



The Society of Apothecaries of London

Standard Setting Procedures for Diploma Examinations

Dec 13/v1

How is a standard defined?

A standard is a statement about whether an examination performance is good enough for a particular purpose. It is based on judgements about the performance of examinees against an adequate level of competence which has been agreed on by experts in the specialty or pre-determined standards of competence agreed on by experts in the specialty.

What is standard setting?

The process of setting the numerical score that serves as the boundary between passing and failing an assessment.

Why is standard setting necessary?

It is impossible to make all questions in an assessment equal in terms of difficulty i.e. some will be more difficult than others. Standard setting takes into account the varying levels of difficulty of each question. Thus, the pass/fail cut off score is determined by predicted or actual candidate performance, rather than determining in advance a set failure rate for the assessment.

On what is standard setting based?

The standard is determined by informed assessors who are practitioners in the relevant field, knowledgeable about the curriculum or syllabus content and understand the candidates' educational process.

What are the characteristics of the borderline pass candidate?

A borderline pass candidate is one who is competent. The characteristics of this candidate were defined by a panel of examiners from clinical diplomas as:

- safe to practise;
- acceptable communication skills;
- ethical, basic technical skills;
- acceptable to patients;
- reasonable insight;
- potential for improvement.

Standard setting by the Modified Angoff method

This method is based on predictions by the panel of the proportion of borderline pass candidates who would (not should) answer a question correctly or demonstrate a skill competently. Judgements are made by individual members of an expert panel of 5-7 individuals. Discussion focuses around differences between judges, particularly between high and low scorers on the item in question. The threshold for discussion is 0.3 or a 30% difference between predictions on any item. Judges may decide to modify their score as a result of the discussion or to stay with their original estimate. The Angoff score for each

single best answer question is the average of all of the judges' scores and the pass mark for the paper is the average of the individual question Angoff scores. For an OSCE, the judges' scores for each mark sheet item are averaged to give an Angoff score and the average of the Angoff scores for all of the items becomes the passing standard for the OSCE station. Determining the passing standard for the entire OSCE examination is described below.

What information is required prior to standard setting?

The panel of judges is provided with the correct answers to knowledge-based questions before standard setting. When it exists, it will also be provided with data on candidate performance in prior assessments to assist in the setting of the pass marks.

What if a question has several parts and requires several answers?

In a structured answer or short answer question, or an OSCE, candidates may be asked for or expected to mention multiple points e.g. drugs, infections etc. or related points. In such cases, the estimate needs to be made about how many marks a borderline pass candidate would achieve in each individual part. In Best of Five questions, the estimate is made only for the proportion of borderline pass candidates who would give the single correct answer.

What determines whether the OSCE has been passed?

Candidates must meet two criteria in order to pass the OSCE. The numerical pass scores for all stations are summed to determine the overall pass standard for the whole exam. They must achieve the numerical pass mark for the whole exam as well as a proportion of the stations (normally two thirds or 8/12 stations). Candidates who fail one third (four or more stations) will fail the exam, even if they have achieved the numerical pass mark. Borderline fail candidates cannot then compensate for several poor performances with a few very good ones.

Standard setting criteria

Credible - with a clear explanation and rationale

Defensible – valid and with a demonstrable process reflecting effort to get it right

Stable – standards should not vary from year to year

Feasible – with available time and personnel

Supported - by evidence from the literature

Standard setting will always be arbitrary to a certain extent as it contains assumptions about the required level of competence, prior learning, teaching and experience and curriculum end goals. The aim is to minimise this by focusing on the borderline pass candidate characteristics and standard-setter training.

Dr Erica Allason-Jones

Ms Angela Hall

Dr Melinda Tenant Flowers

Updated December 2013