

Lecture 5 Psychological Testing and Measurement
Sunthud Pornprasertmanit

Validity

What is Validity?

Classic Definition

- The validity of a test is the extent to which a test measures what it purports to measure.

What is Validity?

Assumption of Classical Definition

- Validity is a property of tests, rather than of test score interpretations
- In order to be valid, test scores should measure some purported construct directly
- Score validity is a function of the test author's or developer's understanding of whatever construct she or he intends to measure

What is Validity?

Current Definition

- According to Messick (1989),
- Validity is an integrative evaluative judgment to degree to which empirical evidence and theoretical rationales support
- The *adequacy* and *appropriateness* of *inferences* and *actions* based on test scores or other modes of assessment.

What is Validity?

- According to current definition of validity, the term construct has been used in two alternate ways
 - Traits, Processes, Knowledge Stores, or Characteristics
 - Inferences that may be made on the basis of the test scores (e.g. test that empirical support)

What is Validity?

- For example, Mechanical Aptitude Test

Type of Validity

- Content Validity
- Criterion-Related Validity
- Construct Validity

Content Validity

- Content validity tell us
 - Does the test cover a representative sample of the specified skill and knowledge?
 - Is test performance reasonably free from the influence of irrelevant variables?

Content Validity

- Commonly used in CRT in educational settings
- The domain under consideration should be fully described in advance

Content Validity

- Validity depends on
- the *relevance of the individual's test responses* to the behavior area under consideration,
- rather than on the apparent relevance of *item content*.

Content Validity

- Content Validity is
- Objectives, Domain → Test Items
- Otherwise, in knowledge testing, cognitive processes used

Content Validity

- Table of Specifications
 - *Domain*: Derived by examination of syllabi or textbooks or consults subject-matter experts or job analysis
 - *Processes to be tested*: Knowledge, Comprehension, Application, Analysis, Synthesis, Evaluation
 - *Number of items*

Content Validity

- The current approach to content validity
- Multiple judges of content validity
- Quantifying judgments of experts, such as content validity ratio
- Qualitative feedback can be supplemented
- Assessing in various dimensions such as relevance, representativeness, specificity, and clarity.

Content Validity

- Are expert judgments appropriate to use as content validity indices in all measures?
 - Personality
 - Interest
 - Aptitude Test

Content Validity

- For aptitude and personality tests, content validity is usually inappropriate and may be misleading.
- Representative and Relevance are important factors, but are not enough for validation.
- It would be impossible to determine the psychological functions measured by the test from an inspection of its content.

Content Validity

- Experts or others people opinions in representative and relevance are essential but insufficient
- Not the evidence of content validity

Content Validity

- Another approach is job analysis
- Finding the representative domain of the test

Content Validity

- Face validity = “looks valid”
- Face validity is a desirable feature of tests.
- Face validity should never be regarded as a substitute for content validity.

Criterion-Related Validity

- Criterion-related validity = predicting criteria
- Commonly used in the selection and classification of personnel and clinical settings

Criterion-Related Validity

- Criteria are which we really want to know.
- Criterion measures are indexes of the criterion that tests are designed to assess or predict
 - Such as indexes of academic achievement
 - Performance in specialized training
 - Indexes of job performance
 - Membership in contrasted group.

Criterion-Related Validity

- Mismatch between criterion and criterion measures
 - Criterion Deficiency
 - Criterion Contamination

Criterion-Related Validity

- Make certain that the test scores do not themselves influence any individual's criterion status
 - Case study in assessment center

Criterion-Related Validity

- The reliability and validity of criterion measures need to be evaluated, just as the reliability and validity of test scores do.
- Some criteria are more complex than others.

Criterion-Related Validity

- Some criteria can be gauged at the time of testing; others evolve over time.
 - Some criteria needed to be mature before analyzing criterion-related validity
 - Intermediate and ultimate criteria
- Currently available tests may be regarded as a criterion measure, if the new test is an abbreviated or simplified form of a currently available test.

Criterion-Related Validity

- Criterion related validity can be differentiated between
 - Concurrent Validity: “Does Smith qualify as a satisfactory pilot?”
 - Predictive Validity: “Does Smith have the prerequisites to become a satisfactory pilot?”

Criterion-Related Validity

- In a number of instances, concurrent validation is employed merely as a substitute for predictive validation
- Sometimes, concurrent prediction is the most appropriate type and can be justified in its own right.

Criterion-Related Validity

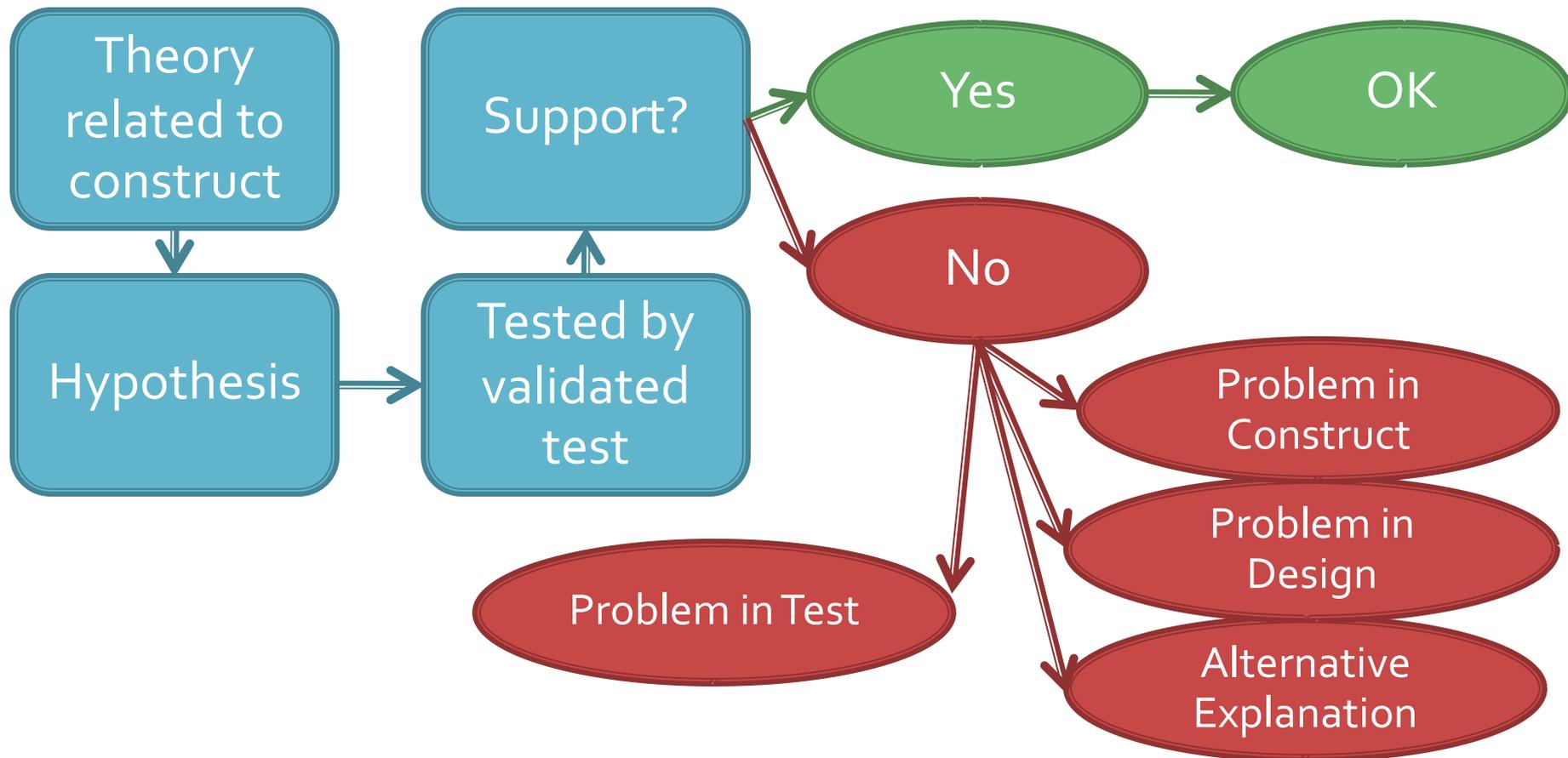
- Criterion-related validity is often used in local validation studies.
- Relationships between test scores and criterion measures may or may not generalize across groups, settings, or time periods.
- Validity coefficients were found to vary widely, a serious limitation in the usefulness of standardized tests in personnel selection.

Construct Validity

- Construct validity is the extent to which the test may be said to measure a theoretical construct or trait
- Construct validation requires the gradual accumulation of information from a variety of sources.

Construct Validity

■ Process of Construct Validity



Procedures for Construct Validation

Developmental Changes (Age differentiation)

- A psychological test validated against such a criterion measures behavior characteristics that increase with age

Procedures for Construct Validation

Developmental Changes (Age differentiation)

- Inapplicable to any functions that do not exhibit clear-cut and consistent age changes.
- Age differentiation is a necessary but not a sufficient condition for validity.

Procedures for Construct Validation

Developmental Changes (Age differentiation)

- Developmental analyses are basic to the construct validation of ordinal scales.
 - Such as moral development scale
- The construct validation of ordinal scales should include empirical data on the sequential invariance of the successive steps.

Procedures for Construct Validation

Correlations with Other Tests

- Correlation between a new test and similar earlier tests
- If the new test correlates too highly with an already available test,
- Without such added advantages as brevity or ease of administration,
- Then the new test represents needless duplication.

Procedures for Construct Validation

Correlations with Other Tests

- Another way to demonstrate that the new test is relatively free from the influence of certain irrelevant factors.
- However, low correlations would not themselves ensure validity.

Procedures for Construct Validation

Convergent and Discriminant Validity

- Convergent Validity – High correlations between measures designed to assess a given construct
- Discriminant Validity – Low correlations between measures that are supposed to differ.

Procedures for Construct Validation

Convergent and Discriminant Validity

- Campbell and Fiske (1959) developed *multitrait-multimethod matrix* to assess convergent and discriminant validity
- This procedure requires the assessment of two or more traits by two or more methods.

Procedures for Construct Validation

| Method | Trait | Self-report | | | Observation | | | Projective | | |
|-------------|-------|-------------|-------|-------|-------------|-------|-------|------------|-------|-------|
| | | ANX | AFF | DOM | ANX | AFF | DOM | ANX | AFF | DOM |
| Self-report | ANX | (.90) | | | | | | | | |
| | AFF | .45 | (.88) | | | | | | | |
| | DOM | .35 | .38 | (.80) | | | | | | |
| Observation | ANX | .60 | .23 | .10 | | | | | | |
| | AFF | .25 | .58 | -.08 | .47 | (.93) | | | | |
| | DOM | .12 | -.12 | .55 | .30 | .32 | (.86) | | | |
| Projective | ANX | .56 | .22 | .11 | .65 | .40 | .31 | (.94) | | |
| | AFF | .23 | .57 | .05 | .38 | .70 | .29 | .44 | (.89) | |
| | DOM | .13 | .10 | .53 | .19 | .26 | .68 | .40 | .44 | (.86) |

Reliability Coefficient

Method variance

Heterotrait-heteromethod correlation

ANX = Anxiety; AFF = Affiliation; DOM = Dominance

Procedures for Construct Validation

Convergent and Discriminant Validity

- If there is common method variance, the test unduly affected by some irrelevant common factor.
- Such as ability to understand the questions, social desirability.

Procedures for Construct Validation

Nomological Validity

- Correlation with other constructs in theory
- Correlation with other tests
- Experimental Manipulation
- Method of Contrasted Group (Known Group)

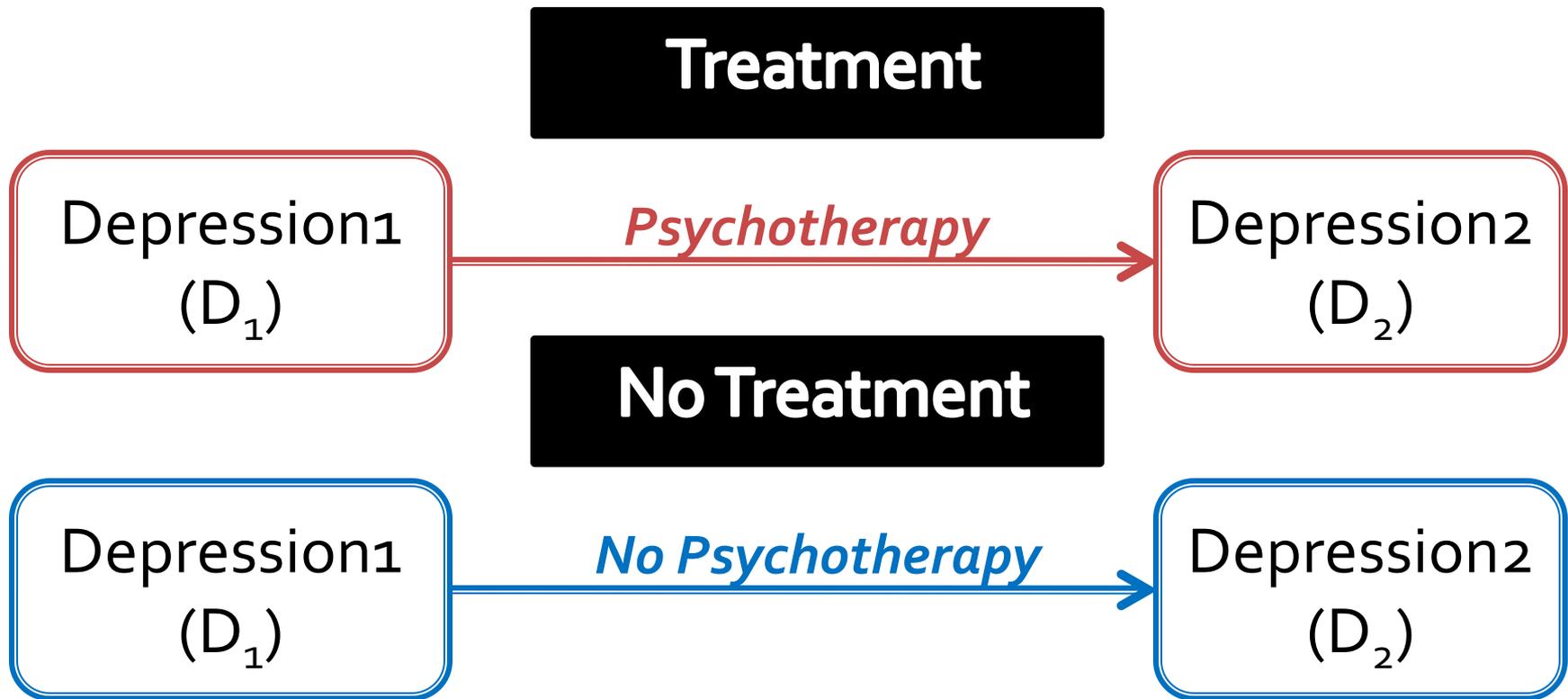
Procedures for Construct Validation

Experimental Intervention

- Experiments on the effect of selected variables on the test scores.
- For example, pretest and posttest.

Procedures for Construct Validation

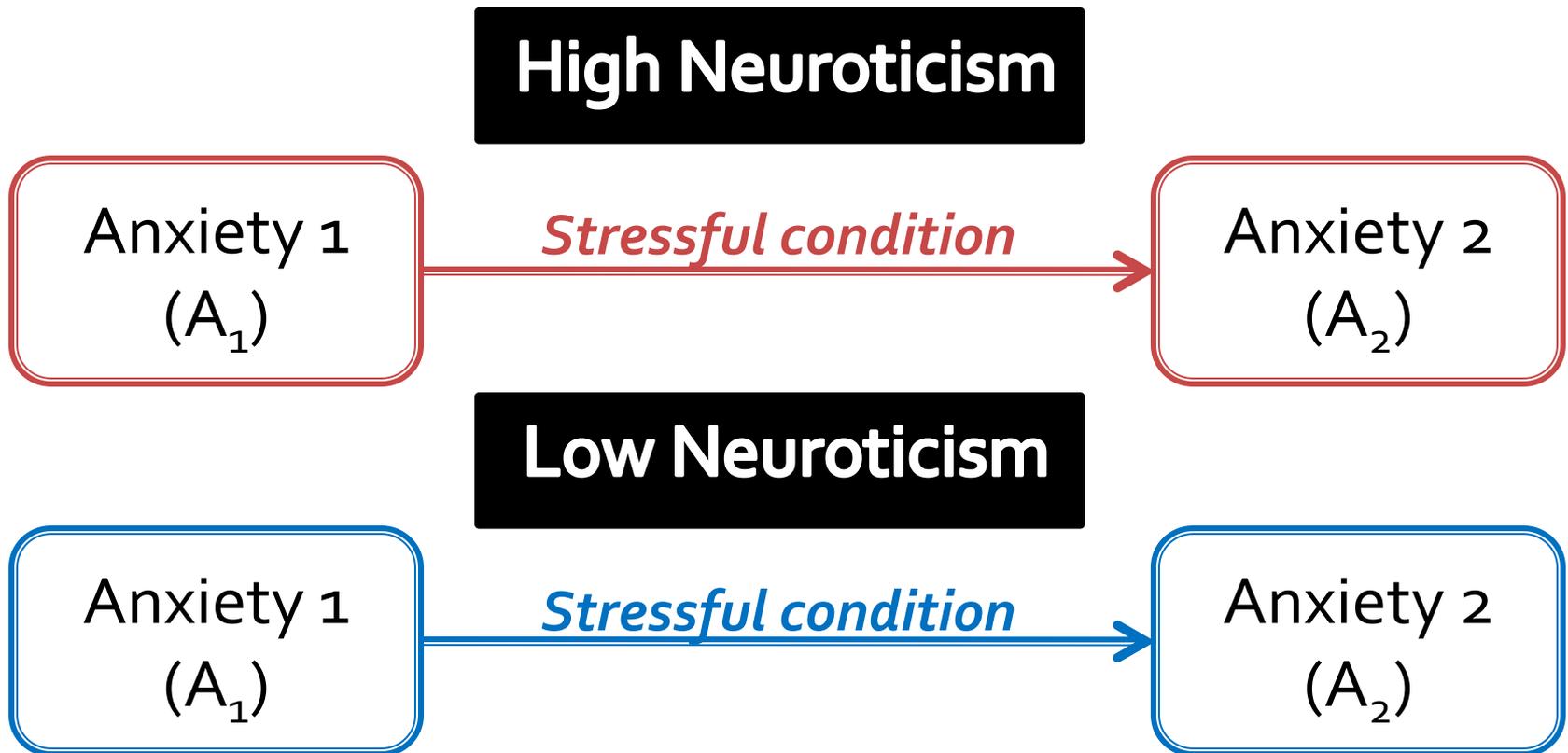
For example, Depression Scale



Validation: $(D_2 - D_1 \text{ in T}) \neq (D_2 - D_1 \text{ in No T})$

Procedures for Construct Validation

For example, Neuroticism (N)



Validation: $(A_2 - A_1 \text{ in High N}) > (A_2 - A_1 \text{ in Low N})$

Procedures for Construct Validation

Factor Analysis

- When tests are analyzed by correlation, the results may reveal certain cluster among tests,
- Suggesting the location of common traits.
- The statistical techniques that locate the common factors, is factor analysis.
- The number of tests is reduced to a relatively small number of factors, or common traits.

Procedures for Construct Validation

Factor Analysis

- How major factors determining its score →
Factor Loading
- This correlation is called factor validity of the test.

Procedures for Construct Validation

Factor Analysis

- For example, verbal comprehension factor correlates .66 with a vocabulary test .

Procedures for Construct Validation

Structural Equation Modeling

- Relations among constructs and of path whereby a construct affects criterion performance.



Procedures for Construct Validation

Structural Equation Modeling

- This technique helps to understand why tests have high or low validity in a given situation.
- Structural Equation Modeling, indeed, analyzes several multiple regression technique simultaneously.

Procedures for Construct Validation

Structural Equation Modeling

- To analyze SEM, the researcher must design a model of the hypothesized casual relations to be tested.
- Then, the correlation matrix of variables is computed.
- The correlation matrix computed by model is compared with empirical correlation matrix by goodness-of-fit test.

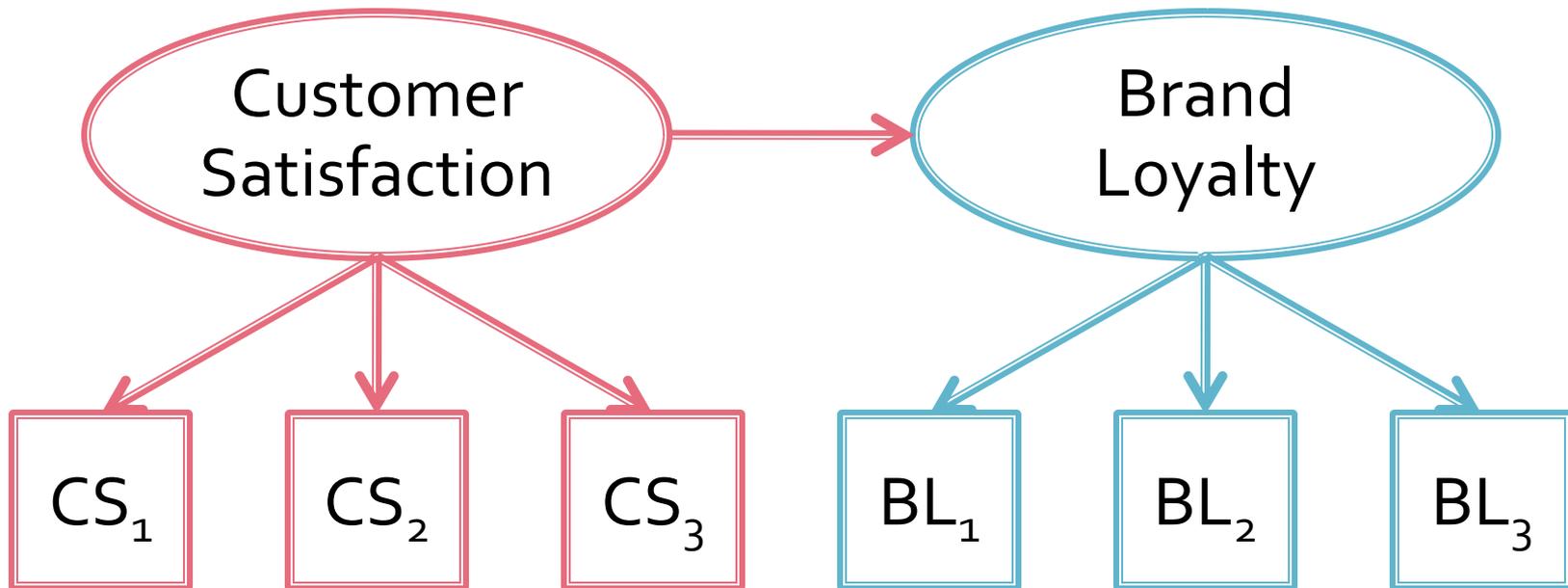
Procedures for Construct Validation

Structural Equation Modeling

- If goodness-of-fit test is not statistically significant, then the model is possible to explain relation between variables.
- Another feature of SEM is that causal relations are typically computed between factor scores.
- The use of constructs provide more stable and reliable estimates.

Procedures for Construct Validation

Structural Equation Modeling



Factors Affect Validity Measured by Correlation

- Unreliability (Attenuation)
- Restriction of Range
- Differential Validity

Case Feedback

3.2 Maid performance assessment sheet

Class Assignment

Case 7.1, 8.1

Case 7.1

Who was assessed in these work?

What is criterion?

What is criterion measure?

Are the criterion and criterion measure congruent?

What is predictor?

How criterion-related validity is calculated?

What is the difference between applicants and incumbents?

What affect criterion-related validity coefficient?

Exercise

Computer Exercise 5.1, 7.3, 8.4

Exercise 8.2

Next Lecture

Item Analysis

Lecture 5 Psychological Testing and Measurement
Sunthud Pornprasertmanit

Item Analysis

- Item analysis is to find problematic items.
 - Item Difficulty
 - Item Discrimination
 - Analysis of Distractors