Item Analysis

Cherdsak Iramaneerat
Department of Surgery
Faculty of Medicine Siriraj Hospital
Mahidol University

Item Analysis

- A group of statistical analyses having two characteristics:
  - The data consist of actual responses of test takers to individual test items
  - The primary purpose is to gain information about the items (rather than about test takers)

Livingston SA. Item analysis. In: Downing SM, Haladyna TM. Handbook of test development. Mahwah, NJ: LEA, 2006, p. 421-444.

Outline

- MCQ item analysis
  - Item statistics
  - Test statistics
- MEQ and OSCE item analysis
- Applications
- Limitations

MCQ Item Analysis

- Item statistics
  - Item difficulty
  - Item discrimination
  - Distractor functionality
- Test statistics
  - Internal consistency reliability
  - Standard deviation and mean
  - Average difficulty
  - Average discrimination

Item Statistics

Looking at individual test items

Item Difficulty

- Proportion of examinees answering an item correctly (p)

$$p = \frac{C}{C + I}$$

C = number of examinees with a correct answer

I = number of examinees with incorrect answers

- Ideal: 0.45 – 0.75
- Good: 0.76 – 0.91
- Acceptable: 0.25 – 0.44
- Problematic: < 0.24 or > 0.91

Item Discrimination

- The ability of an item to discriminate high scorers from low scorers
- Point-biserial correlation (r)

$$r = \frac{M_p - M_q}{SD}\sqrt{pq}$$

Mp	=	Mean score of examinees with a correct answer
Mq	=	Mean score of examinees with incorrect answers
SD	=	Standard deviation of test scores
p	=	Proportion of examinees with a correct answer
q	=	Proportion of examinees with incorrect answers

Point-Biserial Correlation

- The correlation between an item score with the total score
  - **Range: −1.0 – 1.0**
  - **Point–biserial of an item should be positive**
    - Ideal: 0.20 or higher
    - Acceptable: 0.1 – 0.19
    - Problematic: < 0

Distractor Functionality

A functioning distractor is an incorrect option that:

1. Is chosen by at least 5 percent of examinees
2. Has a negative point-biserial correlation with the total score

Siriraj Hospital's IA report

No. : 1		p Value : 0.64				r_{pbi} : 0.23			
A		B		C		* D		E	
r_{pbi}	%	r_{pbi}	%	r_{pbi}	%	r_{pbi}	%	r_{pbi}	%
0.02	6.98	-0.18	5.08	-0.17	8.57	0.23	63.81	-0.07	15.56

Exercise

- ให้อาจารย์พิจารณาผลการวิเคราะห์ข้อสอบ 72 ข้อที่ได้รับ
- ข้อสอบที่มีปัญหามากที่สุด 3 อันดับแรกคือ...

Test Statistics

Looking at the whole test

Reliability

- Consistency of test scores
  - If we test the students again, will they get the same scores?
  - Range: 0 – 1
  - High values: highly consistent test scores

KR-20

$$KR20 = \left(\frac{n}{n-1}\right)\left(1 - \frac{\sum pq}{Var}\right)$$

n = number of items

Var = Variance of the whole test

p = Proportion of people passing the item

q = Proportion of people failing the item

Some Confusion

- Cronbach's Alpha
  - Polytomous or dichotomous items
- KR-20
  - Dichotomous items
- KR-21
  - Dichotomous items, when difficulties of all the items on the test are equal

How Much is Enough?

- Depends on test scores uses
  - High-stakes exam: 0.9 or higher
  - Medium-stakes exam: 0.80 – 0.89
  - Low-stakes exam: 0.70 – 0.79

Improving Reliability

- Increase the number of test items
- Adjust item difficulty to obtain larger spread of test scores
- Adjust testing conditions to eliminate interruptions, noise, and other disrupting factors
- Eliminate subjectivity in scoring

Mean and Standard Deviation

- Effective instruction => All students can do the test well.
  - High mean scores
  - Low standard deviation
- High standard deviation: Wide range of students' scores
  - Some students can solve the problems in the tests, while some students cannot do.
- Too difficult test => Most students fail to get correct answers.
  - Low mean scores
  - Low standard deviation

Average Difficulty

- Average of p values of all items on the test
- Small group of students:
  - Difficult to interpret
  - Depends on the ability distribution of students
- Large group of students:
  - Assume a fair sampling of students
  - Indicates the average difficulty of the whole test

Average Discrimination

- Average point-biserial correlation of the whole test
- Indicates how good the items on the test can differentiate high scorers from low scorers.
- High values generally indicate a good test.
- Effective instruction: All students can do well on the test.
  - —A low value does not necessarily indicate bad items.

MEQ and OSCE Analysis

- MEQ and OSCE
  - Item difficulty
  - Item discrimination
  - Reliability

Item Difficulty

- Score of an item is on an ordinal scale instead of a dichotomous scale
  - Item difficulty
    - P-value => Percentage score
  - Item discrimination
    - Point biserial correlation => Pearson correlation

Item Difficulty

- Score percentage
  - An easy item: High value
  - A difficult item: Low value
  - Which part is difficult?
    - Subscale analysis of each part

Item Discrimination

- Pearson correlation

Excel = Pearson (Range 1, Range 2)

=PEARSON(BM2:BM275,H2:H275)

Pearson Correlation

หากท่านเห็นผลการวิเคราะห์เช่นนี้ในข้อสอบ MEQ ท่านจะทำอย่างไร

	Item 1	Item 2	Item 3	Item 4	Item 5
Pearson	0.68	0.71	0.40	0.63	0.63

Internal Consistency Reliability

- Consistency of test scores: If we test the students again, will they get the same scores?
- In written exam, one commonly reported index of reliability is Cronbach's Alpha

$$\alpha = \frac{n}{n-1}\left(1 - \frac{\sum \sigma_{x_i}^2}{\sigma_x^2}\right)$$

n = number of testlets

σ_x^2 = score variance of total scores

$\sigma_{x_i}^2$ = score variance of the i^{th} testlet

Applications

1. Posttest score adjustment
2. Item revision
3. Item pool management
4. Improvement of instruction

Limitations

1. Sample dependency
2. Reliability is the property of test scores, not test items.
3. Numbers are there to serve us, not the other way around.

Summary

- MCQ item analysis
  - Item statistics
  - Test statistics
- MEQ and OSCE item analysis
- Applications
- Limitations

Questions and Comments

Cherdsakiramaneerat@gmail.com

"We all need people who will give us feedback. That's how we improve."

Bill Gates