

# Setting the Standard in Examinations: How to Determine Who Should Pass

Mark Westwood

# Introduction

- Setting standards
  - What is it we are trying to achieve
- How to do it (methods)
  - Angoff
  - Hoftsee
  - Cohen
  - Others

# Types of examination

- Formative
- Summative
  - High stakes
- Viva type examinations
  - Can be subjective
- Essay type examinations
  - Model answers, time consuming
- MCQ's
  - Objective, probably least flawed

# Formative vs Summative

- Formative
  - Self feedback
  - To check learning at the end of a chapter
  - Low stakes
  - Low rigour
- Summative
  - High stakes
  - Consistency
  - Accuracy

# Formative vs Summative

- Formative

- Passmark can be relatively arbitrary

- “What is the purpose of the test?”

The key question is:

“What is the purpose of the test”

- Passmark critical

- Needs clear methodology

- Or liable to challenge

- No one single method

# Where to start: Basic Points

- Set the test (ideal world)
  - Meaningful/essential performance criteria
  - Candidates provide evidence by taking the test

Standard setting is a set task/process

It is imperfect

Therefore must be robust

Especially if tied to promotion/job prospects

- Full mastery not required (unrealistic)
  - Also medicine complex/imperfect
- Need a cut off between competent/non competent

# Standard setting: Reference

- Norm referenced (relative)

Norm group must be

Fixed pass rate

heterogenous, large

Rela

roup

est

- Criterion referenced (absolute)

This standard is fixed

Fixed pass mark

Can be partially re-evaluated over time

# Standard setting: Reference

- Norm referenced
  - Standard not content related
  - Fixed fail rate
  - Examinees ability influences standard

But its easier (generally)

- Diagnostic feedback relative to performance unclear



# Standard setting: Experts

- Expert panels
  - Need to be true experts!
  - Need to be familiar with test takers
  - Need to be familiar with exam methods

**It is easier to be harsh when setting standards!**

- Often higher than they use in practice
- Even borderline candidates
- Often pass mark set too high

# Standard setting: Multiple points

- Compensatory

May let candidate pass who lacks key skill

Generally less failures

Easier to administer

- Coniunctive

Candidate needs all key skills to pass

Generally more failures

Good candidates failing (sampling error)

—In effect multiple sub-passmark

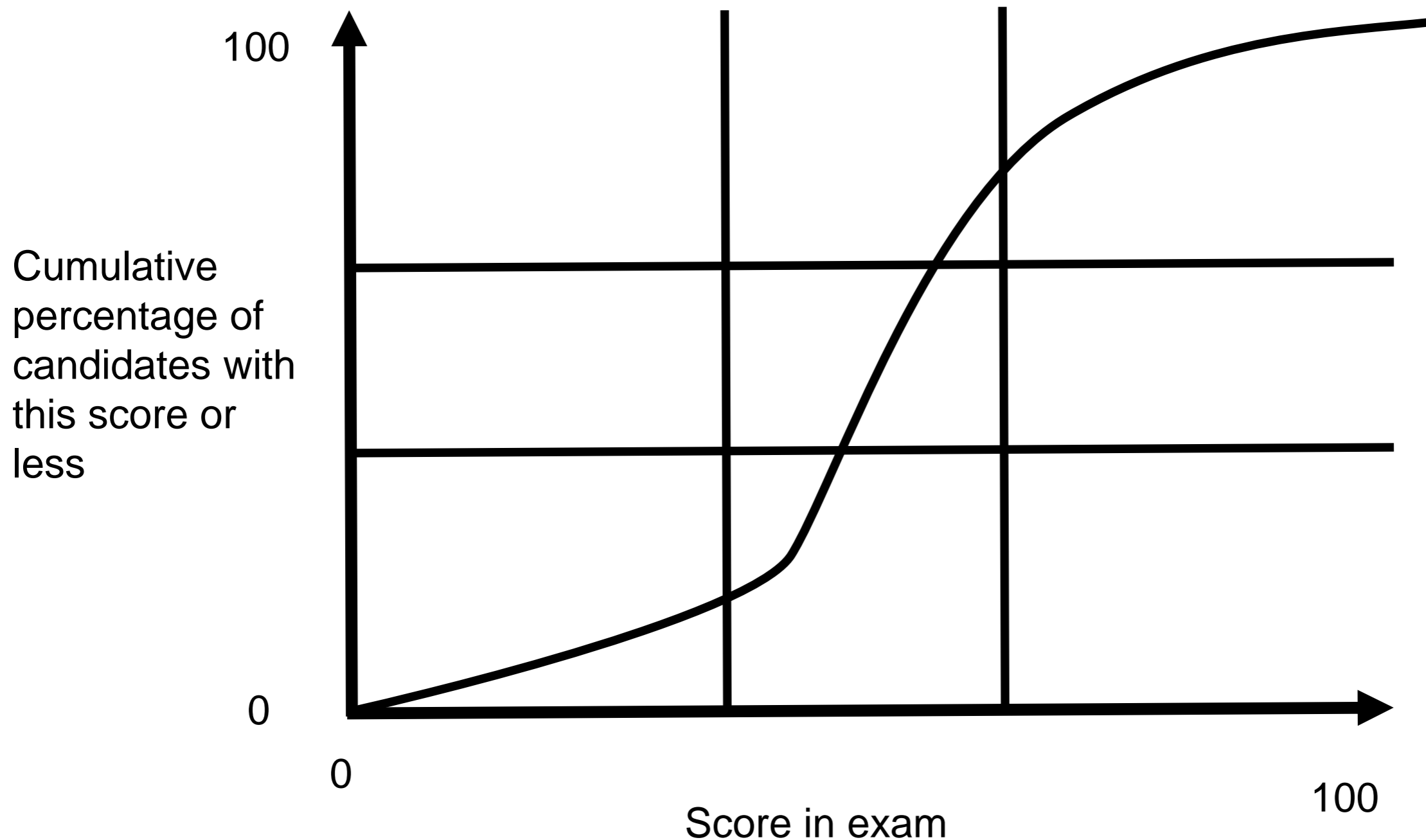
# Standard setting: Models

- Test centred
  - Judges set standards on test items
  - Provide judgements on the ‘just adequate level’ of performance these items
  - Eg: Angoff, Ebel, Nedelsky, Jaeger
- Examinee centred
  - Panelists make pass/fail by identifying a score consistent with test purpose
  - Eg: Borderline-Group, Contrasts by Group, Body of work

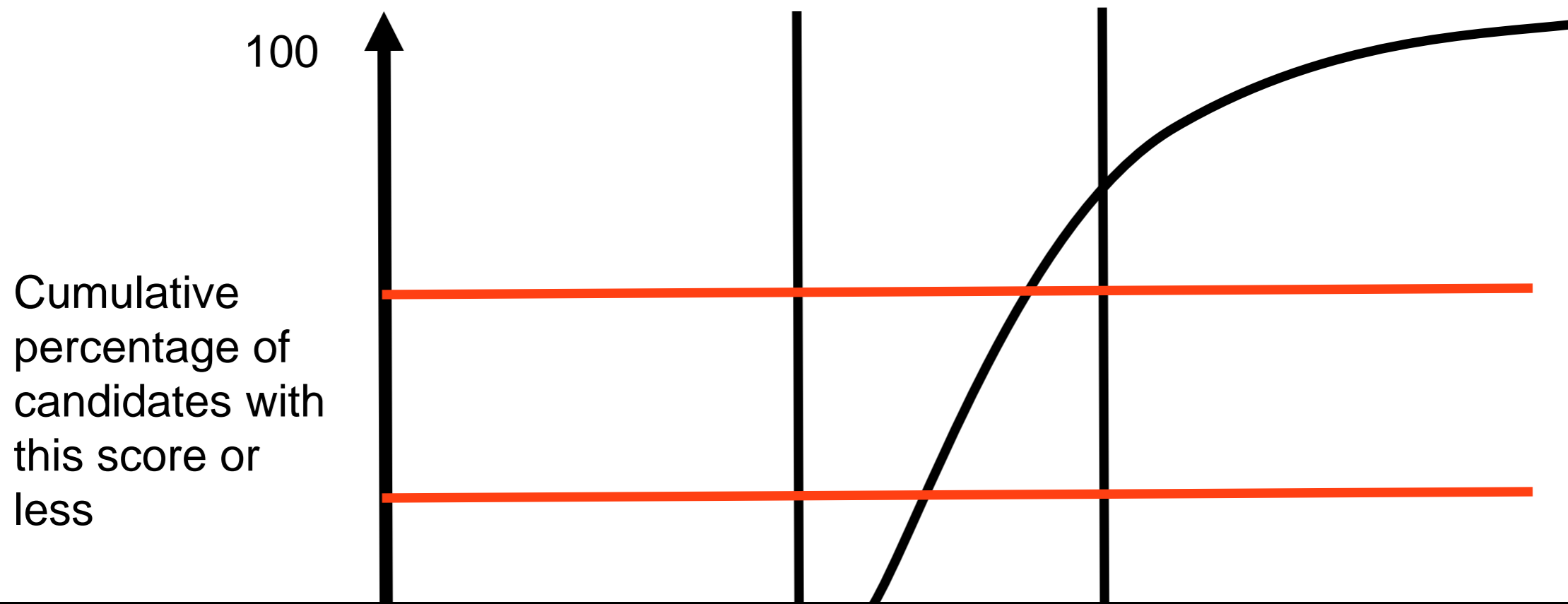
# Standard setting: Compromise

- Relative/Absolute compromise method
- Hoftsee
- Panelists determine 4 scores:
  - Minimum fail rate
  - Maximum fail rate
  - Minimum passing point
  - Maximum passing point
- Median of each taken

# Hoftsee: Setting the pass mark

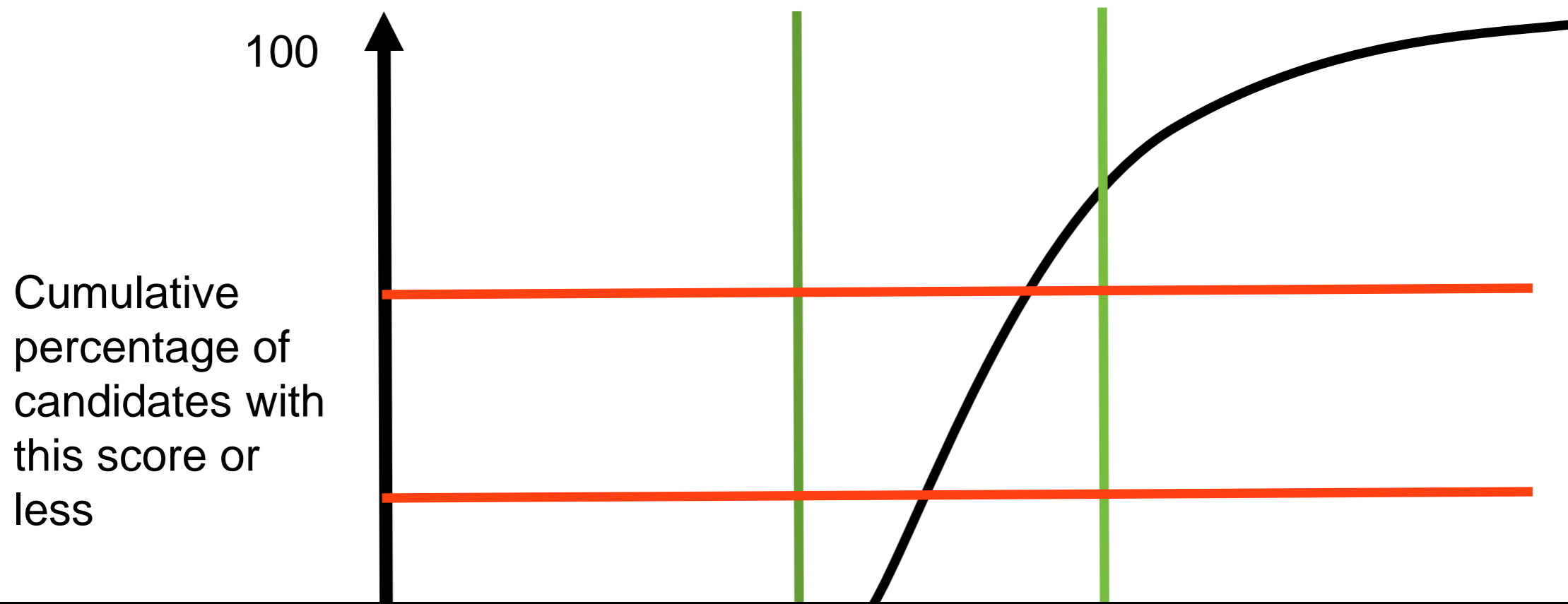


# Hoftsee: Setting the pass mark



The Red lines are the minimum and maximum failure rates as predetermined by the standard setting group

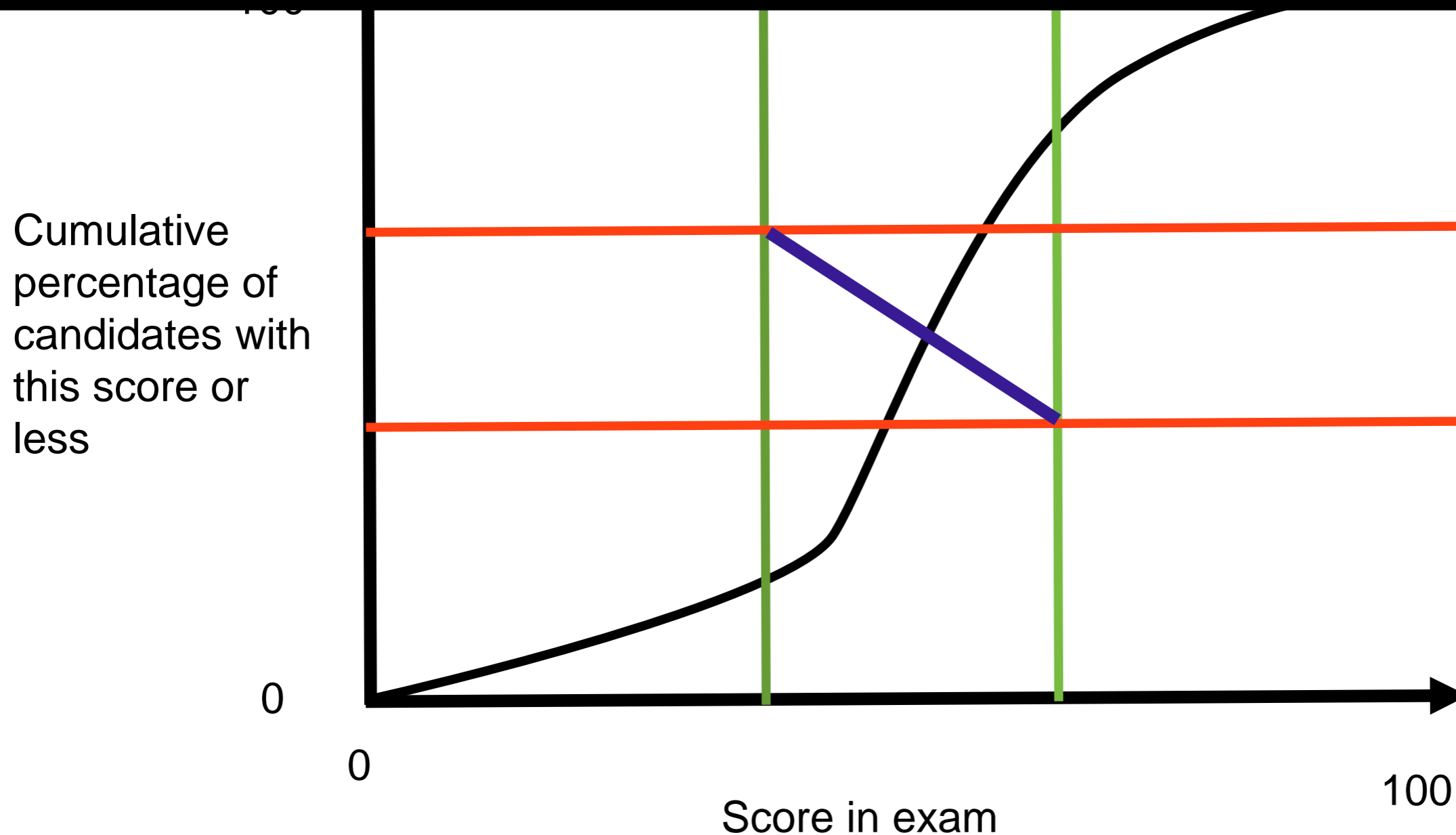
# Hoftsee: Setting the pass mark



The Green lines are the minimum and maximum pass marks as predetermined by the standard setting group

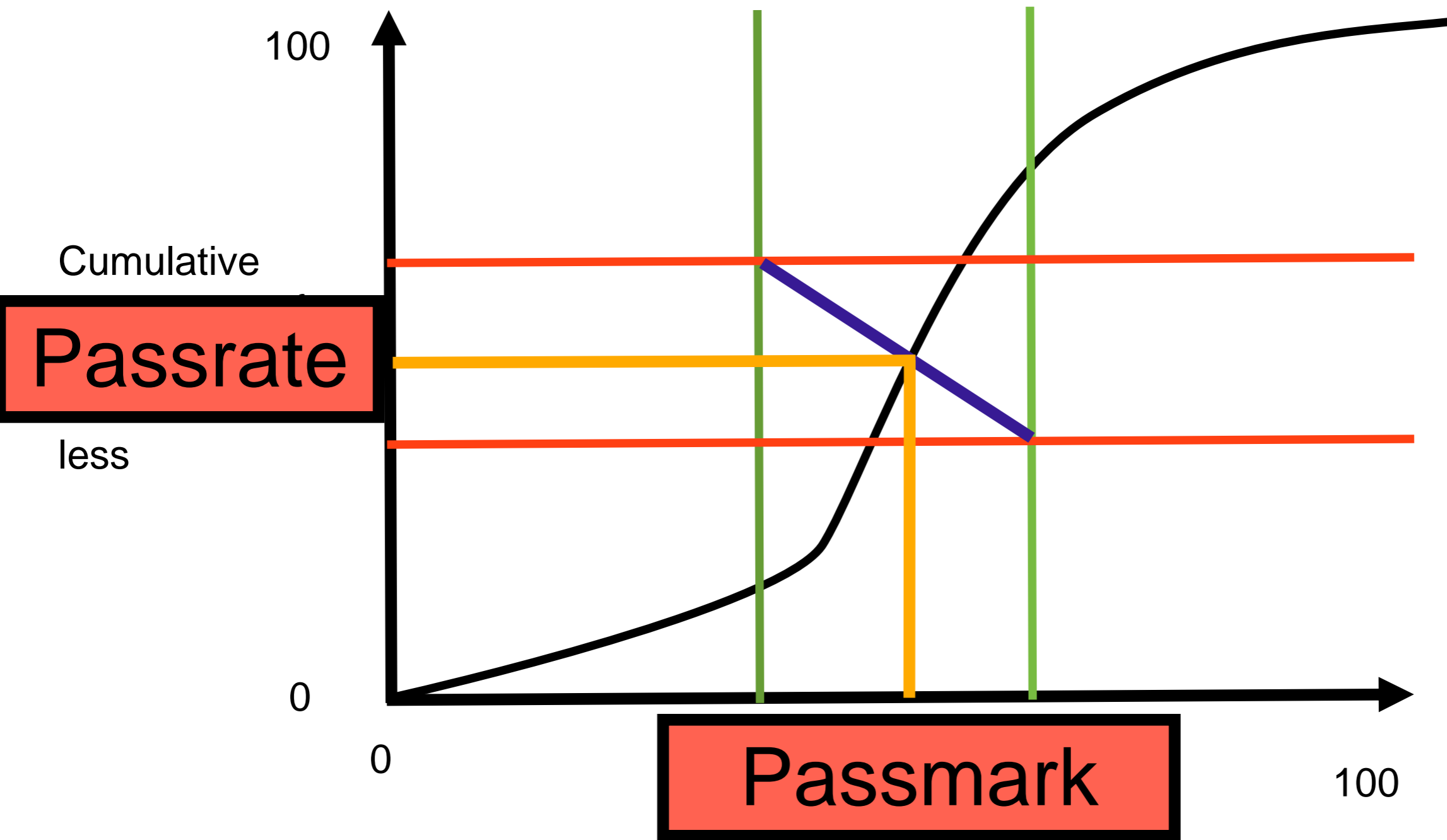
# Hoftsee: Setting the pass mark

Draw the blue line as shown





# Hoftsee: Setting the pass mark



# Standard setting: Compromise

- Relative/Absolute compromise method

Very quick and easy

Really a normative method

Based on performance of single examinee

Values reliability above all else

- Modified Cohen

- 90th centile

- Use cut score of exams set via modified Angoff

# Angoff: Setting the pass mark

- Judgemental approach
- Panel of experts
- The borderline candidate

~~Low proportion of borderline candidates will~~

Panel must be clear in advance of the characteristics of the borderline candidate

of answering a number of items correctly

- Average over judges
- Sum over content
- Cut score

# Modified Angoff: Setting the pass mark

- Provide item difficulty
  - Real performance data
  - Eg: after the examination
- Judges must be familiar
  - The content
  - The minimally competent test taker
    - Neither qualified/unqualified to pass test
  - Borderline candidates characteristics
  - Degree of difficulty of task
  - Tendency to be harsh

# Others: Setting the pass mark

- Ebel
  - Matrix
  - Difficulty (easy, medium, hard) vs
  - Relevance (essentail, important, acceptable)
- Nedelsky
  - How many distractors does the just passing candidate recognise as incorrect
- Jaeger
  - Multiple panels, iterative
  - Focus on passing the candidates

# Others: Setting the pass mark

- Borderline group
  - Examinee centred
  - Judgements of test takers not items
  - Uses judges global ratings
- Contrasting groups
  - Divide takers into 2 groups (pass/fail)
  - Standard is best discriminatory score
- Body of work
  - Similar to contrasting groups
  - Uses durable work (essays, portfolios etc)

# Settings standards: 8 steps

- Select standard setting method
- Select panel/judges
- Set performance criteria (pas/fail/dist.)
- Train judges
- Collect ratings/judgements
- Feedback to facilitate discussion
- Evaluate standard setting process
- Provide results/evidence to final decision makers

# Conclusion

- Setting standards
  - What is it we are trying to achieve

The key question is:

“What is the purpose of the test”

- Hotisee
- Cohen
- Others