

Lecture 6 Psychological Testing and Measurement
Sunthud Pornprasertmanit

Item Analysis

Class Assignment

Case 8.1 Emotional Intelligence

Some evidence for emotional intelligence
construct validity

Class Assignment

- Why consider the test before using?
 - The result of the score
 - The result of research
 - How much believable it is

Class Assignment

Which are considered?

- What does it mean?
 - Definition
 - Domain
- Objective of this test
- Reliability of the score for this objective
- Validity of the score for this objective

Class Assignment

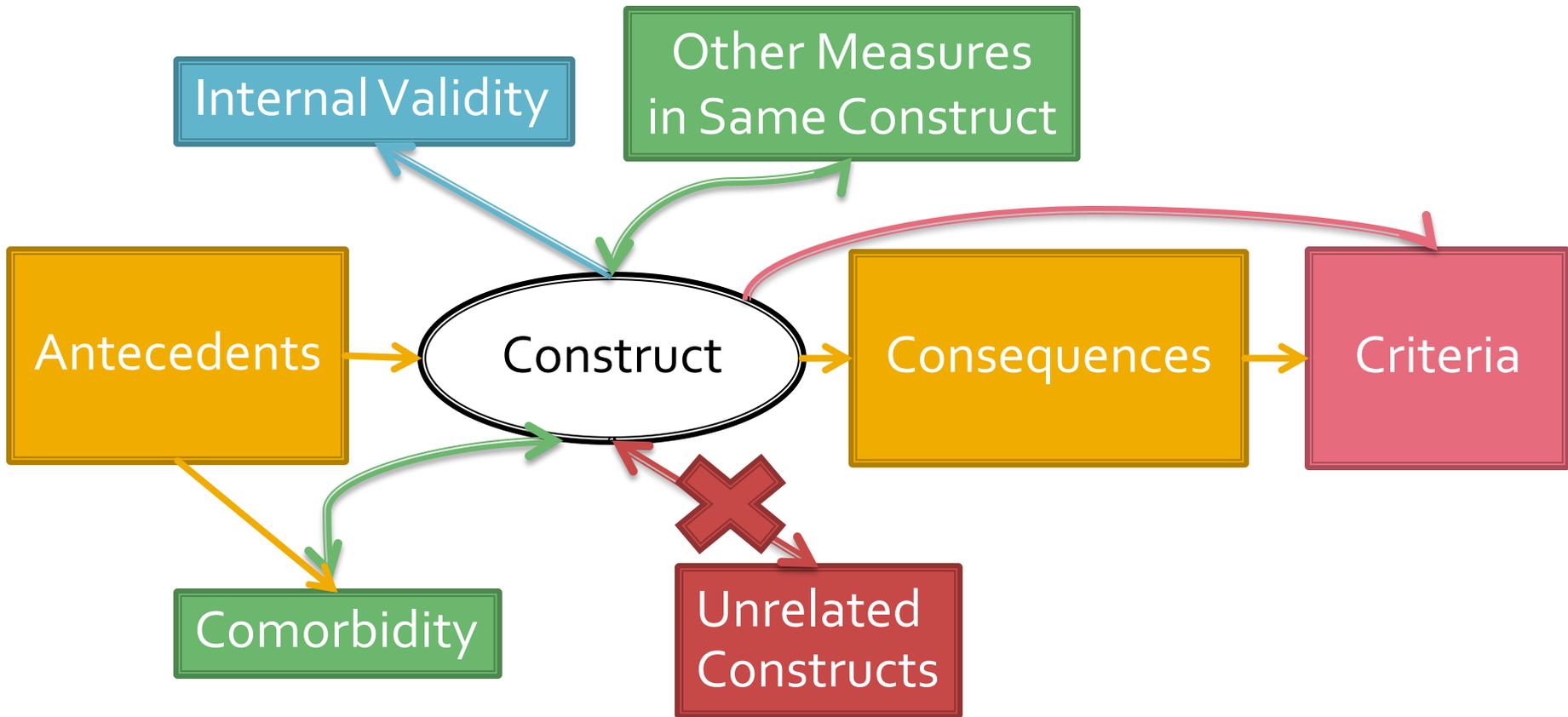
Enhance the construct validity evidences

- Theory; literature review of emotional intelligence

Mayer, J. D., Roberts, R. D., & Barsade, S. G. (2008). Human abilities: Emotional Intelligence. *Annual Review of Psychology, 59*, 507-536. [Downloadable from Blackboard]

Class Assignment

Enhance the construct validity evidences



Class Assignment

Homework

- Finding the problem of EQ definition
- Find the ways to validate the EQ measures as much as possible from the article
- Which concepts are new for you in this article

Case Feedback

3.2 Maid performance assessment sheet

- Checking the improvement
- Evaluate the performance assessment sheet
- How to prove that the score from this test can be used in the real situation?

Exercise

Exercise 8.2

Item Analysis

Lecture 6 Psychological Testing and Measurement
Sunthud Pornprasertmanit

Item Analysis

- Reliability and validity of the test depends on characteristics of items.
- Item analysis makes it possible to shorten a test and at the same time to increase its validity and reliability, in contrast with Spearman-Brown formula.

Item Difficulty

- The difficulty of an item is defined in terms of the percentage (or proportion) of persons who answer it correctly.
- Arrange items in order of difficulty.
 - Improve test takers confidence
 - Reduces the likelihood of their wasting much time on items beyond their ability.

Item Difficulty

- Reason for measuring item difficulty is to choose items of suitable difficulty level.
- Very easy or very difficult items do not provide information about individual difference.
- These items contribute nothing to reliability and validity of the test.

Item Difficulty

- However, very easy items improve test takers' morale.
- For maximum differentiation, items at the .50 test difficulty level should be selected.
- As a result, it improve item intercorrelation.

Item Difficulty

- Possibility of guessing in multiple-choice items
- Lord (1952) suggested that, for a five-option multiple-choice item, the average proportion correct should be approximately .69.

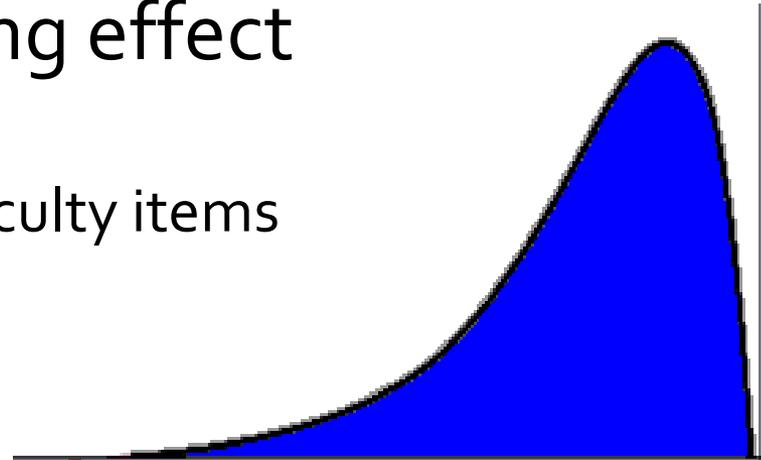
Item Difficulty

- The difficulty of the test as a whole is, directly dependent on the difficulty of the items that make up the test.

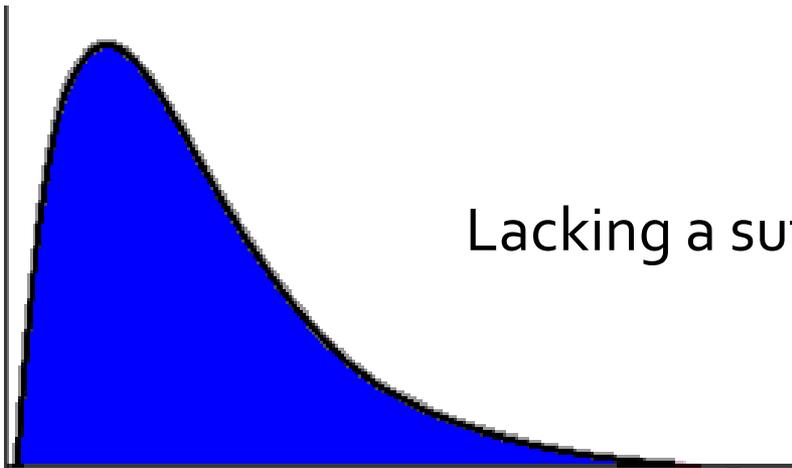
Item Difficulty

- Floor effect and ceiling effect

Lacking a sufficient of difficulty items



Lacking a sufficient of easy items



Item Difficulty

- When the standardization sample yields a markedly nonnormal distribution on a test,
- The difficulty level of the test is ordinarily modified until a normal curve is approximated.
- Only in this way can the maximum differentiation between individuals at all ability levels be obtained with the test.

Item Difficulty

- Types of discrimination objectives determine
 - The choice of *appropriate* item difficulties,
 - The optimal form of the *distribution* of test scores.
- Screening Purposes: The difficulty values should be close to the *desired selection ratio*.

Item Difficulty

- The choice of item difficulty in *mastery testing* should probably be at .80 or .90 level, in post training.
- For pre-training group, item difficulties should be low percentage.

Item Discrimination

- Item discrimination refers to the degree to which an item *differentiates* correctly among test takers in the behavior that the test is designed to measure.
- Which items should discriminate?
 - External Criteria (such as performance, mastery)
 - Internal Criteria (Total score)

Item Discrimination

- Statistical analysis that related to Item discrimination
 - Correlation (Pearson, Point-biserial, Biserial, Phi)
 - Corrected Item Total Correlation (CITC)
 - Regression
 - Mean difference (t -test)
 - Categorical Criterion
 - Extreme groups

Item Discrimination

- Statistical analysis that related to Item discrimination
 - Proportion difference (Chi-square, Fisher exact test)
 - Alpha if item deleted

Item Discrimination

- When using extreme groups,
 - The more extreme the groups, the sharper will be the differentiation.
 - The use of very extreme groups would reduce the reliability of the results because of the small number of cases utilized.
 - Kelly (1939) found that the optimum point is upper and lower 27%, in a normal distribution.
 - ... found that this optimum point the rejected items are similar to those analyzed by CITC.

Item Discrimination

- However, caution of using extreme groups approach

If a researcher does not have reason to use EGA [extreme group approach] beyond the fact that it increase **the odds of achieving statistical significance**, we strongly **caution** against the use of EGA. (Preacher, Rucker, MacCallum, & Nicewander, 2006, p. 190)

Item Discrimination

- Under certain conditions, the external and internal discrimination may lead to opposite results.

Item Discrimination

Internal discrimination

- Make more internal consistency

Item Discrimination

External Discrimination

- When using regression techniques to relate external criterion, the most satisfactory items are
 - those with the highest external validities
 - the lowest coefficients of internal consistency.
- Its content will be heterogeneous that preclude meaningful interpretation of test score.

Item Discrimination

External Discrimination

- The item selection from external criterion is affected by sampling error.
- The result of item selection by external criterion should be cross-validated.

Item Discrimination

- For many testing purposes, a satisfactory compromise is to sort the relatively homogeneous items into *separate tests or subtests*,
- Each of which covers a different aspect of the external criterion.

Item Discrimination

- The item discrimination indices are *not independent of item difficulty* but are biased in favor of intermediate difficulty levels.
- In mastery testing, the useful items are that can discriminate *between mastery and nonmastery* people.

Analysis of Distracters

- The analysis of distracters in multiple-choice may reveal some knowledge from test takers.
- The distracter that cannot deceive any test takers to choose may be changed.
- However, when changing distracters, the item difficulty and discrimination will change.

Item Analysis of Speeded Tests

- Item indices computed from a speeded test may be misleading.
- The item indices found from a speed test will reflect the position of the item in the test
- Rather than its intrinsic difficulty or discriminative power.

Cross-Validation

- Cross checking with different sample of people.
- Any validity coefficient computed on the same sample that used for item-selection purposes (by external criterion) will capitalize on *random sampling errors* within that particular sample and will consequently be *spuriously high*.

Cross-Validation

- Classic example from Kurtz (1948): A research test on the Rorschach test for selection effective managers for insurance companies.
 - Item analysis group: correctly classified 70 of the 80 managers
 - Cross validate group: correctly classified 21 of the 41 managers

Cross-Validation

- When the number of original items is large and the proportion retained is small, there is more opportunity to capitalize on chance.
- Another condition affecting amount of shrinkage is size of sample

Exercise

Computer Exercise 5.1, 7.3, 8.4, 12.2

Sales.sav

- Reliability (Alpha)
- Criterion-related validity (Correlation b/w cognitive test and sales performance)
- Item Difficulty (Percent Correction)
- Corrected Item-total Correlation
- Item Discrimination with external criterion (Regress sales performance on items)

Classroom Assignment

Case 12.2

- Examine: Item difficulty, Item discrimination and Distracters Statistic

Next Lecture

Use of Psychological Test

Lecture 5 Psychological Testing and Measurement
Sunthud Pornprasertmanit

Use of Psychological Testing

- Setting the Cutoff
- Setting the Grade
- Taylor and Russell (1939) Table
- Guessing
- Social Desirability
- Test Administration
- Test Coaching
- Test Security

Changing Reading List

Lecture 7

- Shultz & Whitney (2005) Unit 13, 16
- ศิริชัย (2548) บทที่ 8

Changing Reading List

Lecture 8

- Brown, K. W., & Ryan, R. M. (2003). The benefits of being present: **Mindfulness** and its role in psychological well-being. *Journal of Personality and Social Psychology*, *84*, 822-848.
- McCarthy, D. M., Pedersen, S. L., Thompsen, D. M., & Leuty, M. E. (2006). Development of a measure of **drinking and driving expectancies** for youth. *Psychological Assessment*, *18*, 155-164.
- Mumford, T. V., Van Iddekinge, C. H., Morgeson, F. P., & Campion, M. A. (2008). The **team role test**: Development and validation of a team role knowledge situational judgment test. *Journal of Applied Psychology*, *93*, 250-267.

Changing Midterm Test

- Choice
- Fill in the blank
- Improving the test