

Summative assessment & standard-setting

Roger Neighbour



Synonyms

For today, ‘summative assessment’
means the same as:

- # End point assessment
 - # Licensing examination
 - # Certification exam
-

Why have summative assessment at the end of training?

- # Patient safety
 - # **Patient safety!**
 - # Trainee motivation: ‘assessment drives learning’
 - # Recognition of achievement
 - # ‘Rite of passage’, end of apprenticeship
 - # Reputation of the discipline
 - # Quality marker for patients
-

A good summative assessment is:

<u>Valid</u>	<u>Reliable</u>	<u>Feasible</u>
Content validity	Accurate	Practicable
Construct validity	Consistent	Cost-effective
Predictive validity	Fair	Proportionate

What is good for what?

	Computer test	Written test	OSCE	Oral exam	Video	Work-based ass't
Factual knowledge	++	+	?	+	-	?
Clinical managem't	+	++	++	+	+	+
Clinical skills	--	?	++	-	--	++
Communi-c'n skills	--	-	+	?	++	+
Attitudes, values	--	--	?	++	+	++
Hidden curriculum	--	?	?	+	+	+

Recruiting examiners

Qualities

- # Credibility
- # Can 'rank order'
- # Trainable
- # Impartial
- # 'Team players'

Incentives

- # Status
 - # Influence
 - # Stimulation
 - # 'Make a difference'
 - # Money
-

Standard-setting

Deciding who should pass
and who should fail
is a question of policy
rather than arithmetic.

Questions to ask about standard-setting before you do the arithmetic

- # What is the main purpose of the assessment?
 - # What is at stake
 - for candidates?
 - for patients?
 - for your organisation?
 - # Who has an interest in the outcome?
 - # What is the message you wish to convey?
 - # What may be the effect of a high or low pass rate?
-

... and some more questions!

- # In a multi-component exam, what are the rules of combination?
 - # Who should set the standard?
 - Examiners?
 - 'Ordinary' practitioners?
 - Patients?
 - # Should the standard be 'absolute' or 'relative'?
 - # What will happen to people who fail?
 - # Is there to be an appeals procedure?
-

High stakes assessments need high defensibility

- # Transparent processes for marking & standard-setting
 - # High indices of reliability (>0.8 , $>+0.4$)
 - # Correction for test variance & error of measurement
 - # Low examiner variance
 - Recruitment
 - Training
 - feedback
 - # Fair appeals procedure
-

Two examples of standard-setting procedures

Modified Angoff procedure

- suitable for multiple choice exam

The Hofstee method

- (a compromise relative/absolute method)
 - suitable for overall pass/fail decision
-

Modified Angoff procedure

- # Reference group of stakeholders discuss characteristics of a borderline candidate ‘only just good enough to pass’
 - # For each item on the test, judges estimate percentage of borderline candidates that would answer item correctly
 - # Pass/fail standard is average of percentages for the items
-

Hofstee method

Reference group of judges agrees:

- # Lowest acceptable fail rate (A)
 - # Highest acceptable fail rate (B)
 - # Lowest permissible passing score (C)
 - # Highest required passing score (D)
-

Hofstee method

