Standard Setting

รศ.นพ.เชิดศักดิ์ ไอรมณีรัตน์

ภาควิชาศัลยศาสตร์

คณะแพทยศาสตร์ศิริราชพยาบาล

Standard

- A score that is set to be a boundary between those who perform well enough on the test (pass) from those who do not (fail).
- Standard = cutpoint

Question

- ในการสอบที่อาจารย์เกี่ยวข้อง อาจารย์ใช้วิธีการใดในการกำหนดเกณฑ์ผ่าน
- วิธีการตั้งเกณฑ์ผ่านดังกล่าว ก่อให้เกิดปัญหาหรือไม่ อย่างไร

Objectives

- เมื่อสิ้นสุดการบรรยายแล้ว ผู้เข้าอบรมสามารถ
  - บอกถึงความสำคัญของการตั้งเกณฑ์ผ่านได้ถูกต้อง
  - บอกถึงขั้นตอนของการตั้งเกณฑ์ผ่านได้ถูกต้อง
  - ยกตัวอย่างวิธีการตั้งเกณฑ์ผ่านได้อย่างน้อยสามวิธี
  - จัดทำเกณฑ์ผ่านการสอบ ด้วยวิธีการ modified Angoff method ในการสอบที่ตนเกี่ยวข้องได้อย่างเหมาะสม

Outline

- Basic concepts
- Steps in setting standards
  - The type of standard
  - The method
  - Selecting judges
  - Standard setting meeting
  - Calculate the standards
  - Checking the standards

Basic Concepts

- A standard is an answer to the question, “How much is enough?”
- The classification of examinees into two groups can result in two types of wrong decisions
  - False positive: Passing an examinee who should fail the exam
  - False negative: Failing an examinee who should pass the exam

Judgment

1. Made by qualified judges
2. Meaningful to the persons who are making the decision
3. Made in a way that takes into account the purpose of the test

Steps in Setting Standards

1. Deciding on the type of standard
2. Deciding on the method for setting standards
3. Selecting judges
4. Holding the standard setting meeting
5. Calculating the standards
6. Checking the standards after test

1. Types of Standards

- Absolute standard
- Relative standard

Absolute Standard

- The standard is fixed, based on specific criteria of performance, but may undergo periodic re-evaluation of the standard
- Strengths
  - A standard is known in advance
  - A stable performance level is required to pass the examination => content-related standard
  - Provide clear feedback to examinees
  - Nobody has to fail the exam if their knowledge/skills is adequate for the purpose of the exam.
  - Promote a collaborative learning environment.

Relative Standard

- The standard is set in reference to the group of examinees. The resulting standard may be reasonable providing a representative heterogeneous group.
- Strengths
  - The failure rate is stable, which in someway is easy for curriculum management

2.Methods for Setting Standards

1. Test-centered methods
2. Examinee-centered methods
3. Compromised methods

Test-Centered Methods

- The judges set standards by reviewing the test items and provide judgments regarding the “just adequate” level of performance on these items.
  - Angoff’s method
  - Nedelsky’s method
  - Ebel’s method

Modified Angoff's Method

- The judgment
  - The probability that a borderline examinee would answer the test item correctly
- The passing score
  - The sum of all the probability of correct answers for all items on the exam

Modified Angoff's Method (2)

Item	Probability
1	0.8
2	0.6
3	0.4
4	0.5
5	0.5
Passing score	2.8

Nedelsky's Method

- The judgment
  - How many options a borderline examinee can eliminate from choosing in an item
- The passing score
  - The probability of correct answer for an item = 1/(the number of options not eliminated)
  - The passing score of the test = the sum of all the probability of correct answers of all items on the test

Nedelsky's Method (2)

Item	A	B	C	D	E	Not eliminated	Probability
1			X	X	X	2	1/2 =0.50
2	X	X				3	1/3 = 0.33
3	X					4	1/4 = 0.25
4	X		X	X		2	1/2 = 0.50
5	X				X	3	1/3 = 0.33
Passing score							1.91

Ebel's Method

- The judgment
  - What is the level of difficulty of an item?
    - Easy/Medium/difficult
  - What is the level of importance of that content in clinical practice?
    - Essential/Important/Acceptable/Questionable
  - The probability that a borderline examinee will answer an item in each category correctly
- The passing score
  - The sum of all the probability of correct answers for all items on the exam

Ebel's Method (2)

	Easy	Medium	Difficult
Essential	0.95	0.85	0.80
Important	0.90	0.75	0.60
Acceptable	0.85	0.60	0.40
Questionable	0.55	0.45	0.35

Ebel's Method (3)

Item	Difficulty	Importance	Probability
1	Easy	Essential	0.95
2	Easy	Importance	0.90
3	Difficult	Essential	0.80
4	Difficult	Acceptable	0.40
5	Medium	Acceptable	0.60
Passing score			3.65

Examinee-Centered Methods

- The judges set a standard by reviewing the overall performance of examinees and determine who should pass and who should fail. The scores of examinees are reviewed and the passing score is set based on these judgments
  - Borderline-group method
  - Contrasting-groups method
  - Borderline regression method

Borderline-Group Method

- The judgment
  - Identify examinees who are “borderline”
- The passing score
  - The median score of this “borderline group”

Contrasting-Groups Method

- The judgment
  - Identify examinees who should “pass” and those who should “fail”
- The passing score

Borderline Regression Method

- Judgment
  - What is the level of performance of this examinee?
- Linear regression

Compromised Method

- Combining relative and absolute standard setting methods
  - Hofstee method

Hofstee Method

- The judgment
  - Minimum failure rate
  - Maximum failure rate
  - Minimum passing score
  - Maximum passing score
- The passing score
  - The intersection of test scores curve with diagonal line drawn from upper left to lower right corner

3. Selecting Judges

- The number of judges
- The qualification of judges

4. Standard Setting Meeting

- Discussion of the purpose of the test, the characteristics of examinees, and the nature of competence.
- Explanation of the method and practice before the real standard setting procedure.

5. Calculating Standard

- Outliers
- Errors of the cutpoint

Do we have to care about error?

- True score theory
  - Each student has a true score, a hypothetical value representing a score free of error.
  - If we test a student repeatedly, the average of the obtained scores would approximate the true score, with a standard deviation of SEM.

SEM

$$SEM = SD\sqrt{(1-r)}$$

SD = standard deviation

r = internal consistency reliability

↑SD (more spread of score): higher SEM

↑r (more accurate measures): smaller SEM

What should we do with students with an SEM around cut score?

- False positive: Passing students who should have fail the examination
- False negative: Failing students who should have pass the examination

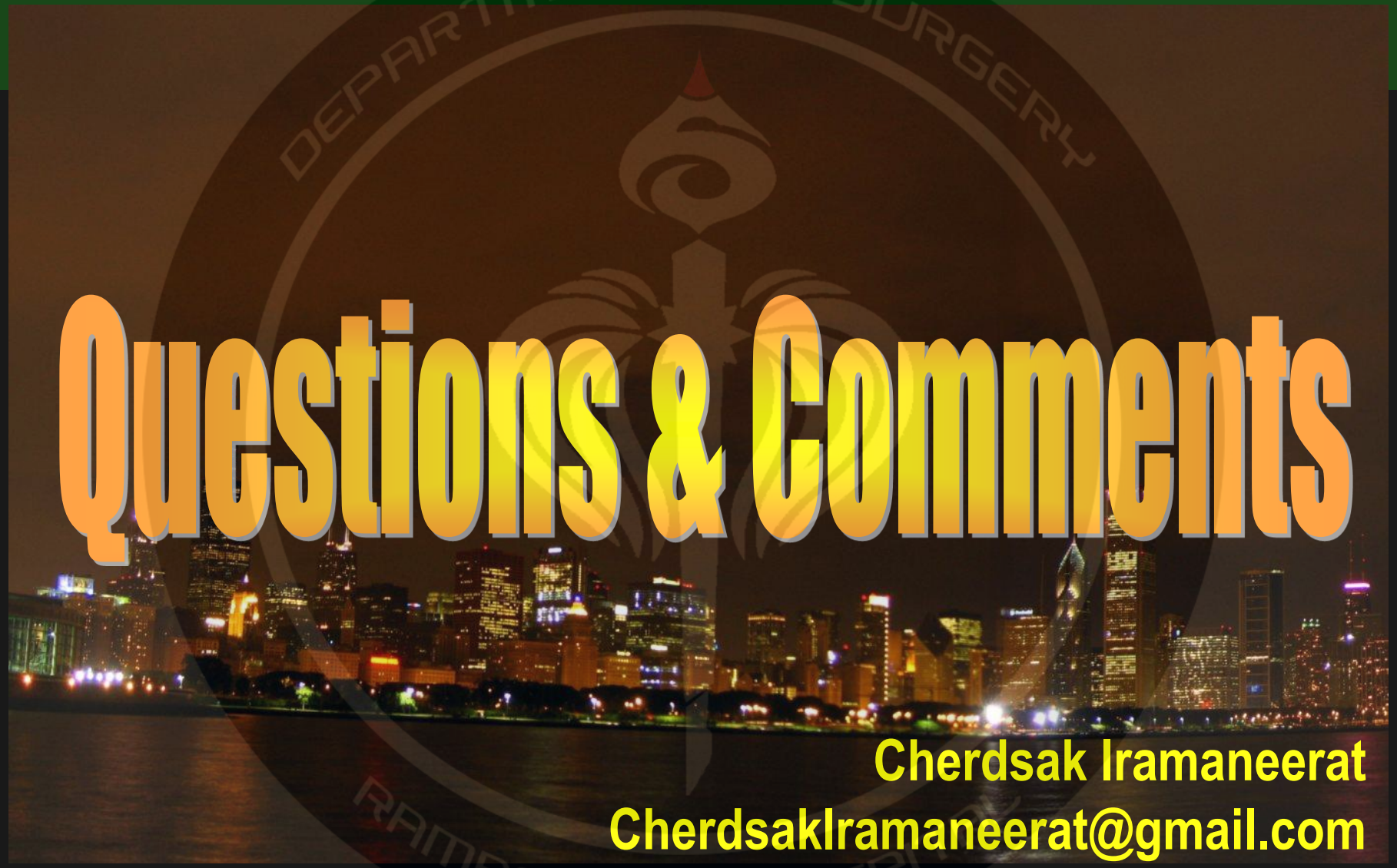
6. Checking Standard

- Stakeholders' acceptance of the results
- Relationship with other markers of competence
- Prediction of future performance

Summary

- Steps in setting up a standard

1. Deciding on the type of standard
2. Deciding on the method for setting standards
3. Selecting judges
4. Holding the standard setting meeting
5. Calculating the standards
6. Checking the standards after test



Cherdsak Iramaneerat
CherdsakIramaneerat@gmail.com

"It does not matter how slowly you go, as long as you do not stop."

Confucius