

# Assessing clinical skills – standard setting in the objective structured clinical exam (OSCE)

**Mash B**, MBChB, MRCOG, DCH, DRCOG, PhD (Stell)

Associate Professor, Division of Family Medicine & Primary Care, Department of Interdisciplinary Health Sciences, University of Stellenbosch

**Correspondence to:** Prof Bob Mash, E-mail: rm@sun.ac.za

## Abstract

Family Medicine training and assessment is becoming more formalized and developed in South Africa. Assessment of competency in relation to clinical skills can involve observation in the clinical setting, but is more usually assessed in an examination. The traditional "long case" has been largely abandoned as it lacks reliability and validity. Summative assessment of family physician's clinical skills now usually includes an Objective Structured Clinical Examination (OSCE). Although a well designed and organized OSCE can have reasonable reliability and validity, a pass mark of 50% may in fact be an arbitrary figure, which does not credibly represent the required competency of a family physician. Standardisation of the OSCE is required to define the pass mark above which a candidate performs at the level expected of a family physician. A number of standardisation processes have been described that either judge the test items prior to the exam or judge the individual during the exam. In this paper we report on an example of the latter called the borderline regression method.

**SA Fam Pract 2007;49(3): 5-7**

## Introduction

Over the last few years Family Medicine in South Africa has reached consensus on the postgraduate curriculum, the specific outcomes for clinical skills<sup>1</sup> and has also provided new resources to assist with their teaching and learning.<sup>2</sup> At the end of the 4-year training period, however, assessment is required to determine if the registrar has achieved competency in these clinical skills.

Competency includes basic knowledge of the skill, for example the indications, contraindications and complications, but ultimately requires the ability to perform the skill in the clinical setting. Assessment of competency can involve observation and certification of competency within the clinical setting, but more commonly is assessed by showing how one would demonstrate the skill in an exam situation.

A number of different methods are available for the assessment of clinical skills within examinations and all methods have both strengths and weaknesses. Written papers and multiple choice questions are insufficient to assess performance of clinical skills. Traditionally the "long case" has been used, but while this assesses the whole consultation, it lacks reliability and validity. The main criticisms are that performance on one specific case does not

reliably predict performance with other cases and different candidates are not assessed against the same cases. In addition procedures and minor surgery cannot be assessed in this way. In order to assess candidates over a range of problems and clinical skills all Family Medicine departments in South Africa make use of the Objective Structured Clinical Examination (OSCE).

This article is intended to describe the technique of borderline regression as a method of improving the OSCE by setting a more valid standard to measure competency. This method has been introduced at Stellenbosch University in the Division of Family Medicine and has contributed towards a higher standard of clinical assessment. The article does not attempt to review the literature or critique alternative techniques, but anticipates that this example will encourage others to improve the quality of assessment.

## The OSCE<sup>3</sup>

The OSCE provides a circuit of stations and at each station the candidate demonstrates a specific skill on a real or simulated patient or if necessary a model. Stations usually last between 5-10 minutes and the OSCE may contain between 8-20 stations. Reliability is increased by sampling a wide range of

different clinical skills and having sufficient testing time. Reliability implies that the test consistently measures what it is supposed to and would provide the same result if repeated. Reliability may be reduced by including items that don't discriminate well between candidates (too easy or too hard), having patients or examiners who perform inconsistently and through poor organization of the OSCE. Examiners and patients may need prior training as reliability is increased by performing and assessing the station in the same way with each candidate. Scoring is performed using a pre-determined score sheet or checklist by an examiner at each station. Validity implies that the test actually measures what it is supposed to, in this case the competence of a family physician. For good validity it is important that the items assessed in the OSCE reflect the outcomes of and learning opportunities in the training programme, are recognized as relevant and important, and are evidence-based. OSCEs are logistically complex and labour intensive as they require a large amount of space, preparation, patients and examiners.

The score in the OSCE is criterion rather than norm referenced. This means that every candidate who performs according to preset criteria will pass. Norm referenced means that the

scores of each cohort of candidates is adjusted around a preset mean (e.g. 60%) and therefore is dependent on the performance of the group rather than the standard required to be a family physician. A candidate's score for the OSCE is usually taken as the mean of each station's score, so that each station contributes equally to the final mark.

The overall pass mark for the OSCE is typically set as 50%, which assumes that a candidate who performs at the level of a family physician should obtain 50% or more. This assumption however may not be correct as stations vary both in their difficulty and in the performance expected of a family physician with the problem presented. This issue is borne out by the frequent concerns of examiners that borderline candidates actually are not competent family physicians. As the content of the OSCE also varies each time it is set, a process for setting standards is required for each OSCE to set the pass mark at the level expected of a family physician, rather than arbitrarily assuming that this will be 50%.

**Standard setting methods**

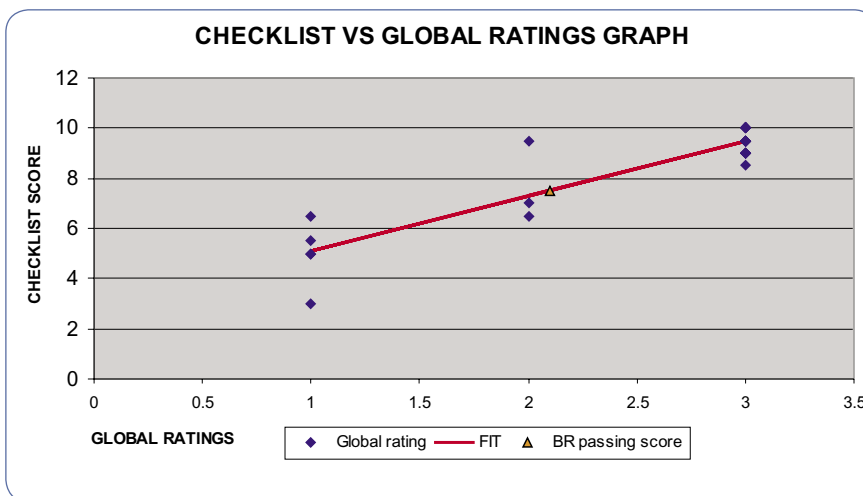
A number of different methods of standardization have been developed. In broad terms these fall into 2 groups, those that judge the test items prior to the exam and those that judge the individual during the exam. The best example of the former is the Angoff method in which a panel, consisting of 12 or more experts, independently evaluate each item in each station prior to the exam and predict how a borderline candidate would perform. The judges then discuss their predictions and if necessary revise them. Finally the average item score is determined and from this the station and OSCE pass marks are calculated. This method has been widely used, but clearly requires significant organizational effort prior to the exam, which may be beyond the resources of most Family Medicine departments.

Therefore, the borderline regression method was introduced at the University of Stellenbosch, which can be performed by the examiners during the exam and calculated automatically once the scores are entered into an Excel spreadsheet. This method has also been shown to provide a more credible and reliable standard than the Angoff method.<sup>4</sup> This method involves the examiners making two separate judgements with each candidate. Firstly, they score the candidate as usual according

**Figure 1: Example of score sheet for one OSCE station**

Student name	
<b>Checklist score</b>	
<b>Please demonstrate and explain the steps you would take to remove and treat an ingrowing toe nail with phenol ablation.</b>	
Injects 2% lignocaine without adrenaline as a ring block	1
Applies a tourniquet – rubber tubing or a glove	0.5
Cleans with betadine	0.5
Removes any inflammatory debris	0.5
Loosens nail	0.5
Uses straight scissors to cut a 3-5mm wide slice	1
Extends below nail fold to the root	1
Removes with a rotational movement and forceps	0.5
Dries hole with cotton wool	0.5
Carefully inserts cotton bud dipped in phenol, rotates for 45 seconds, repeat x2-3	1
Cleans and irrigates with normal saline	0.5
Removes tourniquet	0.5
Covers with jelonet / dressing	0.5
Informs patient to shower twice a day in order to clean the wound	1
Warns patient of or monitors them for signs of infection	0.5
Allow for alternative reasonable methods	
<b>Global Rating (Tick the appropriate box)</b>	
<b>Is NOT at the level of a Family Physician</b>	<b>Is at the level of a Family Physician</b>
<b>Unsure</b>	

**Figure 2: Regression of global rating against checklist score for results in Table 1**



Note: Checklist passing score is 7.5 for this station.

to the checklist and then, secondly, they give a global rating of the candidate's overall performance (See Figure 1 for an example). In our OSCE we used a 3-point global rating scale:

- 1-clearly not a family physician,
- 2-not sure / borderline,
- 3-clearly a family physician.

The global ratings for each candidate at a station are then regressed against the checklist scores (See Table 1 and Figure 2 for an example). A score of 2.1 (55% or just above "unsure") on the global rating axis was used to determine

the pass/fail checklist score for that station. This score represents the minimum that a family physician should obtain for that station. Stations that expect a high performance or are too easy will have a higher pass mark and stations that were too hard a lower pass mark (Table II). The pass mark for the OSCE is then established as the mean of the station pass marks (Table II). In this OSCE the final pass mark was set at 56%.

The borderline regression method still relies on the expertise of the examiner to make an appropriate judgement and will be affected by the number of candidates

**Table I:** Results of candidates for one OSCE station

	Checklist score	Global rating
Candidate 1	7	2
Candidate 2	9	3
Candidate 3	9.5	3
Candidate 4	10	3
Candidate 5	9.5	3
Candidate 6	9.5	3
Candidate 7	9	3
Candidate 8	9	3
Candidate 9	10	3
Candidate 10	9.5	3
Candidate 11	5	1
Candidate 12	9.5	3
Candidate 13	9	3
Candidate 14	5.5	1
Candidate 15	10	3
Candidate 16	9.5	3
Candidate 17	9	3
Candidate 18	9.5	3
Candidate 19	6.5	2
Candidate 20	10	3
Candidate 21	6.5	1
Candidate 22	9	3
Candidate 23	8.5	3
Candidate 24	3	1
Candidate 25	9.5	2
Candidate 26	9.5	3
Candidate 27	9.5	3
Candidate 28	9.5	3
Candidate 29	10	3
Candidate 30	10	3
Candidate 31	5	1

**Table II:** Pass scores for each station and OSCE

Station	Pass score
Station 1	7.54
Station 2	7.51
Station 3	4.71
Station 4	4.27
Station 5	4.70
Station 6	5.89
Station 7	4.96
Station 8	5.22
Station 9	6.87
Station 10	5.93
Station 11	6.81
Station 12	5.35
Station 13	5.14
Station 14	5.79
Station 15	5.22
Station 16	5.34
Station 17	5.37
Station 18	3.88
<b>OSCE PASSING SCORE</b>	<b>5.58</b>

contributing data to the standard setting calculation. The examiner's global rating score will also be influenced to some extent by their knowledge of the checklist score – although in our OSCE we did not total the checklist until after the exam.

### Conclusion

This article sets out the rationale for having a standard setting procedure as part of high-stakes OSCEs and illustrates the use of the borderline regression method in the OSCE at Stellenbosch University. The use of standard setting procedures

has not been widely used in South Africa and it is hoped that this article will encourage others to improve the credibility of their assessments of clinical skills.

### References

1. Mash B, Couper I, Hugo J. Building consensus on clinical procedural skills for South African family medicine training using the Delphi technique. *SA Fam Pract* 2006;48(10):14
2. Mash B and Blitz-Lindeque J. *South African Family Practice Manual* (2<sup>nd</sup> ed). Pretoria: Van Schaik, 2006.
3. Smee S. Skills based assessment. *BMJ* 2003;326:703-706
4. Kramer A, Muijtjens A, Jansen K, Düsman H, Tan L, Van Der Vleuten C. Comparison of a rational and an empirical standard setting procedure for an OSCE. *Medical Education* 2003;37: 132-139

It's the shell that makes Ecotrin safer.

Safety-Coated Ecotrin 81mg

The miracle of Aspirin made safer.

Each tablet contains Aspirin 81mg. Reg. No. 29/27/0767  
Pharmatica (Pty) Ltd, 33 Hubert Road, New Centre, Johannesburg 2001  
Under licence from Goldshield Pharmaceuticals Ltd, U.K.