

Standard Setting the NEBDN OSCEs (3rd edition)

Dr Gareth Holsgrove

July 2011

Introduction

The NEBDN OSCEs are designed in line with contemporary best practice. Amongst other things, this means that they focus on essential competencies, skills and professional behaviours (including communication skills and professional appearance). Unlike the written component, where the aim is to have the majority of items classified as moderately difficult with a smaller proportion easy or difficult, OSCEs tend to have a majority of easy items – this is because they are essential and, therefore, most candidates should be good at them. The passmark for OSCEs therefore tends to be higher than for written examinations.

Station marking

The task at each station is marked according to the main component parts of the task. Also, an **Overall Judgement** is indicated by placing the candidate into one of three categories: pass, borderline or fail.

Marks for the component parts of the task are marked either in respect of whether the component was done (meets the standard) or not done - for example, greeting the patient; or for more complex elements how well it was done. These complex elements are rated in one of four categories:

- exceeds standard
- meets standard
- just below standard;
- well below standard or not done.

The last of these categories carries zero marks. The other three categories carry marks which can vary according to the importance of the respective component of the overall task, the difficulty in completing the component, and the time a reasonable competent candidate would be expected to take to carry it out. However, these marks are not shown on the examiners' marksheets. This is because it has been found that when examiners know about the allocation of marks they are inclined to adjust their own marking to suit their pre-conceived outcome.

Where an actor is involved, for example, as a simulated patient, they also rate the candidate, but on different points than the examiner (for example, on the clarity of explanations or answers to questions).

Thus, each candidate is marked for the main component parts of the task by the examiner and, if one is involved, by the actor, too. In addition, the examiner indicates their **Overall Judgement** – pass, borderline, or fail.

It is not necessary for the marks on a station to come to any particular total. In the past it was usual for each station to be marked out of the same total (usually 10 or 20) but this is now seen as unnecessary and an obstacle to good station design. Instead, the marks for each station are converted into percentages before statistical analysis and the production of the final marks and results.

The **Overall Judgement** does not contribute directly to the marks, but is used for two other purposes:

1. As a check on the passmark, using the 'borderline group' method of standard setting.
2. As an additional QA measure, providing information about whether candidates' marks for the station are congruent with the pass/borderline/fail *Overall Judgement*.

Standard setting

The passmarks for the written examination will be set using a modified Ebel method and is described in detail elsewhere. That for the OSCE will be set on a station by station basis, with the final OSCE passmark being calculated from the sum of the station passmarks (as percentages) divided by the number of stations.

Individual station passmarks are set with reference to the total marks available for the 'meets the standard' column and, where elements have just a done/not done option, the marks available in the 'done' column. However, as it stands this method can result in a candidate gaining marks in all but one of the 'meets the standard' and 'done' columns but fail the station. Consequently, in setting the station passmark the total available in the 'meets the standard' and 'done' columns is calculated and the passmark is this total minus (usually) 10% and rounded down to the nearest whole number. This method is used successfully in a number of high stakes OSCEs including, for example, the MJDF.

For example:

EXAMINER'S MARKSHEET

Action	Above standard	Meets standard	Just below standard	Well below standard or not done
1. xxxx		3		0
2. xxxxxxxxxxxx		5		0
3. xxxxxxxxxxxx		3		0
4. xxxxxxxxxxxxxxxxxxxx xxxxxxx	5	3	1	0
5. xxxxxxxxxxxxxxxxxxxx xxxxxxxxxxxx	3	2	1	0
6. xxxxx		3		0
7. xxxxxxxxxxxx	5	3	1	0
8. xxxxxxxxxxxxxxxxxxxx xxxxxxxxxxxxxxxx	3	2	1	0

Note that actions 1, 2, 3 and 6 are either done (meets standard) or not done, whereas the other actions are marked on the basis of performance in one of four categories.

Total of 'meets the standard' column = 24

Total minus 10% = $24 - 2.4 = 21.6$

The total minus 10% (21.6) rounded down to the nearest whole number gives the passmark for this station of 21 out of a possible maximum mark of 30 (ie 70%).

This passmark can be checked *post hoc* using the borderline group method which is described elsewhere.