

Rubrics and Standardization

Caren M. Stalburg, MD MA
Clinical Assistant Professor
Obstetrics and Gynecology and Medical Education



Unless otherwise noted, this material is made available under the terms of the
Creative Commons Attribution Share Alike-3.0 License: <http://creativecommons.org/licenses/by-sa/3.0/>

Intended Learning Outcomes

- Understand how to create scoring rubrics
- Become familiar with standard setting techniques

SCORE = performance on the assessment

STANDARD = acceptable score to
indicate desired level of performance

Scoring Rubrics

	Performance				
Assessed domain	Descriptor				
		Novice	Intermediate	Competent	
	Historical component	Obtains most relevant patient history, missing key components	Most of relevant patient history, most key elements, not all	Includes all relevant components of patient history	
	Score	○ ○ ○ 1 2 3	○ ○ ○ 4 5 6	○ ○ ○ 7 8 9	

VERBAL SKILLS

	5	4	3	2	1
Clarity	Clear, easy to understand	Occasionally difficult to understand	Audience must put forth effort to listen, poor pronunciation	Unclear, difficult to understand	Loses audience's attention due to lack of clarity
Evaluation					
Volume	Easy to hear, doesn't overpower audio equipment	Overall appropriate, some sentences trail off or are hard to hear	Generally audible, often hard to hear	Difficult to hear, poorly positioned audio equipment	Generally inaudible, not using audio equipment
Evaluation					

Standards

- What do the scores ‘mean’
- Thoughtful judgment by experts
 - Content of the assesement/exam
 - Purpose of the exam: stakes?
 - Criteria that can be explained and justified
 - Understand the learners or group being tested

Cut-off scores

- The number below which performance is deemed unacceptable.
 - Can have significant ramifications for the individuals and for you.
- **Relative standards** = Norm-referenced, based on the performance of a group
 - Exam mean is set as a C, bottom 10th percentile fails. If the mean is 60%, people still pass
- **Absolute standards** = Criterion based
 - Independent of group performance. If learner gets 70% of exam questions correct, then that demonstrates they have mastered enough of the material to have an adequate performance. If none of the class scores 70% then no one passes.

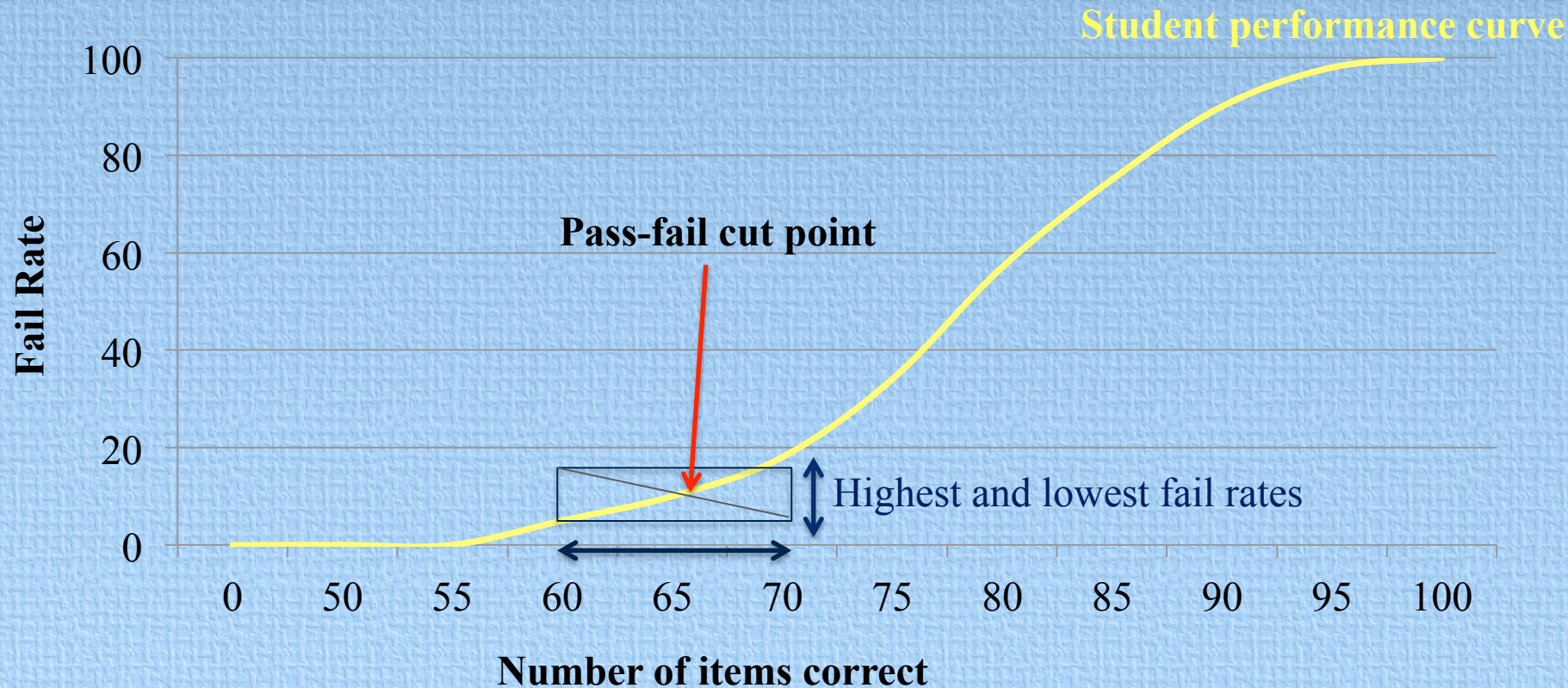
Angoff Method: test centered

- How will a borderline student perform?
- Panel of expert judges each question
- Gives an estimate of the chance that a **BORDERLINE** student would answer correctly
- Average the estimates of the raters per item
- Average estimate for each item is then averaged across the exam providing a cut-off score

Hofstee method

- Judges make 4 judgments:
 - lowest & highest acceptable passing scores
 - lowest & highest acceptable fail rates
- Passing score = mean of 4 judgments plotted against cumulative score distribution

Hofstee



Cut-points

- May have unique ramifications
 - Too many passes
 - False-positives: you pass those who shouldn't
 - Practical aspects of administering a test
 - Dependent on 'rectangle' and 'curve' of student performance line actually intersecting....

Standard setting

- Difficult to do
- The higher the stakes, the more intentional the process should be
- Defensible, reproducible, contextual
- May need to be revisited over time