Exam Analysis

Dr Ghaffarifar S, M.D, M.M.Ed, D.H.E.P
Associate professor of Medical Education
Tabriz University of Medical Sciences

Tabriz Medical Education Research Center

Tabriz Uiversity of Medical Sciences

Learning objectives

After this presentation, you will be able to:

1. Discuss the importance of exam analysis
2. Differentiate components of exam analysis
3. Criticize the role of exam analysis in quality education
4. Describe your plan for effective exam analysis

Key goal of assessment in ME

"Minimization of all errors influencing a test in order to produce an observed score which approaches a learner's 'true' score, as reliably and validly as possible"

Objective tests

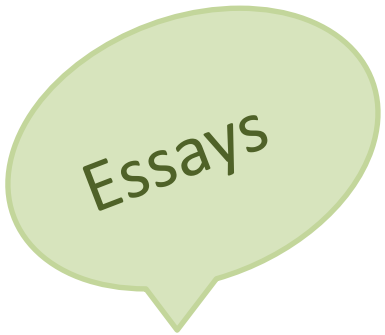
Multiple choice questions (MCQs)

Objective structured clinical examinations (OSCEs)

Direct observation of procedural skills (DOPS)

Mini-clinical examination (mini-CEX)

Subjective interpretations of performance



Item-difficulty(facility) index=P

The percentage of the total number of students who answered the test question correctly

Ranges from 0 (if no one answered a question correctly) to 1 (if everyone answered a question correctly

The larger the value, the easier the question

Item-difficulty(facility) index=P

$$\frac{\text{Correct choices}\left(n_{\text{top group}} + n_{\text{bottom group}}\right)}{N_{\text{top group}} + N_{\text{bottom group}}} * 100$$

- Between 0.3 and 0.8: the question is considered to be a good question

Question variance= P* (1-P)
Max=0.25 when P=0.5

Item-discrimination index=d

"A value of how well a question is able to differentiate between students who are high performing and those who are not, or between 'strong' and 'weak' students"

Dividing students into two groups ('high' and 'low')
"27% of the students are categorized as a strong group and 27% as a weak group"

The range of d is 1.00 to 1.00

Item-discrimination index=d

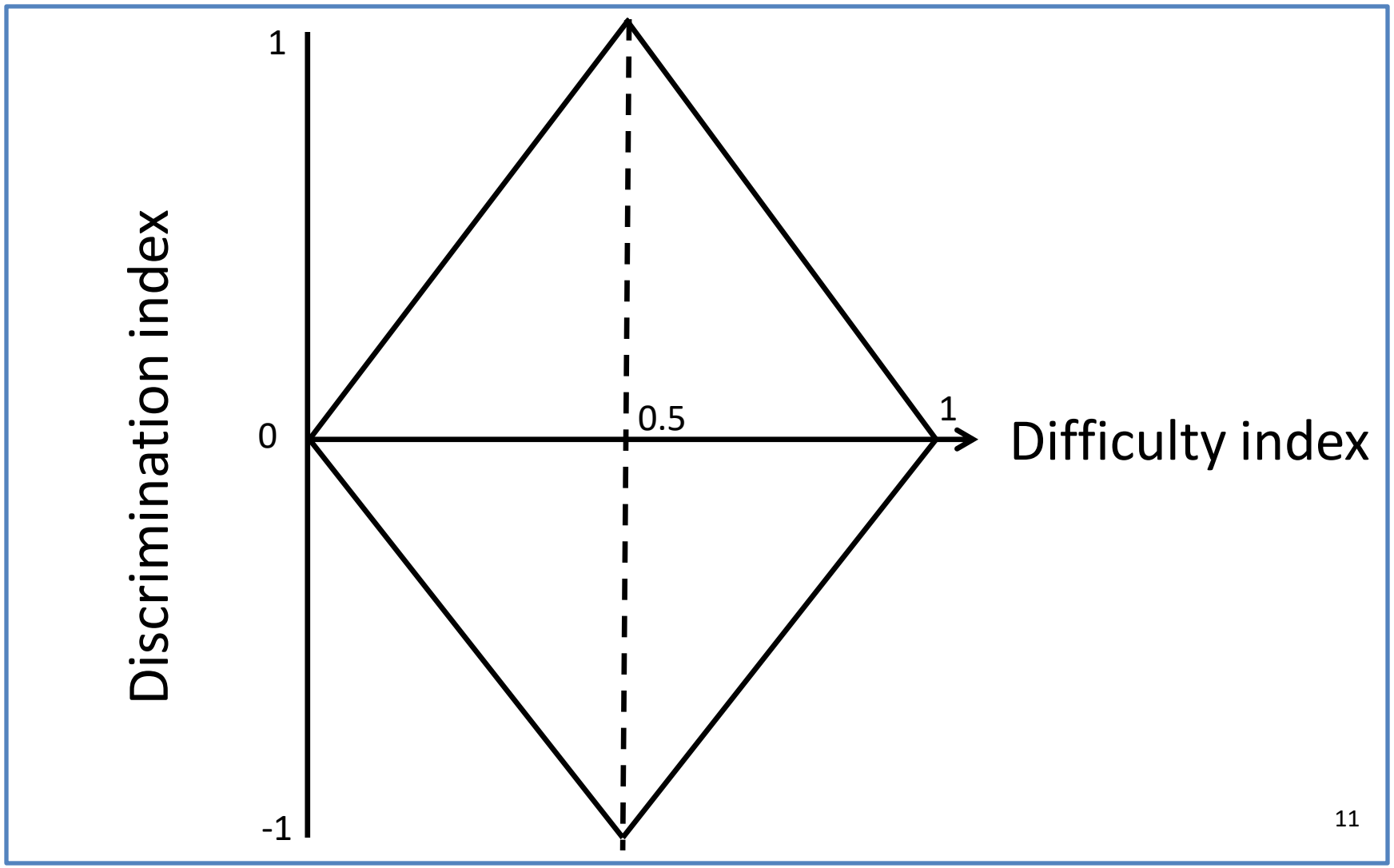
$$\frac{\text{Number of correct choices}(^{n}\text{ top group } - {}^{n}\text{ bottom group})}{{}^{N}\text{ one group}}$$

- d less than 0.3: item revising or discarding
- Negative d: ‘strong’ students answered the question incorrectly & the ‘weak’ students answered it correctly

Interpretation of discrimination index

d	Interpretation
>= 0.4	Excellent
0.3- 0.39	Good
0.11 – 0.29	Moderate
0.00 – 0.1	Poor
Negative	Faulty

Relationship between difficulty and discrimination indexes



Reliability assessment

Point-bi- serial correlation co-efficient

Cronbach's Alpha

Kuder-Richardson (KR20, KR21)

Reliability of exam scores is always less than inter-rater reliability

At least 0.7

The point bi-serial correlation coefficient

"Relationship between a particular question (correct or incorrect) on a test and the total tests score"

Range from -1.0 to +1.0

The higher the value, the better the question is discriminatin

CC= 0.4

$4^2 = 16\%$ of the exam scores variance

The standard Error of measurement (SEM)

- To determine the discrepancies between an individual's observed score on the test and his/her true score

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

Analyzing distractors

Each distractor should be chosen by at least one student in the bottom group

Each distractor should be chosen by students in the bottom group more than the peers in the top group

Important point about measures of exam analysis

All the rules about norm-referenced questions

Analysis of essays and performance scores

Marking the responses in order for all the learners

P= Mean score of the question / (Upper score- lower score)

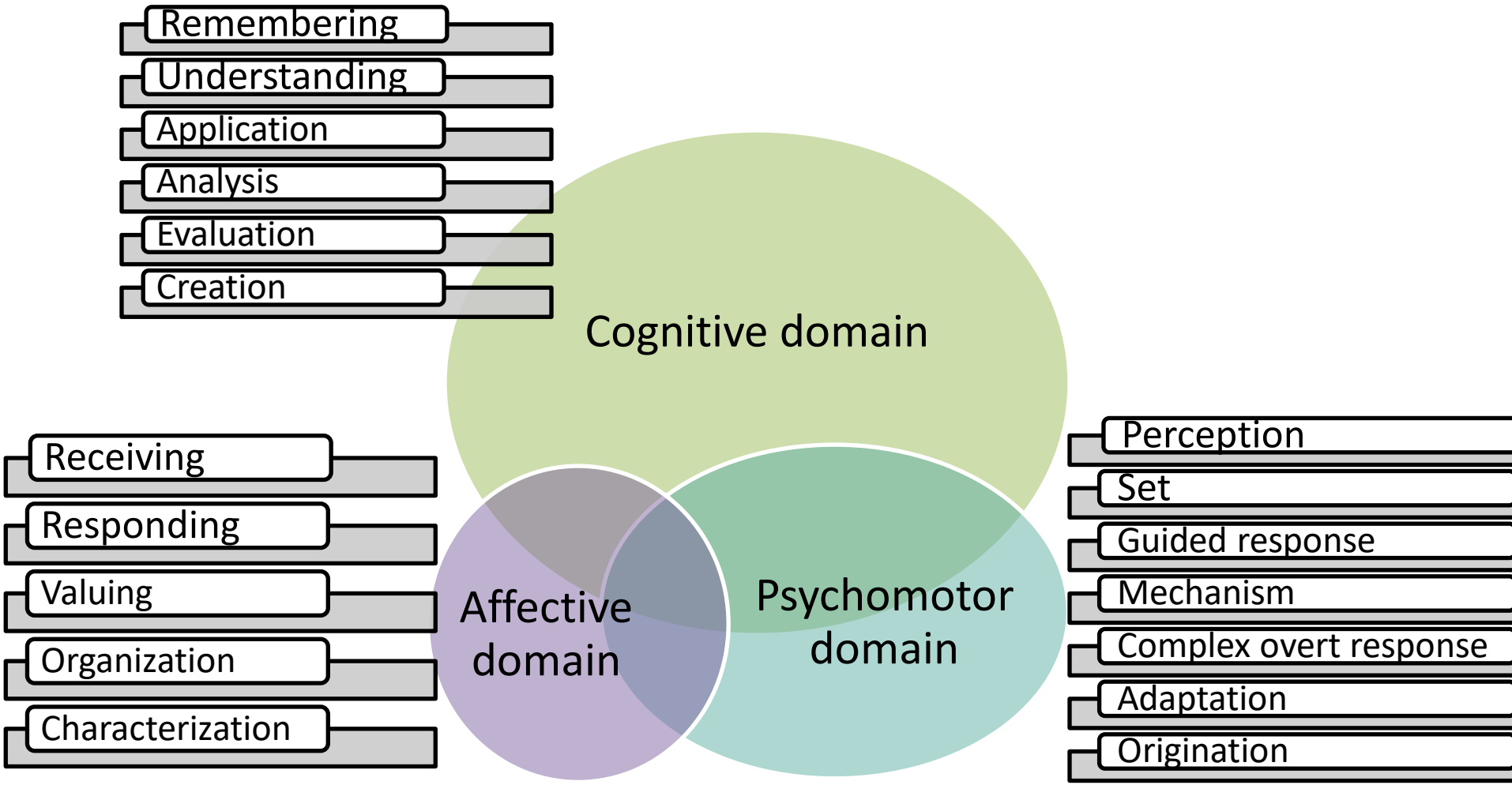
d= (Mean of U group – Mean of L group) / Upper score – lower score)

The importance of quality analysis

"Obtaining and reporting a numerical index has no meaning in itself unless we interpret and value the index"

(Morrow et al. 2006).

Classification of behavioral outcomes by Bloom (1956)



A sample blueprint

	Examination cycle	Psychometrics	Quality analysis	Reflective critique
Remembering	1	1	1	
Understanding		2	2	
Application		1	1	1
Analysis		1		
Evaluation			1	
Creation			1	1

A blueprint for an OSCE

	History	Examination	Health promotion	Practical skills
Cardiovascular	History of palpitations			ECG interpretation
Respiratory	History of breathlessness		Smoking cessation advice	
Gastro		Abdominal examination	Explain high-fibre diet	
Neuro		Gait examination		Lumbar puncture on manikin

Assessment of a blueprint

	Examination cycle	Psychometrics	Quality analysis	Reflective critique	Others
Q1	*				
Q2		*			
Q3					*
Q4			*		
Q5		*			
Q6		*			
Q7				*	
Q8					*
Q9			*		
Q10				*	
Q11					*
Q12			*		
Q13					*
Q14					*

Question writing

Related to a defined learning outcome

Clear and unambiguous

Case and Swanson techniques of question writing

The order of questions in exam drafting

1. True-false
2. Matching items
3. MCQs
4. Short answer
5. Essay

Exam drafting

From easy to hard questions in each category

Appropriate organization of items according to exam chapters

Exam's proper guide: duration, question values, penalties, maximum choices

Planning for good implementation

Piloting of assessments

To eliminate any gross content problems

'leakage' of the exam content

An external examiner's advice on the ease, difficulty or appropriateness of questions

Using parallel-forms of the test

Standard setting

Determining a pass-mark

Determining standard, which is the minimum adequate level of performance, indicating the boundary between those who perform acceptably and those who do not

A number of different standard setting methods

Define a fairer and more appropriate pass mark or cut-score by standard setters

Item sampling

The size of many exams is based on tradition

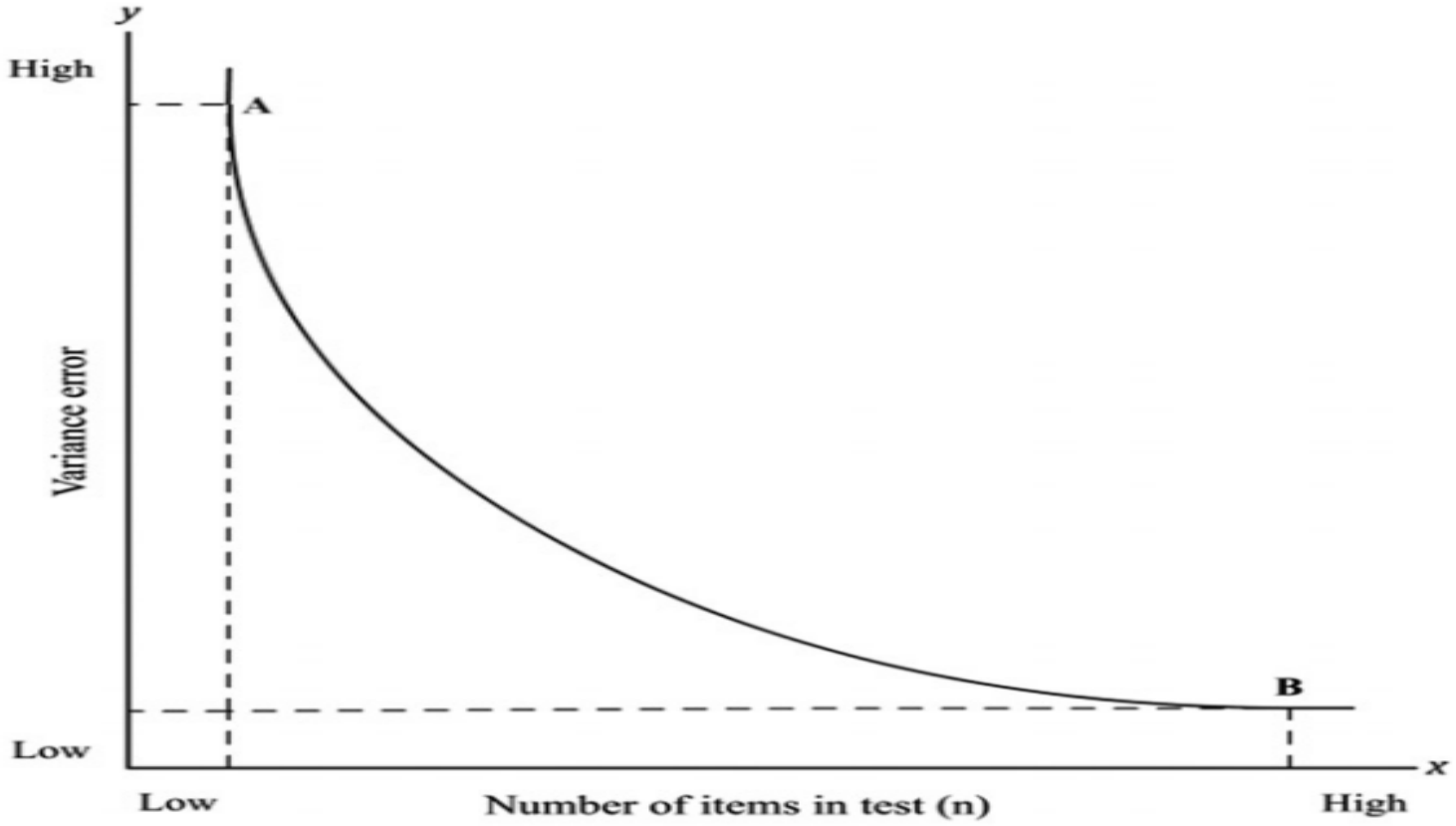
length of time

Appropriate sample size

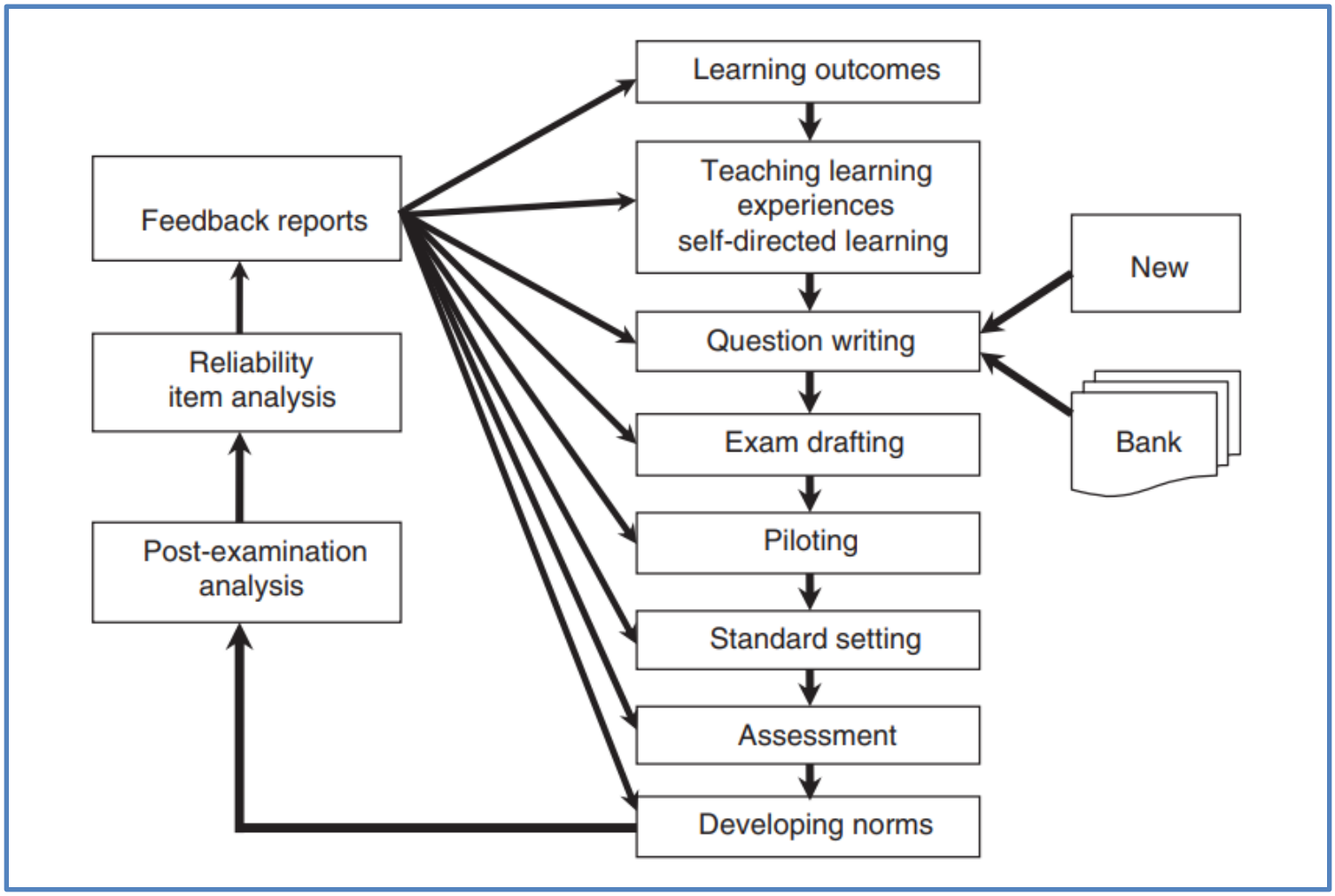
Z2 is a **confidence level** indicating how much the sample size is influenced by chance (1.64 for 90% confidence, **1.96 for 95%** and 2.57 for 99%), **SD** is an estimation of **standard deviation** in the population of items, **e2** is the error of the sample size, e.g., **0.03 or 0.05**.

$$n = \frac{Z^2(SD)^2}{e^2}$$

Relationship between the number of items in the test and variance error



The examination cycle



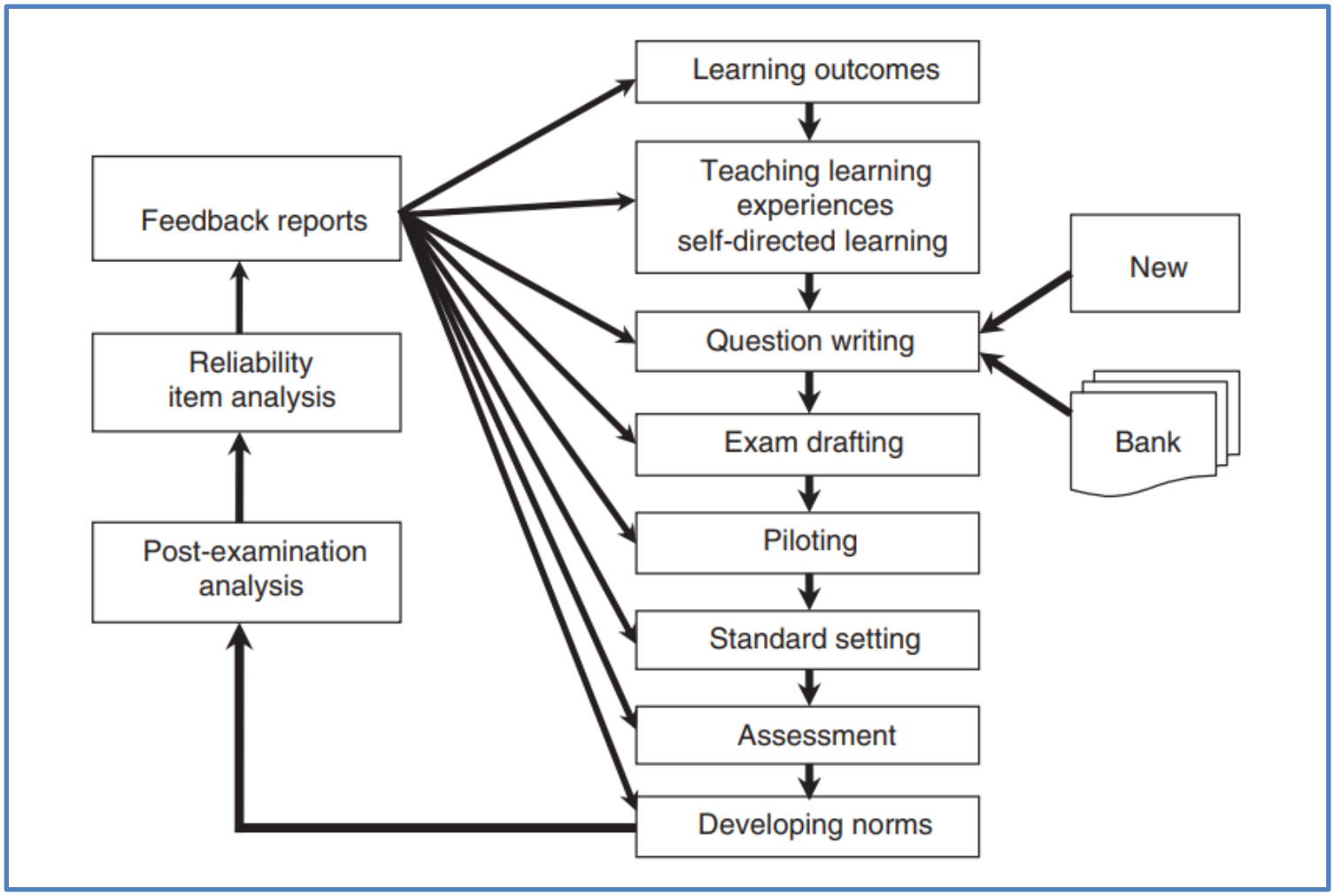
Summary

Listing potential biases that can influence all components of the assessment cycle from question creation to the interpretation of exam scores

Analyzing the data generated by assessments

Proper interpretation of measures of exam analysis in order to revise or discard the items or questions

Summary


Learning outcomes
Teaching learning experiences self-directed learning
Question writing
Exam drafting
Piloting
Standard setting
Assessment
Developing norms
New
Bank
Feedback reports
Reliability item analysis
Post-examination analysis

References

1) Tavakol M, Dennick R. Post-examination analysis of objective tests. Medical Teacher. 2011 Jun 1;33(6):447-58.
2) Tavakol M, Dennick R. Post-examination interpretation of objective test data: monitoring and improving the quality of high-stakes examinations: AMEE Guide No. 66. Medical teacher. 2012 Mar 1;34(3):e161-75.
3) Tavakol M, Dennick R. Post-examination interpretation of objective test data: monitoring and improving the quality of high–stakes examinations–a commentary on two AMEE Guides. Medical teacher. 2012 Mar 1;34(3):245-8.
4) Fino E, Hanna-Khalil B. Psychometric Post-Examination Analysis in Medical Education Training Programs. InHandbook of Research on the Efficacy of Training Programs and Systems in Medical Education 2020 (pp. 221-242). IGI Global.
5) Google images


Really appreciate your attention.