


Topic-VALIDITY AND RELIABILITY

4TH SEMESTER


BY:-

MRS.RINKI KUMARI
DEPTT.OF EDUCATION
MMHA&PU, PATNA
DATE:-7TH MAY,2020

INTRODUCTION

- ▶ The validity and Reliability of a test is important for data collection process.
 - ▶ Before and after collecting the data, the researcher need to consider the validity and reliability of their data.
- 


validity

- ▶ Validity refers to how accurately a method measures what is intended to measure. If research has high validity, that means it produces results that correspond to real properties, characteristics, and variations in the physical or social world.
 - ▶ High reliability is one indicator that a measurement is valid. If a method is not reliable, it probably isn't valid.
- 

Types of validity

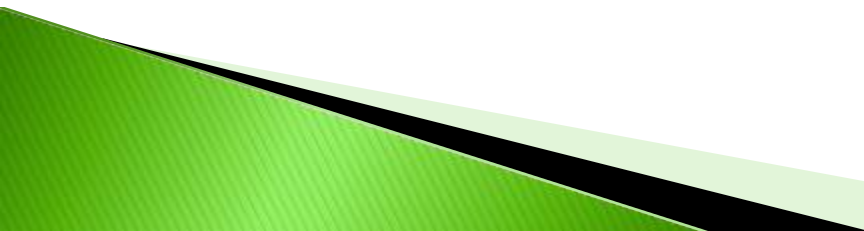
- Three Cs (conventionally)
 - Content
 - Criterion
 - Concurrent
 - Predictive
- Construct –
 - Convergent, discriminant, trait etc.,
- Others (face validity)

All types of validity are addressing the same issue of the degree of confidence we can place in the inferences we can draw from the scales

- **Face validity**
 - On the face of it the tool appears to be measuring what it is supposed to measure
 - Subjective judgment by one/more experts, rarely by any empirical means
 - **Content validity**
 - Measures whether the tool includes all relevant domains or not
 - Closely related to face validity
 - aka. ‘validity by assumption’ because an expert says so
 - Certain situations where these may not be desired
- 

Content validity

- Example – cardiology exam;
 - Assume it contains all aspects of the circulatory system (physiology, anatomy, pathology, pharmacology etc., etc.,)
 - If a person scores high on this test, we can say ‘infer’ that he knows much about the subject (i.e., our inferences about the person will ring true across various situations)
 - In contrast, if the exam did not contain anything about circulation, the inferences we make about a high scorer may be wrong most of the time and vice versa

- Generally, a measure that includes a more representative sample of the target behaviour will have more content validity and hence lead to more accurate inferences
 - Reliability places an upper limit on validity (the maximum validity is the square root of reliability coeff.) the higher the reliability the higher the maximum possible validity
 - One exception is that between internal consistency and validity (better to sacrifice IC to content validity)
 - The ultimate aim of scale is inferential which depends more on content validity than internal consistency
- 

Criterion validity

- Correlation of a scale to an accepted 'gold standard'
- Two types
 - **Concurrent** (both the new scale and standard scale are given at the same time)
 - **Predictive** – the Gold Standard results will be available some time in the future (eg. Entrance test for college admission to assess if a person will graduate or not)
- Why develop a new scale when we already have a criterion scale?
 - Diagnostic utility/substitutability (expensive, invasive, dangerous, time-consuming)
 - Predictive utility (no decision can be made on the basis of new scale)
- Criterion contamination
 - If the result of the GS is in part determined in some way by the results of the new test, it may lead to an artificially high correlation

Construct validity

- Height, weight – readily observable
- Psychological - anxiety, pain, intelligence are abstract variables and can't be directly observed
- For eg. Anxiety – we say that a person has anxiety if he has sweaty palms, tachycardia, pacing back and forth, difficulty in concentrating etc., (i.e., we have a *hypothesize* that these symptoms are the result of anxiety)
- Such proposed underlying factors are called hypothetical constructs/ constructs (eg. Anxiety, illness behaviour)
- Such constructs arise from larger theories/ clinical observations
- Most psychological instruments tap some aspect of construct

Establishing construct validity

- IBS is a construct rather than a disease – it is a diagnosis of exclusion
- A large vocabulary, wide knowledge and problem solving skills – *what is the underlying construct?*
- Many clinical syndromes are constructs rather than actual entities (schizophrenia, SLE)

- Initial scales for IBS – ruling out other organic diseases and some physical signs and symptoms
 - These scales were inadequate because they lead to many missed and wrong diagnoses
 - New scales developed incorporating demographical features and personality features
- Now how to assess the validity of this new scale
 - Based on theory high scorers on this scale should have
 - Symptoms which will not clear with conventional therapy
 - Lower prevalence of organic bowel disease on autopsy

- ***Convergent validity*** - If there are two measures for the same construct, then they should correlate with each other but should not correlate too much.
E.g. Index of anxiety and ANS awareness index
- ***Divergent validity*** – the measure should not correlate with a measure of a different construct, eg. Anxiety index and intelligence index

FACTORS AFFECTING VALIDITY

1. Environmental factors
 2. Research factors
 3. Instrumentation factors
 4. Subject factors
- 

CONT....


- **Environmental factors:** Some environmental factors such as room temperature, lighting and noise can influence the error rate
- **Research factors:** The researcher can influence the results of the study in many ways
- **Instrumentation factors:** An inadequate sampling of items is the other common source of instrumentation errors.
- **Subject factors:** Any changing physical, emotional or psychological state of the subject could introduce error into the measurement process.

Reliability

The reliability of a measure indicates the extent to which it is without bias (error free) and hence ensures consistent measurement across time and across the various items in the instrument.

In other words, the reliability of a measure is an indication of the *stability* and *consistency* with which the instrument measures the concept and helps to assess the “goodness” of measure

Definition

- ▶ In the word of Anne Anastasi, “Reliability refers to the consistency of scores obtained by the same individuals when re-examined with the same test on different occasions or with different sets of equivalent items or under other variable examining conditions.”
- 

Methods of determining the reliability

- **Internal consistency procedures** – items or questions measuring the same phenomenon, should produce similar results irrespective of their number in an instrument
 - The split-half technique
- **External consistency procedures** – compare findings from two independent process of data collection with each other as a means of verifying the reliability of the measure
 - Test/retest
 - Parallel form of the same test

Split-half technique

- To correlate half of the items with the other half in a research instruments
- Questions are divided in half in such way that any two questions intended to measure the same aspect fall into different halves.
- The scores obtained by administering the two halves are correlated.
- Reliability is calculated using product moment correlation between scores


Test/ retest (repeatability test)

- An instrument is administered once, and then again, under the same or similar conditions.
- The ratio between test and retest score is an indication of the reliability of the instrument
 - The greater the value of the ratio, the higher the reliability of the instrument
- Advantage - it permits the instrument to be compared with itself.
- Disadvantage – a respondent may recall the responses that they gave in the first round
 - Overcome by increasing the time span between two tests

Parallel forms of the same test

- Two instrument intended to measure the same population is constructed and administered to two similar population.
- The results obtained from one test is compared with another
 - If similar, the instrument is reliable.
- Advantage – does not suffer from the problem of recall and time lapse between two test is not required.
- Disadvantage – need to construct two instrument instead of one.

Factors Affecting Reliability

- ▶ Length of the test
 - ▶ Range of individual differences in the Sample Tested.
 - ▶ Average Ability Level of the Examinees.
 - ▶ Testing conditions and Scoring.
 - ▶ Guessing and chance errors.
 - ▶ Construction of the test.
- 

RELATIONSHIP BETWEEN VALIDITY & RELIABILITY

Validity and reliability are closely related.

A test cannot be considered valid unless the measurements resulting from it are reliable.

Likewise, results from a test can be reliable and not necessarily valid.

Reliability vs. Validity



Reliable
Not Valid



Low Validity
Low Reliability



Not Reliable
Not Valid



Both Reliable
and Valid