

# External Perspective on Analysis Results

- UMP Graduation Exam & Licensing Exam -

**Youkyoung Oh**

Department Medical Education  
Yonsei University College of Medicine

*Severance*

# Contents

- 1** What is a Test?
- 2** UMP Graduation Exam
- 3** Licensing Exam
- 4** Conclusion

65

**USMLE**  
(United States Medical Licensing Examination)

**KMLE**  
(Korean Medical Licensing Examination)

**65**

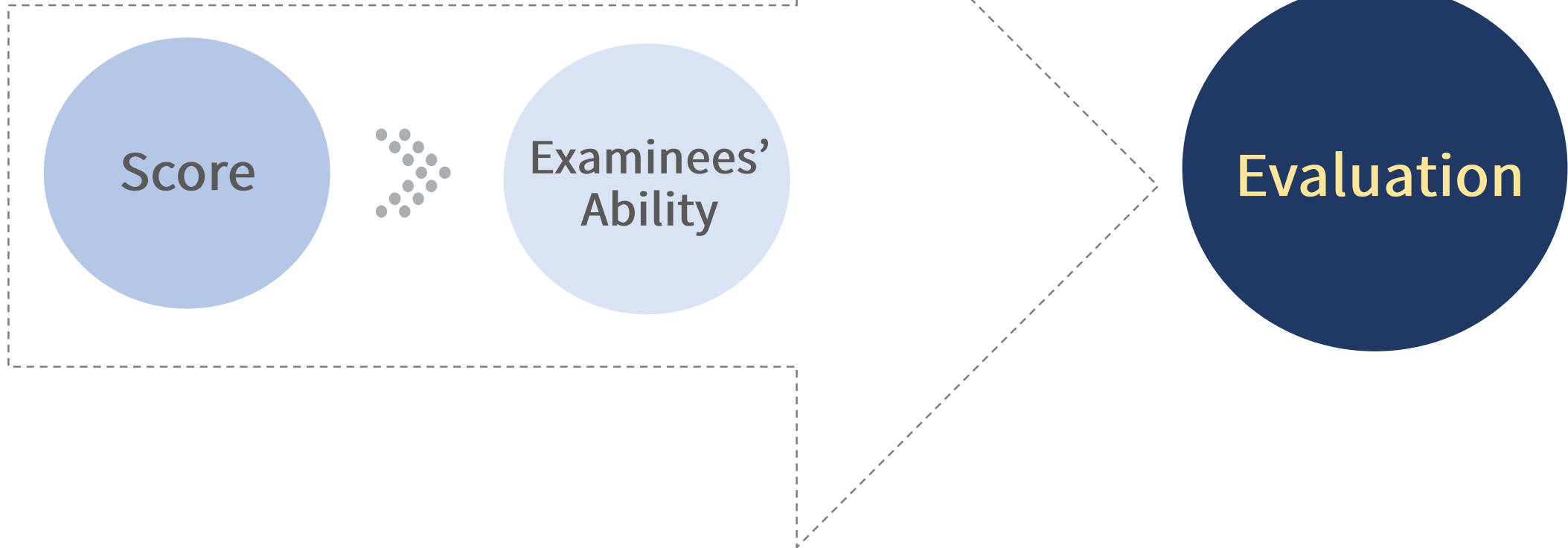
**UMP Graduation Exam**

# 65

*“How are the scores interpreted?”*

# What is a test?

with Competency-Based Learning,



- ◆ The key to test development is..

**Accuracy**

**Fairness**

# What is a test?

- ◆ The key to test development is..

Ensures fair outcomes for all students

**Accuracy**

**Fairness**

Accurately measures the student's ability



# What is a test?

◆ To check Accuracy & Fairness,

## UMP Graduation Exam

Checking the quality of test,

- ① Item Difficulty
- ② Item Discrimination
- ③ Test reliability *or*  
Item reliability

## Licensing Exam

Checking the quality of test,

- ① Item Difficulty
- ② Item Discrimination
- ③ Test reliability *or*  
Item reliability
- ④ Test Equating

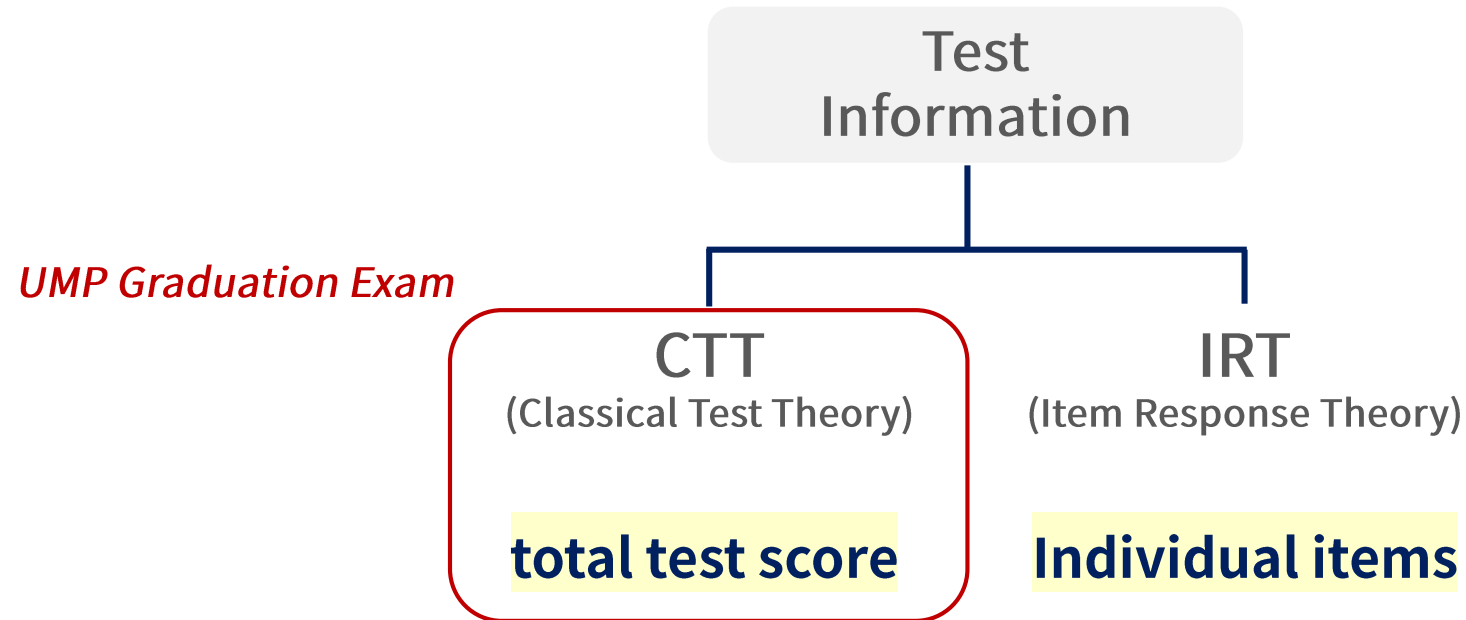
◆ UMP Graduation Exam Results can be reviewed as,

1. Item Difficulty

2. Item Discrimination

3. Test Reliability

## ◆ Difference between CTT & IRT



## 1. Item Difficulty

- An index indicating the level of ease or difficulty of an item
- The proportion or probability of students who answered correctly among all students


	item 1					item 6
Student	1	0	0	0	0	0
	1	1	0	1	0	0
	1	1	1	0	0	0
	1	1	1	1	0	0
	1	1	1	1	1	0
	1	1	1	1	1	1
		Easy				

## 1. Item Difficulty

- Test 1

Average Item Difficulty: **0.61**

Item Difficulty		# of items	Percentage(%)
$0 \leq diff < 0.25$	A very difficult item	9	6.00
$0.25 \leq diff < 0.50$	A difficult item	39	26.00
$0.50 \leq diff < 0.75$	A moderately difficult item	55	36.67
$0.75 \leq diff < 0.95$	An easy Item	40	26.66
$0.95 \leq diff < 1.00$	A very easy Item	7	4.67
<b>Total</b>		<b>150</b>	<b>100.00</b>

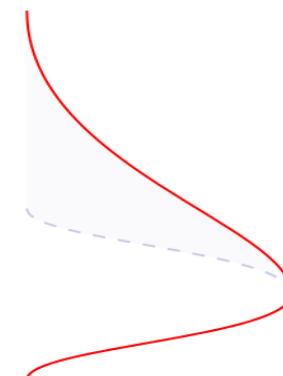


## 1. Item Difficulty

## • Test 2

Average Item Difficulty: **0.68**

Item Difficulty		# of items	Percentage(%)
$0 \leq diff < 0.25$	A very difficult item	7	4.67
$0.25 \leq diff < 0.50$	A difficult item	28	18.66
$0.50 \leq diff < 0.75$	A moderately difficult item	43	28.67
$0.75 \leq diff < 0.95$	An easy Item	60	40.00
$0.95 \leq diff < 1.00$	A very easy Item	12	8.00
<b>Total</b>		<b>150</b>	<b>100.00</b>



## 1. Item Difficulty

- Item with very **low** ↓ *or* **high** ↑ difficulty

⇒ lead low item discrimination



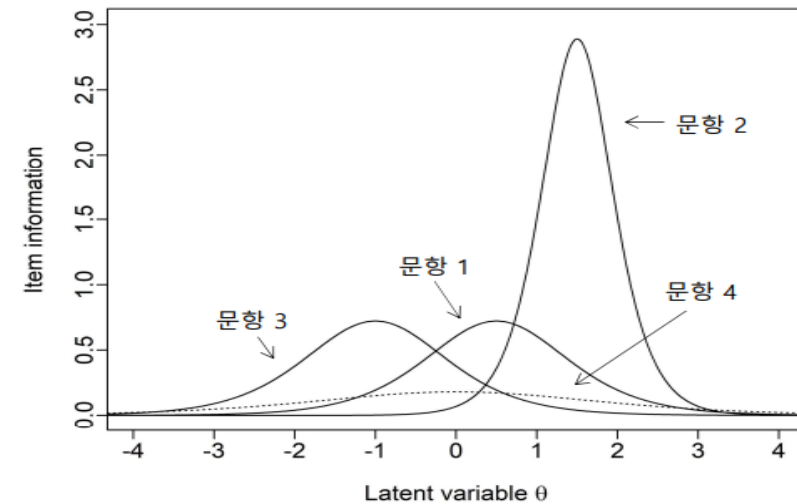
**Item difficulty and discrimination are correlated with each other**

## 2. Item Discrimination

- An index of how well an item discriminates between students
- Index: The correlation coefficient between item scores and total scores



Usually treated as Item Information





## 2. Item Discrimination

- Test 1

Average Item Discrimination: **0.23**

Item Discrimination		# of items	Percentage(%)
$disc < 0$	A non-discriminative item	12	8.00
$0 \leq disc < 0.10$	A very low discriminative item	21	14.00
$0.10 \leq disc < 0.20$	A low discriminative item	40	26.67
$0.20 \leq disc < 0.40$	A discriminative item	51	34.00
$0.40 \leq disc$	A very discriminative item	26	17.33
Total		150	100.00

## 2. Item Discrimination

- Test 2

Average Item Discrimination: **0.20**

Item Discrimination		# of items	Percentage(%)
$disc < 0$	A non-discriminative item	10	6.67
$0 \leq disc < 0.10$	A very low discriminative item	23	15.33
$0.10 \leq disc < 0.20$	A low discriminative item	44	29.33
$0.20 \leq disc < 0.40$	A discriminative item	59	39.33
$0.40 \leq disc$	A very discriminative item	14	9.34
Total		150	100.00

## 2. Item Discrimination

- ✓  $disc < 0$  discrimination item means in CTT ?
  - Total Score  $\uparrow$  with **Incorrect Answer** or Total Score  $\downarrow$  with **Correct Answer**
  - Categorized into issues such as
    - ① **multiple correct answers**
    - ② **no correct answer**
    - ③ **system errors**      *and so on..*

## 2. Item Discrimination

- Test 1

Item	Difficulty	Discrimination	A	B	C	D
106	0.11	-0.04	0.19	0.63	0.11	0.07
111	0.97	-0.03	0.03	0.00	0.97	0.01
45	0.99	0.00	0.00	0.99	0.00	0.01

Wrong answer > Correct answer

Too easy item

- Test 2

Item	Difficulty	Discrimination	A	B	C	D
86	0.09	-0.11	0.31	0.02	0.59	0.09
146	0.20	-0.05	0.03	0.20	0.47	0.30
99	0.32	-0.19	0.32	0.06	0.13	0.48

Too difficulty item

Wrong answer > Correct answer



Difficulty and Discrimination are related to each other and response option rate

## 2. Item Discrimination

- **Low discriminative** power causes **low test reliability**

⇒ The test fails to differentiate between high-ability and low-ability students

⇒ The test results may not be consistent



The reliability of the test decreases

## 3. Reliability

- Reliability increases as the number of items increases

Reliability	Test1	Test2
Cronbach's $\alpha$	0.864 (with 150 items)	0.852 (with 150 items)

- ⇒ It is important to achieve optimal reliability with an appropriate number of items
- ⇒ The 'Reliability after item deletion' index can be used to check a test's quality

2023 Test

70

Student A

2024 Test

75

Student B

Does 'Student B' have a higher ability?

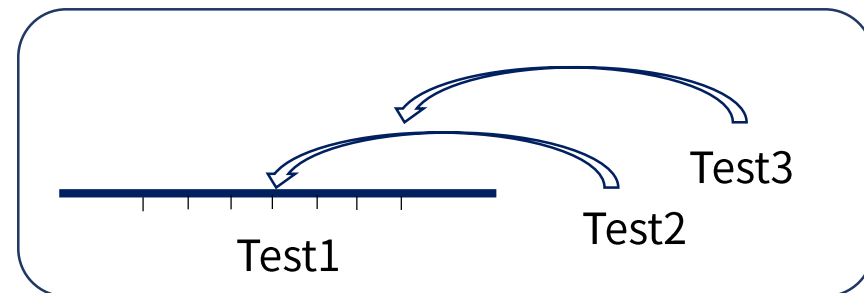


Without equating, it cannot be determined

## 03 Licensing Exam

### ◆ Test Equating

- The difference in test scores
  - ① **student's ability?** or ② the effect of test difficulty?
- Statistically adjust for differences in test difficulty between test forms



After equating, **differences in test scores** can be interpreted as **changes in students' ability**

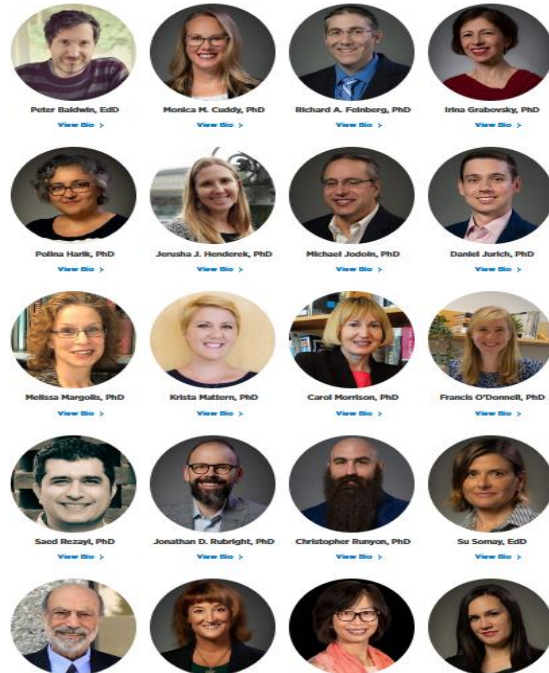


## ◆ Test Equating in USMLE

^

4. How are scores from different forms of a Step examination made comparable?

Within each Step examination, USMLE creates various forms that are similar in difficulty and content. Scores on individual examination forms are made comparable through equating, a psychometric process that adjusts scores based on the difficulty of the questions. This can be thought of as small score increases applied to examinees who see somewhat more difficult sets of test questions, and small score decreases applied to test takers who see somewhat less difficult sets of test questions. This process ensures examinees who take different forms are held to the same passing standard.



More than **20 Psychometricians** at NBME

<https://www.nbme.org/research/meet-our-experts>

◆ The key to test development is..

Ensures fair outcomes for all students

**Accuracy**

**Fairness**

Accurately measures the student's ability



*We look forward to progressing to the test equating stage in the future*

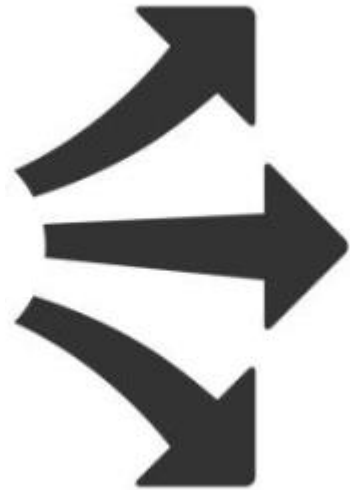
## 04 Conclusion

- ◆ To make Accurate and Fair UMP Graduation Exam,
  - ✓ Check test Blue-print, item writer, item reviewer, item analysis results, training Process...



## 04 Conclusion

- ◆ To make Accurate and Fair Licensing Exam,



**Medical Licensing Exam in Vietnam**

**With UMP's extensive expertise, experience and in-depth know-how**

◆ Through an Accurate and Licensing Exam,



**Verification of competency as a physician**



# Thank you for Listening

