



Principles of Standard Setting

Katharine Boursicot
Trudie Roberts

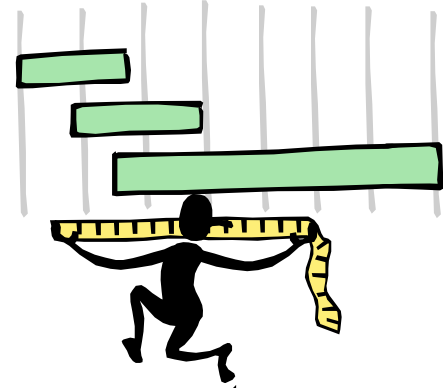
Setting Standards

- Scores and standards
- Characteristics of credible standards
- Methods
 - Relative standard setting methods
 - Absolute standard setting methods
 - Compromise methods
- Steps in implementation

A maths test

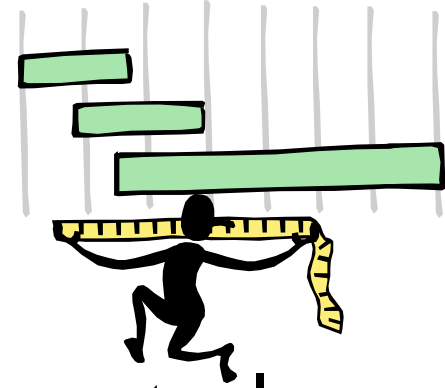
$$\begin{array}{r} 2683 \\ \times \quad 57 \\ \hline 15781 \\ 13415 \\ \hline 149931 \end{array}$$

Definition of Scores



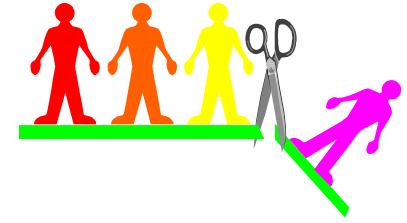
- A score is a number or letter that represents how well an examinee performs along a continuum
 - The degree of correctness for a response or group of responses

Definition of Scores



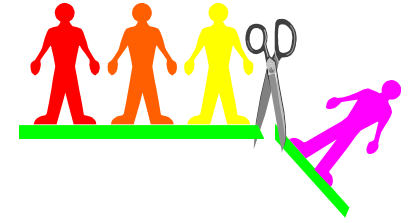
- For e.g. MCQs a score is based on the actual responses of examinees - a count
- For formats reproducing complex clinical situations with high fidelity
 - May involve weighting (degrees of correctness)
 - May involve an interpretation of the examinee's responses (e.g., oral exam)

Definition of Standards



- A standard is a statement about whether an examination performance is good enough for a particular purpose
 - A special score that serves as the boundary between passing and failing
 - The numerical answer to the question “How much is enough?”

Standards



- Standards are based on judgments about examinees' performances against a social or educational construct
e.g. Competent practitioner or student ready for graduation

The Standard Setting Problem

		Competent	Incompetent
		Pass	Fail
Test Result	Pass	Green	Red
	Fail	Red	Green

Setting the pass mark: characteristics of credible standards

The method has to be:

- Defensible
 - Credible
 - Supported by body of evidence in the literature
 - Feasible
 - Acceptable to all stakeholders
-
- Norcini, J. J. (2003). Setting standards on educational tests. *Medical Education*, 37, 464-469.
 - Norcini, J. J. & Shea, J. A. (1997). The credibility and comparability of standards. *Applied Measurement in Education*, 10, 39-59.

Classification Scheme

Relative methods

- based on judgments about groups of test takers

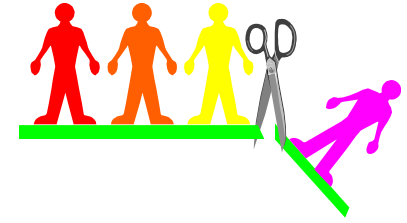
Absolute methods

- based on judgments about test questions
- based on judgments about the performance of individual examinees

Compromise methods

- Livingston, S.A. & Zeiky, M.J. (1982) *Passing scores: a manual for setting standards of performance on educational and occupational tests* Educational Testing Service, Princeton

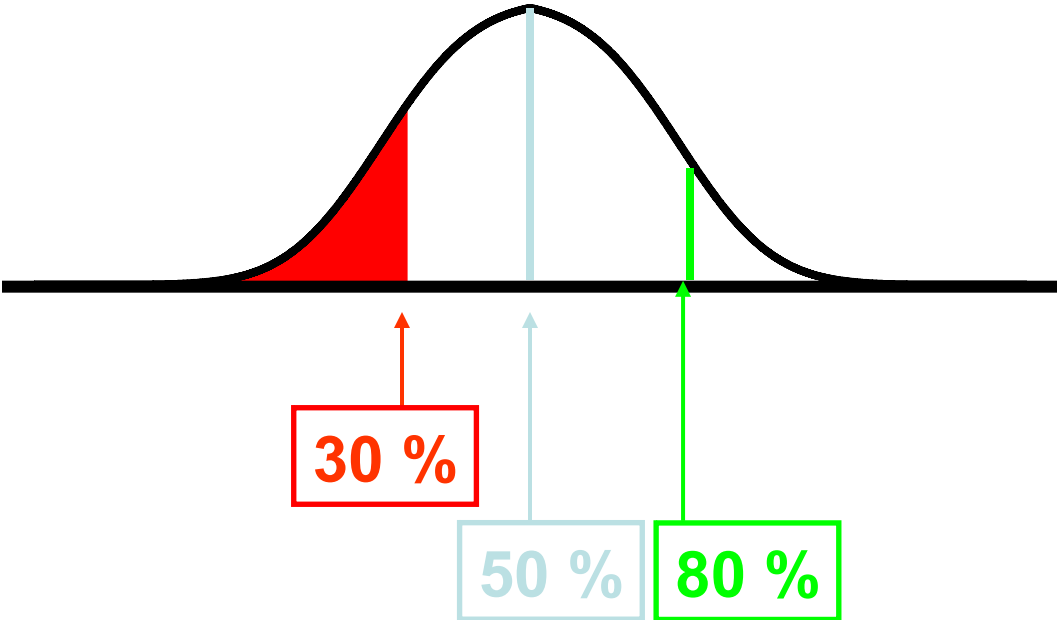
Types of Standards



- **Relative standards/ norm referenced methods:**
 - Based on a comparison among the performances of examinees
 - A set proportion of candidates fails regardless of how well they perform e.g. the top 84% pass
- **Absolute standards/ criterion referenced methods:**
 - Based on how much the examinees know
 - Candidates pass or fail depending on whether they meet specified criteria e.g. examinees must correctly answer 70% of the questions

Norm-referenced standard

Test score distribution

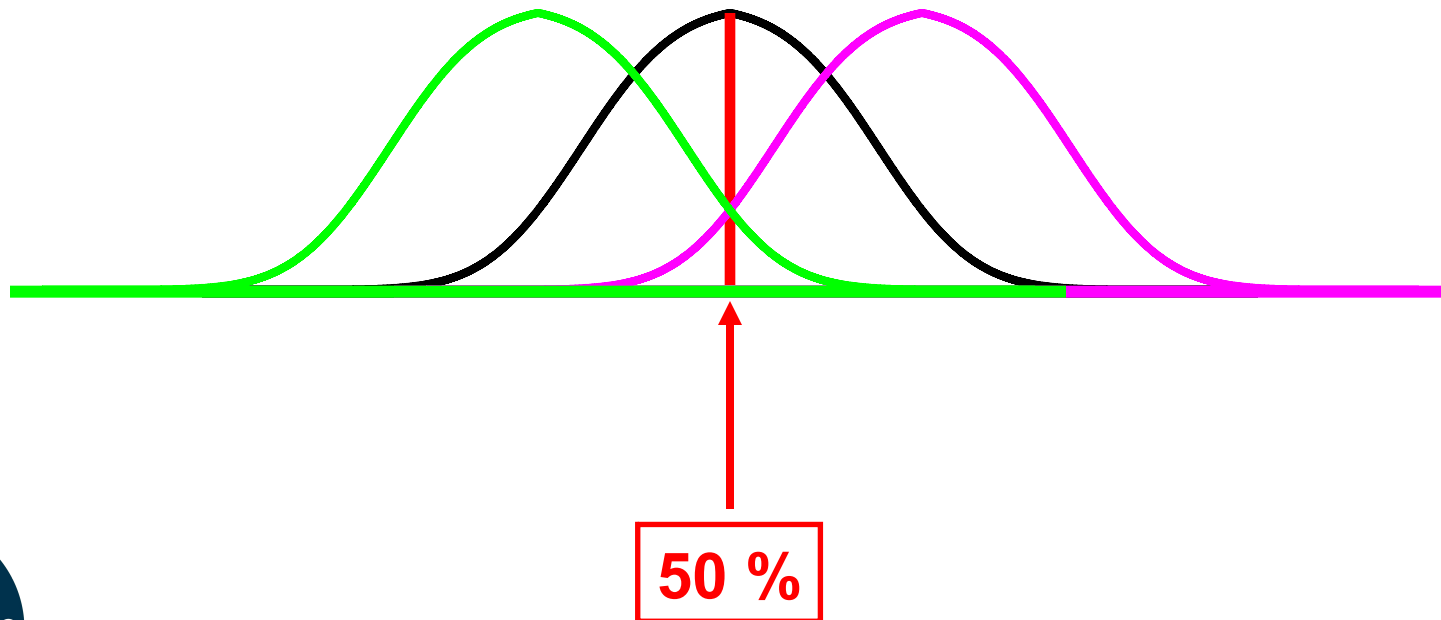


Criterion referenced standard

Test score distribution (average group)

Test score distribution (poor group)

Test score distribution (good group)



Absolute Methods: Judgments About Individual Test Items

- Methods
 - **Angoff's** method
 - **Ebel's** method

Angoff's method - I

- Select the judges
- Discuss
 - Purpose of the test
 - Nature of the examinees
 - What constitutes adequate/inadequate knowledge
 - The borderline candidate

Angoff's method - 2

- Read the first item
- Estimate the proportion of the borderline group that would respond correctly
- Record ratings, discuss, and change
- Repeat for each item
- Calculate the passing score

Ebel's Method - I

- Difficulty-Relevance decisions
 - Judges read each item and assign it to one of the categories in the classification table
 - They make judgments about the percentages of items in each category that borderline test-takers would have answered correctly
 - Calculate passing score

Ebel's method - 2

Easy

Medium

Hard

Essential

Important

Acceptable

Ebel's method - 3

	Easy	Medium	Hard
Essential	95%	80%	70%
Important	90%	80%	75%
Acceptable	80%	60%	50%

Ebel's Method

<u>Category</u>	<u>% Right</u>	<u># Questions</u>	<u>Score</u>
Essential			
Easy	95	3	2.85
Hard	80	2	1.60
Important			
Easy	90	3	2.70
Hard	75	4	3.00
Acceptable			
Easy	80	2	1.60
Hard	50	<u>3</u>	<u>1.50</u>
		17	12.25

Absolute Methods: Judgments About Individual Test Items

- **Advantages**
 - They focus attention on item content
 - They are relatively easy to use
 - There is a considerable body of published work supporting their use
 - They are used frequently in high stakes testing

Absolute Methods: Judgments About Individual Test Items

- **Disadvantages**
 - The concept of a "borderline group" is sometimes difficult to define
 - Judges sometimes feel they are "pulling numbers out of the air"
 - The methods can be tedious

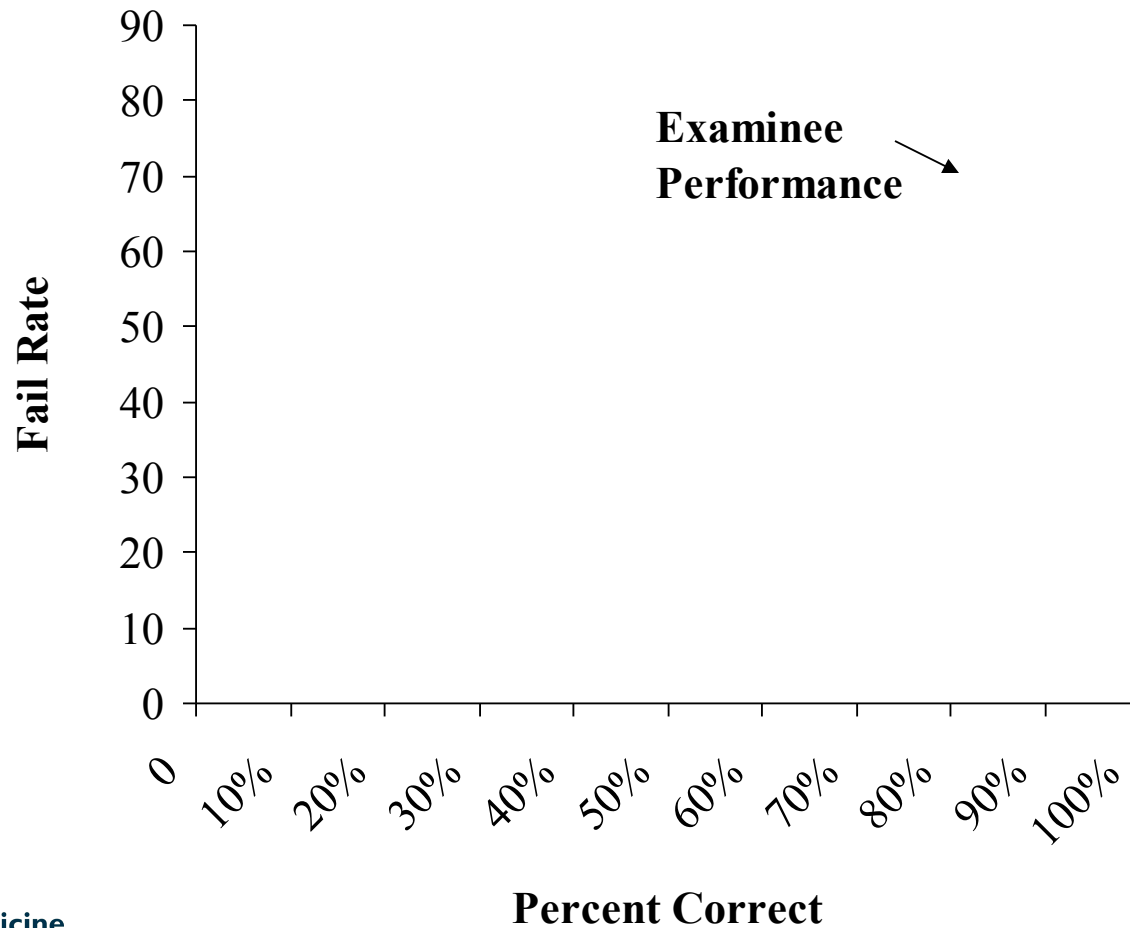
Compromise Methods

- **Hofstee Method**
 - Select the judges
 - Discuss
 - Purpose of the test
 - Nature of the examinees
 - What constitutes adequate/inadequate knowledge
 - Review the test in detail

Hofstee's method - I

- Ask the judges to answer four questions:
 1. What is the minimum acceptable cut score?
 2. What is the maximum acceptable cut score?
 3. What is the minimum acceptable fail rate?
 4. What is the maximum acceptable fail rate?
- After the test is given, graph the distribution of scores and select the cut score

Hofstee's method - 2



Compromise Methods

- **Advantages**

- Easy to implement
- Educators are comfortable with the decisions

- **Disadvantages**

- The cut score may not be in the area defined by the judges' estimates
- The method is not the first choice in a high stakes testing situation

Implementation Guidelines for Setting Standards

- Select the judges
 - Assign an appropriate number (at least 6-8 for high stakes testing)
 - Select the characteristics the group should possess
 - Develop an efficient design for the exercise

The choices

- There is no perfect standard setting method
- Make a decision based on the most important criteria for a particular circumstance

Practical implications

- Choice of standard setting methods depends on:
 - Credibility
 - Resources available
 - High stakes level of exam

Standard setting

- Not so much
 - the **METHOD** as the **PROCESS**
 - Suitable judges on the panel
 - Due diligence applied
 - Defensible rationale

References

- Berk, R.A. (1986). A consumer's guide to setting performance standards on criterion-referenced tests. *Review of Educational Research*, 56, 137-172.
- Cizek, G. J. (2001). *Setting Performance Standards: Concepts, Methods, and Perspectives*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Jaeger, R.M. (1989). Certification of student competence. In R.L. Linn (Ed.), *Educational Measurement*. New York: American Council on Education and Macmillan Publishing Company.
- Kane, M. (1994). Validating the performance standards associated with passing scores. *Review of Educational Research*, 64, 425-461.
- Livingston, S.A. and Zeiky, M.J. (1982). *Passing scores: A manual for setting standards of performance on educational and occupational tests*. Princeton, NJ: Educational Testing Service.

References

- Norcini, J.J. and Guille, R.A. (2002). Combining tests and setting standards. In Norman, G., van der Vleuten, C., and Newble, D. (Eds.): *International Handbook of Research in Medical Education* (pp. 811-834). Dordrecht: Kluwer Press.
- Norcini, J. J. (2003). Setting standards on educational tests. *Medical Education*, 37, 464-469.
- Norcini, J. J. & Shea, J. A. (1997). The credibility and comparability of standards. *Applied Measurement in Education*, 10, 39-59.
- Zeiky, M. J. (2001). So much has changed. How the setting of cutscores has evolved since the 1980s. In G.J.Cizek (Ed.), *Setting Performance Standards: Concepts, Methods, and Perspectives* (pp. 19-52). Mahwah, NJ: Lawrence Erlbaum Associates.