in the number of fails. The authors
need to validate this new pass mark before concluding and accommodating the study results.

3. Several statements in the manuscripts need to be referenced.

4. Answer the reviewers’ comments point by point and submit their revised manuscript with highlights on the changed/updated areas. A letter to the editor explaining point by point the changes done need to be submitted.

Thank you, each of these issues have been addressed above.

---------------