




VALIDITY

Lecture 5 Psychological Testing and Measurement
Sunthud Pornpresertmanit



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?

- *Construct* is anything that is devised by the human mind but not directly observable.
- Construct = Latent trait



What is Validity?

- By classical definition, any evidence labels as test validity came to be seen as proof the test was valid and worthy of use
- Regardless of the nature of the link between test score data the interferences that were to be drawn from them.

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)



Type of Validity

- Content Validity
- Criterion-Related Validity
- Construct Validity



Content Validity

- Content validity determine whether test content covers a representative sample of the behavior domain to be measured.
- 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.
- Test users should be guarded against any tendency to overgeneralize regarding the domain sampled by the test.
- There is possible inclusion of irrelevant factors in the test scores.



Content Validity

- Content validity is built into a test from the outset through the choice of appropriate items.



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

- Use multiple judges of content validity and quantify judgments using formalized scaling procedures
 - Every element of an assessment instrument should be judged by multiple experts on applicable dimensions such as relevance, representativeness, specificity, and clarity.
 - Quantitative indices of content validity can be supplemented with qualitative feedback from evaluators.

Content Validity

- Empirical procedures that supplement to content validation
 - Total scores and individual items can be checked for grade progress.
 - Analysis of types of errors commonly made on a test and observation of work methods employed by test takers (by thinking aloud)
 - Scores on the test may be correlated with scores on a reading comprehension test



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

- 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 and inspection of its content.



Content Validity

- Face validity pertains to whether the test “looks valid” to the examinees who take it, the administrative personnel who decide on its use, and other technically untrained observers.
- Face validity is a desirable feature of tests.
- Face validity should never be regarded as a substitute for content validity.

Discussing Question?

- Is experts' judgment of content validity useful?





Criterion-Related Validity

- Criterion-related validity indicate the effectiveness of a test in predicting an individual's performance in specified activities.
- 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 and that are gathered independently of the test in question.
- Such as indexes of academic achievement or performance in specialized training, indexes of job performance, membership in contrasted group.



Criterion-Related Validity

- Careful attention must be given to the selection of criteria and criterion measures.
- Criterion deficiency and criterion contamination
 - An essential precaution in finding the validity of a test is to make certain that the test scores do not themselves influence any individual's criterion status
 - No person who participates in the assignment of criterion ratings has any knowledge of the examinees' test scores.



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.

Discussing Question?

- Are these criterion measures good indices for academic achievement?
 - GPA
 - Percentiles
 - Educational Level
- Is job performance a good criterion for good employees?
- Are current criterion measures of job performance good indexes of job performance?
- Are contrast groups good criterion measures?

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.

Criterion-Related Validity

- Hunter and Schmidt found that the variation of validity coefficient resulting from small sample size, criterion unreliability and restriction of range.
- The meta-analysis procedures can be used for investigating validity generalization.
- Meta-analysis is statistical technique used for integrating findings from different studies.



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

- Lord and Novick (1968) note that construct must defined in two levels
 1. Operational Definition, usually by specifying the procedures used to measure construct
 2. Syntactic definition (Theoretical construct), by the postulation of specific relationship between measures of construct with (1) measures of other constructs in theoretical system and (2) measures of specific real-world criteria



Construct Validity

- The general steps to gain construct validity
 - Formulate hypotheses about how those who differ on the construct are expected to differ on demographic characteristics, performance criteria, or measures of other constructs whose relationship to performance criteria has already been validated, based on theory that underlies construct
 - Select (or develop) a measurement instrument, which consists of items representing behaviors that are specific, concrete manifestations of the construct



Construct Validity

- The general steps to gain construct validity
 - Gather empirical data which will permit the hypothesized relationships to be test
 - Determine if the data are consistent with the hypotheses and consider the extent to which the observed findings could be explained by rival theories or alternative explanations (and eliminate these if possible)

Procedures for Construct Validation

Developmental Changes (Age differentiation)

- The criterion of age differentiation is 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)

- A psychological test validated against such a criterion measures behavior characteristics that increase with age under conditions existing in the type of environment in which the test was standardized.



Procedures for Construct Validation

Developmental Changes (Age differentiation)

- Developmental analyses are basic to the construct validation of ordinal scales.
- 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 is evidence that the new test measures approximately the same general area of behavior as other tests designated by the same name.
- 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

- Correlations with other tests are employed in still 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.
- Campbell and Fiske (1959) developed multitrait-multimethod matrix to assess convergent and discriminant validity



Procedures for Construct Validation

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	(.90)					
	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, it might indicate that a person's scores on this inventory are unduly affected by some irrelevant common factor.
- Such as ability to understand the questions, social desirability.



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

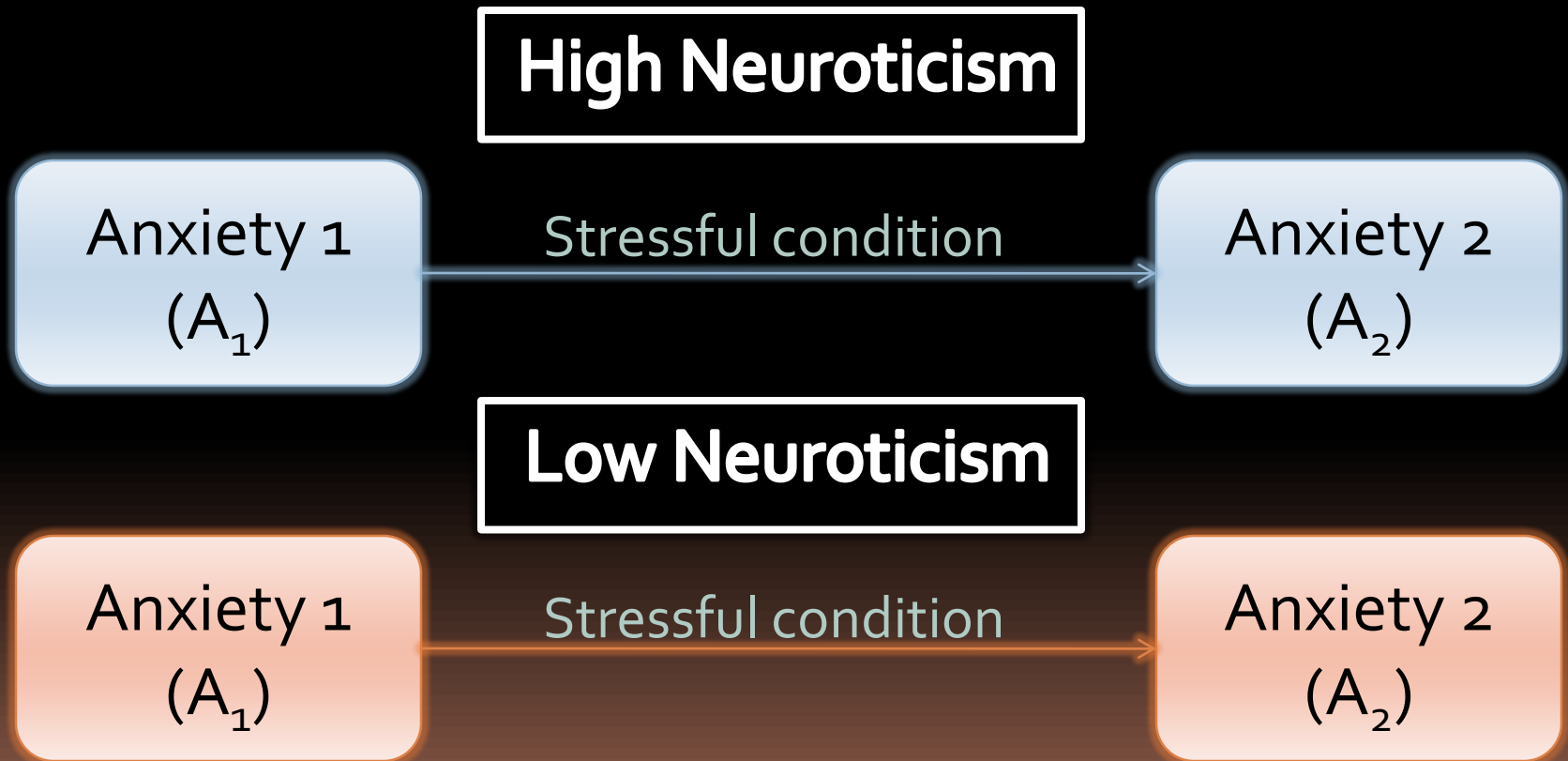
For example, Depression Scale



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

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

- After the factors have been identified, they can be utilized in describing the factorial composition of a test.
- Each test can be characterized in terms of the major factors determining its score, together with the weight or loading of each factor and the correlation of the test with each factor.
- 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

- An important advance in test validation was the consideration of 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 constructs, rather than between isolated measured variables.
- The use of constructs provide more stable and reliable estimates.



Procedures for Construct Validation

Contributions to Cognitive Psychology

- Embretson (1983) proposed two principle aspects of test validation
 - Construct Representation
 - Nomothetic Span



Procedures for Construct Validation

Contributions to Cognitive Psychology

- The object of construct representation is to identify specific information-processing components (procedural skills) and knowledge stores (declarative knowledge) needed to perform the tasks set by the test items.
- Task decomposition can be applied experimentally to conduct such analyses.

Procedures for Construct Validation

Contributions to Cognitive Psychology

- Mathematical models have been developed to measure the contribution of different response components to item performance.
- Another widely used procedure for cognitive task analysis is protocol analysis (thinking aloud).
- It may find that the same item may evoke quite different cognitive processes in respondents who vary in experience.

Procedures for Construct Validation

Contributions to Cognitive Psychology

- Analyzing test performance in terms of specific cognitive processes should certainly strengthen and expand our understanding of what the tests measure.
- It may be pinpoint each person's sources of weakness and strength and thereby enhance the diagnostic use of tests.

Relationship between Reliability and Validity

- The maximum correlation coefficient between two variables is equal to square root of the product of their reliabilities.

$$r_{12\max} = \sqrt{r_{11}r_{22}}$$

- The random error of measurement suppress this correlation.

Relationship between Reliability and Validity

- Correction of Attenuation: Correlation between two true scores

$$r_{T_1T_2} = \frac{r_{12}}{\sqrt{r_{11}r_{22}}}$$

Integration of Validity Evidence

- Example: Arithmetic Test for different purposes

Testing Purposes	Illustrative Question	Evidence of Validity
Achievement test in elementary school arithmetic	How much has Dick learned in the past?	Content Validity
Aptitude test to predict performance in high school mathematics	How well will Jane learn in the future?	Predictive Validity
Technique for diagnosing learning disabilities	Does Bill's performance indicate specific disabilities?	Concurrent Validity
Measure of quantitative reasoning	How does Helen's score relate to other indicators of her reasoning ability?	Construct Validity



Integration of Validity Evidence

- This example highlights the fact that the choice of validation procedure depends on the use to be made of test scores.

Integration of Validity Evidence

- Content, criterion-related, construct validity do not correspond to distinct or logically coordinate categories.
- Construct validity is a comprehensive concept, that includes the other types of validity.
- All test use and all interpretation of test scores imply construct validity.