

Validity

the validity (content validity) is measured first, and then reliability.

Validity refers to a **test's accuracy**. “The soundness or appropriateness of a test or instrument in measuring what it is designed to measure” e.g: the exam is supposed to test the knowledge.

A test is valid when it measures what it is **intended to measure**. The intended uses for most tests fall into one of **three categories**, and each category is associated with a different method for establishing validity: (these types will be discussed in details later on)

- The test is used to obtain information about an examinee's familiarity with a particular content or behavior domain: **content validity**. **Face validity can be considered as part of content validity**.
- The test is administered to determine the extent to which an examinee possesses a particular hypothetical trait: **construct validity**.
- The test is used to estimate or predict an examinee's standing or performance on an external criterion: **criterion-related validity**.

Types of Experimental Validity

- Internal

- Is the experimenter measuring the effect of the independent variable on the dependent variable?

- External

- Can the results be generalised to the wider population?

Validity of Measurement

- **Meaning analysis**: the process of attaching meaning to concepts
 - Face validity
 - Content validity
- **Empirical analysis**: how well our procedures (operational definition) accurately observe what we want to measure
 - Criterion validity.
 - Construct validity

Face Validity

Done before content validity and it's not performed alone.

- A brief and hasty examination of an instrument. قراءة سطحية وسريعة للاستبيان
- Refers to whether the instrument looks as though it is measuring the appropriate construct. هل كل العناصر لهم علاقة بالمتغير، ويشملوا كل جوانبه أم لا؟
(attribute/ variable) The construct in face validity is usually qualitative.
- Based on judgment of experts in the content area, no objective criteria for assessment.

Content Validity

- The degree to which an instrument has an appropriate sample of items for the construct being measured.
- Concerned with the scope or range of items used to measure the variable, i.e. number and type of items to measure the concept.
- Infers that the test measures all aspects contributing to the variable of interest
- Evaluated by expert evaluation, via the content validity index (CVI)
e.g: the compliance of Tb patients to take antibiotics (monitoring them while taking the drug by specialists). The existence of other diseases that may affect the disease of interest such as HIV (increased susceptibility)..etc

Content Validity

first we make a form containing all items that are measured for our variable, and then give the form to at least 3 experts, they rate each item from 1 to 4 according to its relevance

(Item content validity index)

- To calculate an I-CVI, experts are asked to rate the relevance of each item, usually on a 4-point scale.
- There are several variations of labeling the 4 ordinal points, but the scale that seems to be used most often is 1 = not relevant, 2 = somewhat relevant, 3 = quite relevant, 4 = highly relevant
- I-CVI is computed as the number of experts giving a rating of either 3 or 4, divided by the number of experts (1,2) equal zero , (3,4) equal 1
- For example, an item rated as “quite” or “highly” relevant by 4 out of 5 judges would have an I-CVI of .80. → For this item
To measure the total CVI we take the average for all items.
CVI of less than .8 is unacceptable.

Criterion Validity

- The degree to which the instrument correlates with an external criterion
- Validity coefficient is calculated by correlating scores on the instrument and the criterion.
- This type of validity is used to measure the ability of an instrument to predict future outcomes.
- Validity is usually determined by comparing two instruments ability to predict a similar outcome with a single variable being measured.
- Two types of criterion-related validity:

(present)

- **Concurrent**: the degree to which a **test correlates** with **an external criteria that is measured at the same time** (e.g. does a depression inventory correlated with clinical diagnoses).
(test) (criterion)

(future)

- **Predictive**: the degree to which a **test predicts (correlates)** with an **external criteria that is measured some time in the future** (e.g. does a depression inventory score predict later clinical diagnosis).

e.g: weather predictions are the test and the weather state in the future is the criterion, if there is no correlation then the test is not valid

Criterion-Related Evidence

- Tells us how well a test corresponds with a particular criterion
 - criterion: behavioral or measurable outcome
 - SAT predicting GPA (GPA is criterion)
 - BDI scores predicting suicidality (suicide is criterion).
- Used to “predict the future” or “predict the present.”

Criterion-Related Evidence

- Predictive Validity Evidence **examples**
 - forecasting the future
 - how well does a test predict future outcomes
 - SAT predicting 1st yr GPA
 - most tests don't have great predictive validity
- decrease due to time & method variance

Criterion-Related Evidence

- Concurrent Validity Evidence **examples**
 - forecasting the present
 - how well does a test predict current similar outcomes
 - job samples, alternative tests used to demonstrate concurrent validity evidence
- generally higher than predictive validity estimates

Quantifying Criterion-Related Evidence

The doctor skipped the slides 12- 15

- Validity Coefficient
 - correlation between the test and the criterion
 - usually between .30 and .60 in real life.
 - In general, as long as they are statistically significant, evidence is considered valid.
- However,
 - recall that r^2 indicates explained variance.
 - SO, in reality, we are only looking at explained criterion variance in the range of 9 to 36%.
- Sound Problematic??

Concurrent Criterion Validity

- Concurrent criterion validity is used when the two instruments are used to measure the same event at the same time.
- Example: Exit polls during an election can be said to have concurrent criterion validity if they predict similar outcomes to the election.

Predictive Criterion Validity

- Predictive validity is used when the instrument is administered then time is allowed to pass and is measured against the another outcome.
- Example: During election season polls are taken of registered voters, the outcome of the polls are measured against the outcome of the election to determine their validity.

Criterion Validity and Quantitative Research

- When doing quantitative research it is important to have criterion validity with regard to survey instruments.
 - Instrument validity produces dependable raw data for further quantitative analysis.
 - Dependable raw data provides strong evidence for reaching a trustworthy conclusions.
- Criterion validity is useful when other instruments can be used for comparison and lead to the strongest relationship for a given variable.

Construct Validity

- The degree to which our operational definitions accurately reflect the concepts they were designed to measure
- Concerned with the questions:
 - What is this instrument really measuring?
 - Does it adequately measure the construct of interest?

Methods of Assessing Construct Validity

- **Known-groups technique** The measurement differentiates between disease x and disease y
- **Relationships based on theoretical predictions** If $x=y$, $y=z$ we can predict that $x=z$
- **Multitrait–multimethod matrix method (MTMM)**
- **Factor analysis**

You are just required to know the names, without the definitions (just read them).

Known-Groups Technique

- Assesses contrast validity.
- In this procedure, the instrument is administered to groups hypothesized to differ on the critical attribute because of a known characteristic.
- It is a method to support construct validity and provided when a test can discriminate between a group of individuals known to have a particular trait and a group who do not have the trait.
- Assess controlled versus uncontrolled blood pressure.

Relationships based on Theoretical Predictions

- It involves testing hypothesized relationships on the basis of theory or prior research.
- A researcher might reason as
 - According to the theory, construct X is positively related to construct Y.
 - Instrument A is a measure of construct X; instrument B is a measure of construct Y.
 - Scores on A & B are correlated positively, as predicted.
 - Therefore, it is inferred that A & B are valid measure of X & Y.

Multitrait–Multimethod Matrix Method

Builds on two types of evidence:

- Convergence
- Discriminability (Divergent)

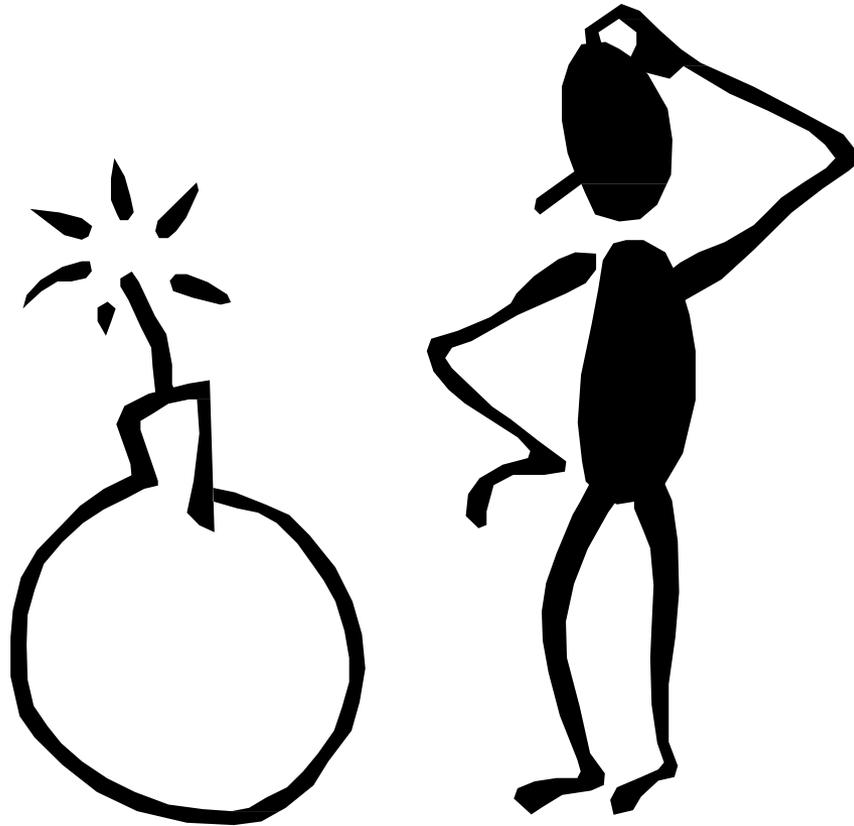
Convergence

- Evidence that different methods of measuring a construct yield similar results
- Convergent validity comes from the correlations between two different methods measuring the same trait

Discriminability

- Evidence that the construct can be differentiated from other similar constructs
- Discriminant validity assesses the degree to which a single method of measuring two constructs yields different results

Was it Clear Enough !



Edited by: Leen Hajeer