PREPARED BY

MRS.KALPANA

ASSO.PROFESSOR

ICON

1. VALIDITY AND RELIABILITY

VALIDITY:

Validity of an instrument refers to the degree to which an instrument measures what it is supposed to be measuring. For example, a temperature-measuring instrument is supposed to measure only the temperature; it cannot be considered a valid instrument if it measures an attribute other than temperature.

Similarly, if a researcher developed a tool to measure the pain, and if it also includes the items to measure anxiety, it cannot be considered a valid tool. Therefore, a valid tool should only measure what it supposed to be measuring.

DEFINITIONS

- According to Treece and Treece, 'Validity refers to an instrument or test actually testing what it is supposed to be testing'.
- According to Polit and Hungler, 'Validity refers to the degree to which an instrument measures what it is supposed to be measuring'.
- According to American Psychological Foundation, 'Validity is the appropriateness, meaning, fullness, and usefulness of the interference made from the scoring of the instrument'.
- Validity is the appropriateness, completeness, and usefulness of an attribute measuring research instrument

TYPES OF VALIDITY

Basically validity is classified into following four categories.

i) **Statistical conclusion validity:** concerns the validity of inferences that there truly is an empirical relationship, or correlation, between the presumed cause and the effect. The researcher's job is to provide the strongest possible evidence that an observed relationship is real

ii) **Content validity:** It is concerned with scope of coverage of the content area to be measured. More often it is applied in tests of knowledge measurement. It is mostly used in measuring complex psychologic tests of a person. It is a case of expert judgment about the content area included in the research instrument to measure a particular phenomenon, Judgment of the content viability may be subjective and is based on previous researchers and experts opinion about the adequacy, appropriateness, and completeness of the content of instrument Generally, this viability is ensured through the judgments of experts about the content

Face validity: Face validity involves an overall look of an instrument regarding its appropriateness to measure a particular attribute or phenomenon.

Though face validity isnot considered a very important and essential type of validity for an instrument. However, it may be taken in consideration while assessing for other aspects of validity of a research instrument.

In simple words, this aspect of validity refers to the face value or the outlook of an instrument.

For example, a Likert scale designed to measure the attitude of the nurses towards the patients admitted with HIV/AIDS; a researcher may judge the face value of this instrument by its appearance, that is it looks good or not, but it does not provide any guarantee about the appropriateness and completeness of a research instrument with regard to its content, construct, and measurement score

iii) **Criterion validity:** This type of validity is a relationship between measurements of the instrument with some other external criteria. For example, a tool is developed to measure the professionalism among nurses and to assess the criterion validity nurses were separately asked about the number of research papers they published and number of professional conferences they have attended. Later a correlation coefficient is calculated to assess the criterion validity. This tool is considered strong with criterion validity if a positive correlation exists between score of the tool measuring professionalism and the number of research articles published and professional conferences attended by the nurses

a) **Predictive validity:** It is the degree of forecasting judgment; for example, some personality tests on academic futures of students can be predictive of behaviour patterns. It is the differentiation between performances on some future criterion and instruments ability. An instrument may have predictive validity when its score significantly correlates with some future criteria.

b) **Concurrent validity:** It is the degree of the measures in present. It relates to the present specific behaviour and characteristics; hence the difference between predictive and concurrent validity refers to timing pattern of obtaining measurements of a criterion.

iv) **Construct validity:** A construct is founded in this type of validity, such as a nurse may have designed an instrument to measure the concept of pain in amputated patients.

v) The pain pattern may be due to anxiety; hence the results may be misleading.

vi) Construct validity is a key criterion for assessing the quality of a study, and construct validity has most often been addressed in terms of measurement issues.

Construct validity involves the validity of inferences "from the observed persons, settings, and cause and effect operations included in the study to the constructs that these instances might represent".

One aspect of construct validity concerns the degree to which an intervention is a good representation of the underlying construct that was theorized as having the potential to cause beneficial outcomes.

Another concerns whether the measures of the dependent variable are good operationalization of the constructs for which they are intended.

Construct validity gives more importance to test relationship predicted on theoreticalmeasurements.

The researcher can make prediction in relation to other such type of constructs. One method of construct validation is known as group technique.

vii) **Internal validity:** concerns the validity of inferences that, given that an empirical relationship exists, it is the independent variable, rather than something else that caused the outcome. Researchers must develop strategies to rule out the plausibility that some factor other than the independent variable accounts for the observed relationship.

viii) **External validity:** concerns whether inferences about observed relationships will hold over variations in persons, setting, time, or measures of the outcomes. External validity, then, is about the generalizability of casual inferences, and this is a critical concern for research that aims to yield evidence for evidence-based nursing practice.

CONTROLLING CONFOUNDING PARTICIPANT CHARACTERISTICS:

Randomization:

It is the most effective method of controlling individual characteristics.

The function of randomization is to secure comparable groups – that is, to equalize groups with respect to confounding variables.

A distinct advantage of random assignment, compared with other strategies, is that it can control all possible sources of confounding variation, without any conscious decision about which variables need to be controlled.

Crossover:

Randomization within a crossover design is an especially powerful method of ensuring equivalence between groups being compared – participants serve as their own controls.

Moreover fewer participants usually are needed in such a design.

Fifty people exposed to two treatments in random order yield 100 pieces of data (50×2); 50 people randomly assigned to two different groups yield only 50 pieces of data (25×2).

Homogeneity:

When randomization and crossover are not feasible, alternative methods of controlling confounding characteristics are needed. One method is to use only people who are homogenous with respect to confounding variables – that is, confounding traits are not allowed to vary.

Stratification/ blocking:

Another approach to controlling confounders is to include them in the research design through stratification.

To pursue our example of the physical fitness intervention with gender as the confounding variable, we could use a randomized block design in which men and women are assigned separately to treatment groups.

Matching:

Matching (also called pair matching) involves using information about people's characteristics to create comparable groups.

If matching were used in our physical fitness example, and age and gender were the confounding variables, we would match a person in the intervention group with one in the comparison group with respect to age and gender.

Statistical control:

Another method of controlling confounding variables is through statistical analysis rather than research design.

A detailed description of powerful statistical control mechanisms will be postponed until chapter 18, but we will explain underlying principles with a simple illustration of a procedure called analysis of covariance (ANCOVA)

Evaluation of control methods:

Randomization is the most effective method of managing confounding variables – that is, of approximating the ideal but unattainable counterfactual discussed in chapter 9 – because it tends to cancel out individual differences on all possible confounders.

FACTORS AFFECTING VALIDITY:-

History:events that occur besides the treatment (events in the environment).

Maturation:physical or psychological changes in the participants.

Testing:effect of experience with the pretest become test wise.

Instrumentation: learning gain might be observed from pre to posttest simply due to nature of the instrument. Particularly a problem in observation studies when observers more likely to give ratings based on expectations (conscious or subconscious).

Statistical Regression: Tendency for participants whose scores fall at either extreme on a variable to score nearer the mean when measured a second time.

Differential Selection: Effect of treatment confounded with other factors because of differential selection of participants, problem in non-random samples.

Experimental Mortality:participants lost from the study, attrition.

Selection-maturation Interaction:similar to differential selection, except maturation is the confounding variable

Experimental Treatment Diffusion: Treatment is perceived as highly desirable and members of control group seek access

Compensatory Rivalry by Control Group: (John Henry Effect) -- control group performs beyond expectations because they perceive they are in competition with experimental group

Compensatory Equalization of Treatments: occurs when experimental group received goods or services perceived as desirable and control group is given similar goods and services on compensate. Not comparing treatment with no treatment but one treatment with another.

Resentful Demoralization of Control Group: Control group becomes discouraged because they perceive experimental group is receiving a desirable treatment that is being withheld from them.

RELIABILITY

The reliability of a quantitative measure is a major criterion for assessing its quality. Reliability, broadly speaking, is the extent to which scores are free from measurement error.

DEFINITION:

Reliability is the extent to which scores for people who have not changed are the same for repeated measurements, under several situations, including repetition on different occasions, by different persons, or on different versions of a measure, or in the form of different items on a multi-item instrument (internal consistency).

RELIABILITY

The first component within the broad reliability domain is simply called reliability. It covers four different approaches to reliability assessment, including the following:

- **Test-retest reliability:** administration of the same measure to the same people on two occasions (repetition over occasions)
- **Interrater reliability:** measurements by two or more observers or raters using the same instrument (repetition over persons)
- **Intrarater reliability:** measurements by the same observer or rater on two or more occasions (repetition over occasions)
- **Parallel test reliability:** measurements of the same attribute using alternate versions of the same instrument, with the same people (repetition over versions)

i) TEST-RETEST RELIABILITY

In test-retest reliability, replication takes the form of administering a measure to the same people on two occasions. If a measure yields a good estimate of the true scores of an attribute, ideally, it will do so comparably on separate administrations.

The assumption is that for traits that have not changed, any differences in people's scores on the two testing are the result of measurement error. When score differences across waves are small, reliability is high. This type of reliability is sometimes called stability or reproducibility--the extent to which scores can be reproduced on repeated administrations.

ii) INTERRATER AND INTRARATER RELIABILITY

When measurements involve the use of an observer who makes scoring judgments, a key source of measurement error can stem from the person making the measurements.

The most typical approach is to undertake an interrater (or interobserver) reliability assessment, which involves having two or more observers independently applying the instrument with the same people.

Reliability assessment involves comparing the observers' scores to see if the scores are comparable.

A less frequently used approach-but one that is appropriate in many clinical situations is an intrarater reliability assessment in which the same rater makes the measurements on two or more occasions, blinded to the ratings assigned previously.

Intrarater reliability is an index of self-consistency. It is analogous to retest reliability, except that the focus in retest situations is the consistency of the person being measured, and intrarater reliability concerns the consistency of the person making the measurements. Like retest reliability, intrarater reliability assessments require a carefully selected interval between testing.

iii) PARALLEL TEST RELIABILITY

Multi-item parallel tests (or alternative form tests) are not common in health care measurement, but there are a few examples.

For instance, the latest version of the Mini-Mental State Examination (MMSE-2), a measure of cognitive impairment, has alternate forms (Folstein et al., 2010).

Parallel tests can be created by randomly sampling two sets of items from a carefully developed item pool.

If the two tests are, indeed, parallel, then they are replicates whose true score are identical.

Having measures that are useful when researchers expect to make parallel is measurements in a fairly short period of time and want to avoid carryover biases.

iv) INTERPRETATION OF RELIABILITY COEFFICIENTS

Reliability coefficients are important indicators of an instrument's quality. Unreliable measures reduce statistical power and hence affect statistical conclusion validity. If data fail to support a hypothesis, one possibility is that the instruments were unreliable—not necessarily that the expected relationships do not exist.

v) SPLIT-HALF METHOD

Divide items of a research instrument in two equal parts through grouping either in odd number question and even number question or first-half and second-half item groups.

MEASURING OF RELIABILITY

There are several ways to measure the reliability for the research tools, which depends on several factors, such as the nature of instrument as well as aspects of reliability the researcher wants to measure.

The main aspects of the reliability considered important in quantitative research include: stability, internal consistency, and equivalence.

One of the following aspects may be considered important to measure the reliability of a research instrument.

Stability: The stability aspect of reliability means research instrument provides same results when used consecutively for two or more times. Stability is estimated to make sure that research instrument is consistent in providing similar results with repeated administration,

It is also known as reliability of test-retest function.

To measure test-retest reliability, the test is given twice at two different points in time. It is used to measure the consistencyof a test across time.

This type of reliability supposes that there will be no change in the quality or construct being assessed.

Test-retest reliability is suitably best used for things that are strong over time (e.g. intelligence). Reliability will be, generally, higher when little time has passed between tests.

The test-retest method is a relatively easy and straightforward approach to establish reliability.

It is used for questionnaire, observation checklist, observation-rating scales, and physiologic measurement tools.

However, it has one basic problem that there are many traits that may change with time, such as attitude, behaviour, mood knowledge, satisfaction, physical conditions, and so on.

2. DIFFERENCE BETWEEN QUALITATIVE AND QUANTITATIVE DATA ANALYSIS

Sl.no	Qualitativedataanalysis	Quantitativedataanalysis

1.	Understandandinterpretsocialinteractions	Testhypotheses,checkthecauseandeffect.developpredictionsforthefuture
2.	Small,selectedintentionally	Largerandselectedrandomly
3.	Words,images,objects	Numbersandstatistics
4.	Open-endedresponses,interviews,participantobservations,fieldnotes	Precisemeasurementsusingstructuresandvaidatedinstrumentsfordatacollection
5.	Patterns,features,themesidentification	Statisticalrelationshipsidentification
6.	Researchermaybeknowntoparticipantsinthestudyandparticipantscharacteristicsmaybeknowntotheresearchers	Researcherandtheirbiasesarenotknowntotheparticipantcharacteristicsarehidden
7.	Particularfindings,lessgeneralizable	Generalizablefindings,canbeappliedtotheotherpopulations
8.	Beginswihmoregeneralopen-endedquestions,movingtowardgreaterprecisionasmoreinformationemerges	Keyexpalnatoryandoutcomevariablesidentifiedinadvance
9.	Pre-definedvariablesarernotidentifiedinadvance	Contextual/confoundingvariablesidentifiedandcontrolled
10.	Peliminaryanalysisisaninherentpartofdatacollection	Datacollectionandanalysisdistinctlyseparatephases Analysisuseformalstatisticalprocedures
11.	Itisasubjectiveanalysisthatismoreconcernedwithnon-statisticaldatathatcannotbecompared	Itisanobjectiveanalysisthatquantitativedata
12.	Typicaldataincludecolour,gender,nationality,religionandmanymore	Typicaldataincludemeasurablequantitiessuchaslength,size,weight,massandmanymore
13.	Theanalysisisusedtounderstandwhyacertainphenomenonoccurs	Theanalysisisconcernedwithhowmanyorhowmuchacertainphenomenonoccurs
14.	Interpretsandunderstandsocialinteractions	Testhypothesesandgivefuturepredictions

15.	Researchmethodologyisexploratory	Researchmethodologyisoftenconclusive.