

Standard Setting

Dr Gay Fagan

WARWICK

Standard setting

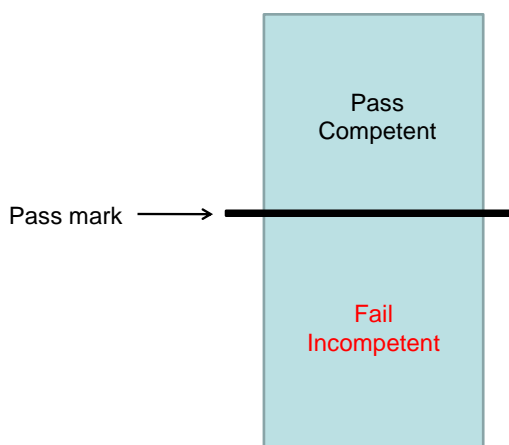
- What is standard setting?
- What methods are available?
- How do we decide which method to use?
- How do we do it?

WARWICK

What is standard setting?

Standard setting is the process of deciding what level of performance is acceptable for a particular level of purpose

i.e. establishing a pass/fail mark



Characteristics of standard setting

Warwick
Medical School

The method has to be

- explicable
- defensible
- reproducible
- feasible

WARWICK

What methods are available?

Warwick
Medical School

- Norm-referenced (relative)
- Criterion-referenced (absolute)
- Compromise methods

WARWICK

Norm-referenced methods

- An individual score is interpreted relative to the score of everyone else who took the exam.
- A set proportion of candidates fails regardless of how well they perform e.g. the top 84% pass.
- Useful for discriminating between high and low performers for ranking, selection.

Criterion-referenced methods

- Individual score is judged by degree to which an acceptable standard is reached
- Useful for tests of competence, to establish that the examinees know enough for a particular purpose e.g. licensure
- Potentially possible for everyone to get 100% (all pass) or everyone to get 0% (all fail)

Criterion-referenced methods

- Angoff
- Ebel
- Borderline Group

Compromise method

- Hofstee

Angoff method

- Select the judges
- Discuss
 - purpose of the test
 - nature of the examinees
 - what constitutes adequate/inadequate knowledge
 - the borderline candidate i.e. minimally competent

Angoff method

- Read the first item
- Each judge decides for each item the probability that a borderline examinee will answer the item correctly
- Record ratings, discuss and change if necessary
- Repeat for each item

Angoff Method

Items	Judges					Mean
	1	2	3	4	5	
1	0.60	0.70	0.55	0.75	0.65	0.65
2	0.80	0.90	0.85	0.95	0.90	0.88
3	0.70	0.75	0.80	0.75	0.40	0.68
4	0.45	0.55	0.50	0.60	0.55	0.53
5	0.90	0.95	0.85	0.95	0.90	0.91
Pass mark						3.77

Ebel Method

Involves two rounds of judgements.

1. Judges classify all the items into six categories based on the level of difficulty (easy, moderate, hard) and two levels of relevance (critical, general)

	Easy	Moderate	Hard
Critical	15	25	5
General	25	25	5

Ebel Method

2. Judges estimate for each category 'the % of items that a borderline examinee will answer correctly'

	Easy	Medium	Hard
Critical	80%	60%	50%
General	60%	50%	30%

Ebel Method

Number of items in a category is multiplied by the percentage correct to give a score for each category and these are summed to give the pass mark:

	Easy	Medium	Hard	Overall total
Critical	$15 \times 0.8 = 12$	$25 \times 0.6 = 15$	$5 \times 0.5 = 2.5$	29.5
General	$25 \times 0.6 = 15$	$25 \times 0.5 = 12.5$	$5 \times 0.3 = 1.5$	28
Pass mark				57.5

Advantages of criterion-referenced methods

- Focus attention on item content
- Relatively easy to use
- There is a considerable body of published work supporting their use
- Used frequently in high stakes testing

Disadvantages of criterion-referenced methods

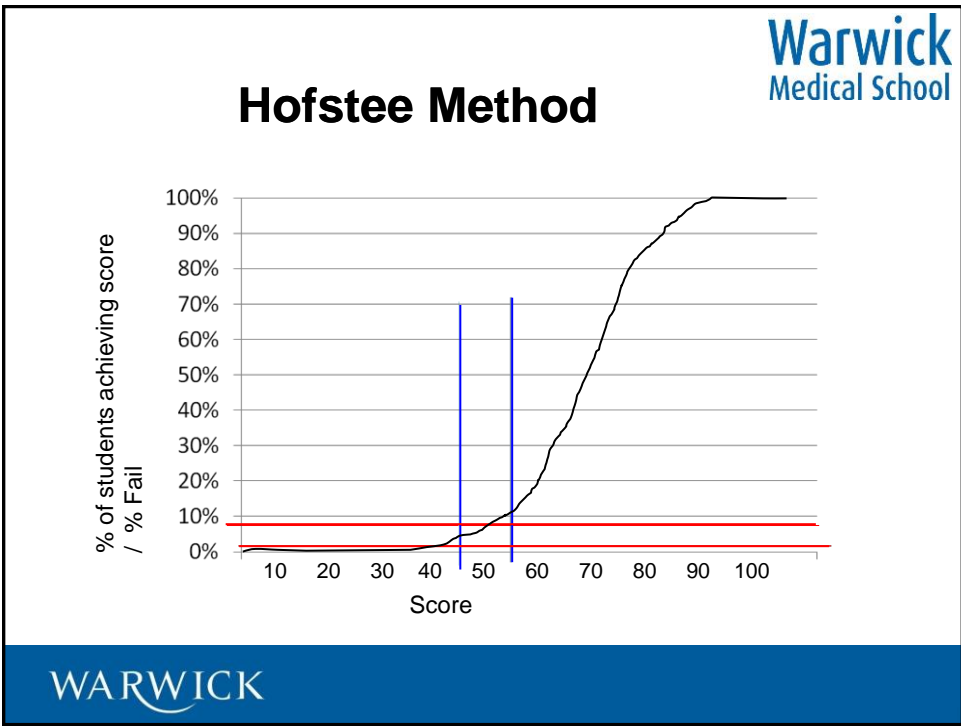
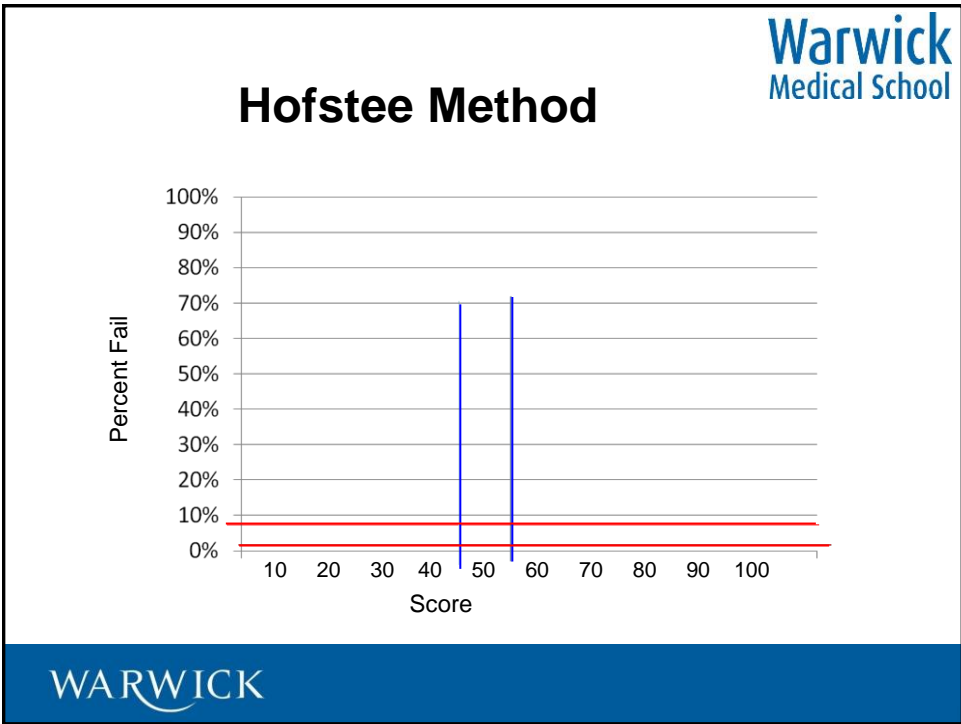
- The concept of the a “borderline candidate” is difficult to conceptualize
- Judges sometimes feel they are “pulling numbers out of air”
- The methods can be labour intensive and time consuming

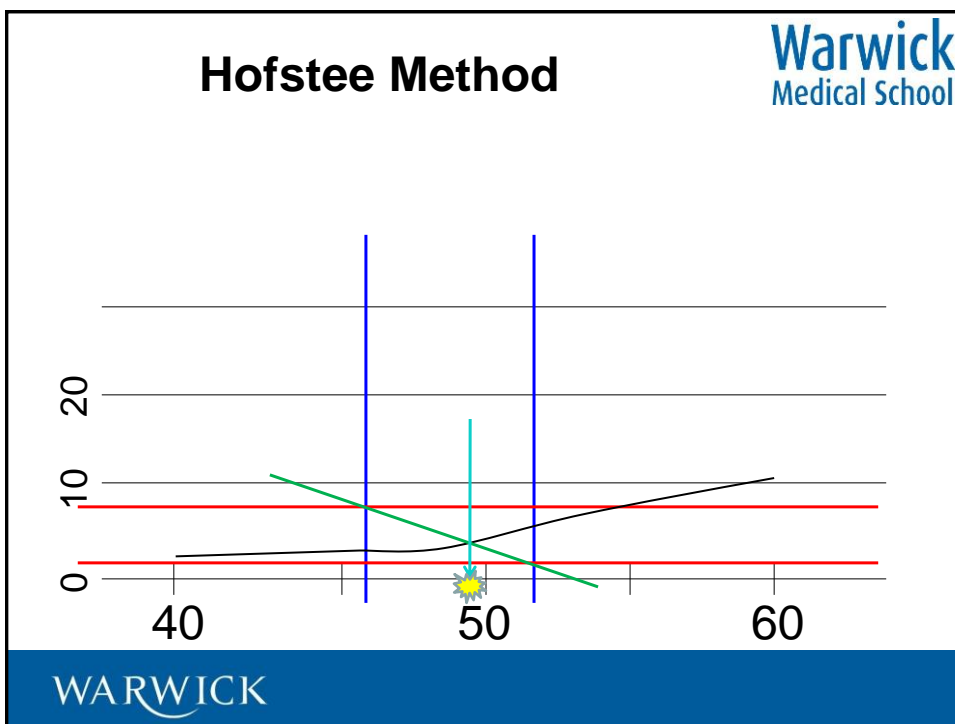
Hofstee method

- Selection and training of the judges
- Each judge reviews the exam in detail – identifying how many items a borderline candidate would answer correctly
- Each judge is asked to answer four questions:
 - What is the **minimum** acceptable pass mark for a borderline student?
 - What is the **maximum** acceptable pass mark for a borderline student?
 - What is the **minimum** acceptable fail rate?
 - What is the **maximum** acceptable fail rate?

Hofstee method

- Discuss results (outliers?)
- Calculate mean (or median) maximum and minimum pass mark and percentage fail
- Draw on graph
 - Example:
 - Pass mark 46 – 52
 - Failure rate 2 – 8%





Hofstee Method

- Advantages
 - Easy to implement
 - One of the most time-efficient methods
 - Educators are comfortable with the decisions
- Disadvantages
 - The graph may not go through the Hofstee box
 - Pass mark not known until after the exam


WARWICK


Borderline Group Method

Uses the examiner's clinical expertise to judge the candidate's performance


Examiner completes the checklist scoring and allocates a **global rating** based on the candidate's overall performance at that station

Unsatisfactory
Borderline
Satisfactory
Excel





Borderline Group Method



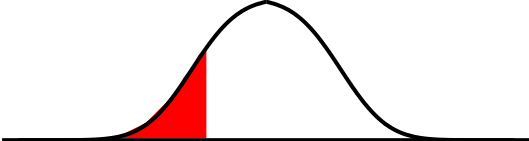
Checklist

1. Ησ σηςσ σφνησσ σφης σφς σφ
2. Κκςσ σκςσςσςσςσςσθλ θλθμθ θ θ σλ
3. Λαλκς κδμ δδκς δλκλ δλλδ
4. Κεγς δδ ε ρ ρρμτ τμκ
5. θφφκς δδ
6. Ησκλςσ σκφ σλς σκς ακ ακλ σλδ
7. Ηδηδδθ σηςσ σηηκςσ κςσ

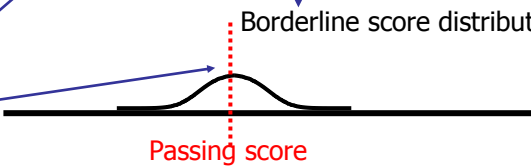
TOTAL Σ

Excel, Satis, Borderline


Test score distribution



Borderline score distribution



Passing score



Borderline Group Method

Warwick
Medical School

Advantages

- Utilises the expertise of the examiners
 - they are observing the performance of the students
 - they are in a position to make a (global) judgement about the performance based on
 - their clinical expertise
 - expected standards for the level of the test
 - knowledge of the curriculum/teaching

WARWICK

Borderline Group Method

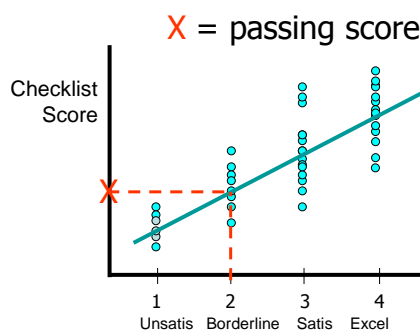
Warwick
Medical School

Disadvantages

- Relies on a large number of well trained examiners
- Global judgements not independent of checklist scoring
- Small number of candidates judged to be borderline make pass mark unstable
- Requires expert processing of marks immediately after the exam

WARWICK

Borderline Regression Method



WARWICK

Compensatory / Conjunctive standards

Compensatory: standard set using the sum of the individual pass marks
Works well with highly correlated items

Conjunctive: uses individual standards for each item

WARWICK

Returning to pass marks

Warwick Medical School

Pass mark is an estimate - error within it

Therefore, we cannot be certain that our pass mark is the exact score where competence is demonstrated

Need to adjust the pass mark - the **standard error of measurement** is used

WARWICK

Pass mark

Warwick Medical School

Pass mark →

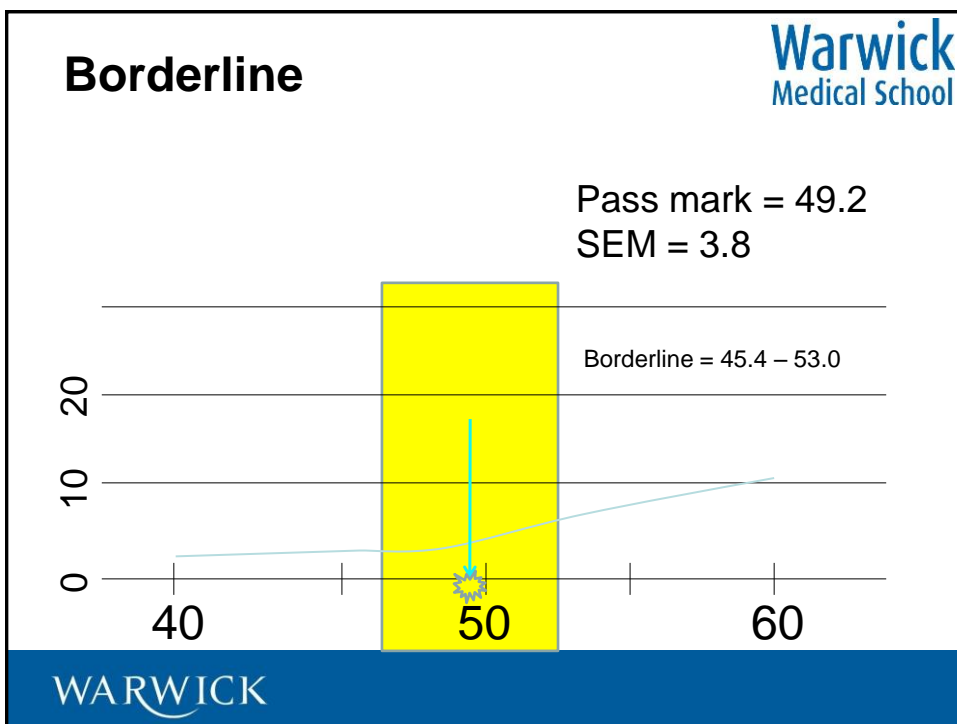
Pass Competent

Borderline Competent or not?


Fail Incompetent

Error of measurement



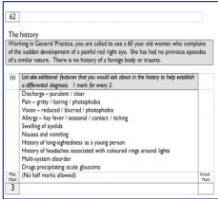
WARWICK



Summary



- Norm-referenced methods**
 - ranking/selection
- Criterion-referenced methods**
 - Angoff & Ebel – written exams
 - Borderline Group/Regression
 - performance based exams
- Compromise methods**
 - Hofstee – written exams

WARWICK

Summary

There is no perfect standard setting method

No method is absolutely accurate

All depend on expert judgements

Not so much the **METHOD** as the **PROCESS**

- suitable judges
- due diligence applied
- defensible rationale