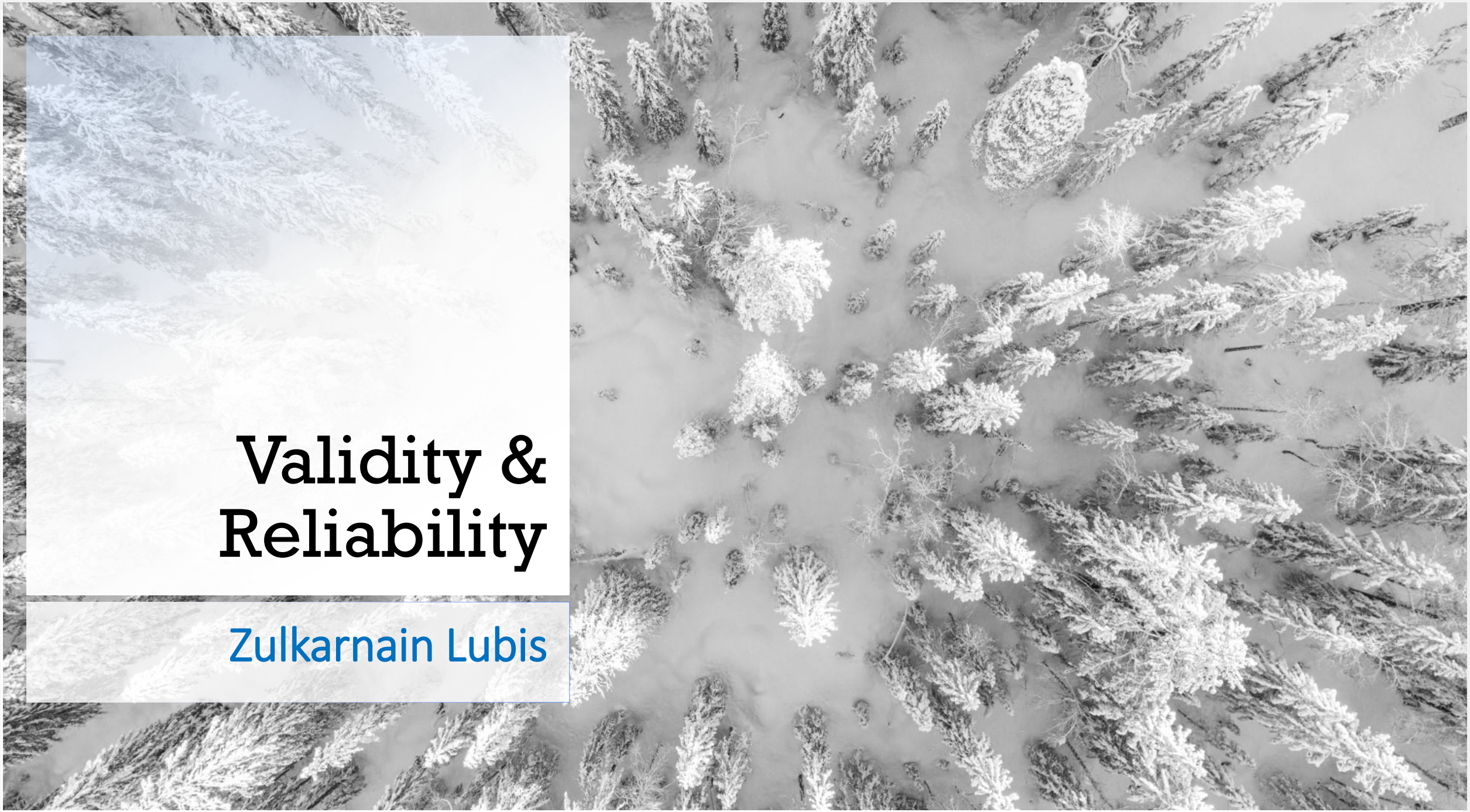


Validity & Reliability

Zulkarnain Lubis

What is Reliability?

Reliability is:

the consistency of your measurement instrument

What do I mean by…?

the degree to which an instrument measures the same way each time it is used under the same condition with the same subjects

- Your ruler…

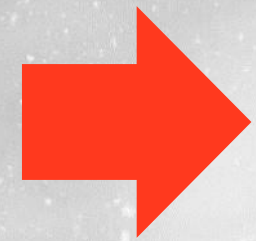
  - was **consistent**
  - measured the same way each time it was used under the same condition with the same object

The book did not change and therefore the ruler reported back the same measurement

Your ruler is **RELIABLE**

Reliability

All of these tools of measurement are reliable given the same temperature, time and volume, they measure the same consistently.



Every time the water begins to boil the kettle whistles

Every time it is 6:30 my alarm clock goes off

Every time I use the little plastic up I have 2 tablespoons of medicine

Reliability

- **Reliability alone does not mean that you have a good instrument however.**
- **Imagine the following reliable instrument:**
  - Every morning Margie gets on the scale and every morning it reads 165 lbs.
  - It seems pretty reliable since Margie hasn't gained or lost any weight
  - One day an elephant got on Margie's scale and it still read 165 lbs.!
- **How is this scale reliable?**
  - It does measure the same way under the same conditions (Margie's) – but a lot of other conditions too (the elephant and who knows what else)

What is Validity?

- Validity asks
  - if an instrument measures what it is supposed to
  - how “true” or accurate the measurement is

Reliable but not Valid

These instruments are very RELIABLE

They both report consistently – <u>too consistently</u>

But, neither measures what it is supposed to:

- The scale is not really measuring weight
- The clock is not measuring time

They are NOT VALID

Putting Reliability and Validity Together

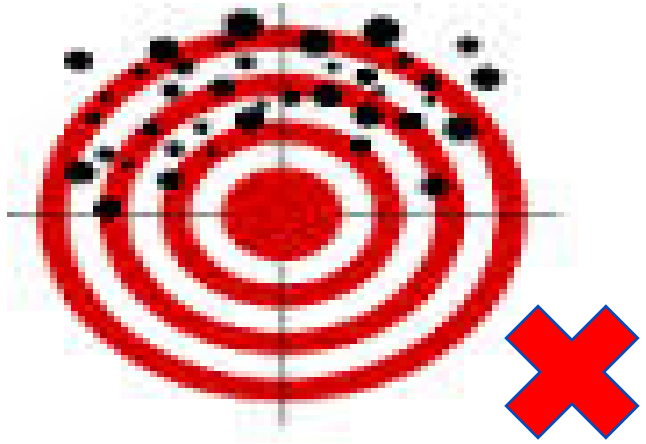
Every instrument can be evaluated on two dimensions:

Reliability

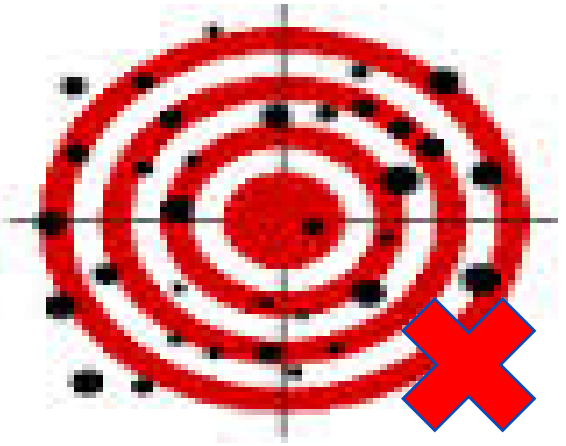
How consistent it is given the same conditions

Validity

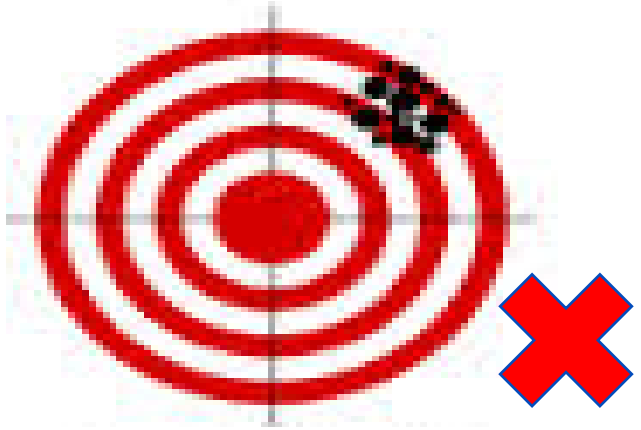
If it measures what it is supposed to and how accurate it is



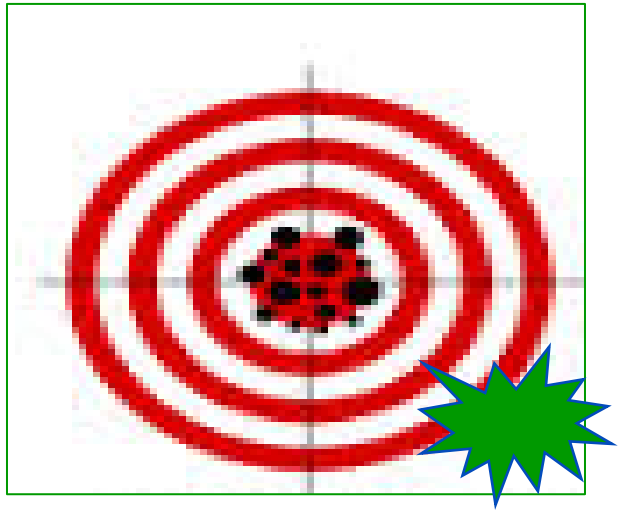
Unreliable and Not Valid



Unreliable but Valid



Reliable but Not Valid



Reliable and Valid

VALIDITY

- *An instrument is valid if measures what it purpose to measure*
- Validity concerns:
  - the form of the instrument
  - the purpose of the instrument
  - the population for whom it is intended
- Validity asks
  - if an instrument measures what it is supposed to
  - how “true” or accurate the measurement is

- Therefore, we cannot ask the general question “Is this a valid instrument?”.
- The question to ask is *“how valid is this instrument for the decision that I need to make?”* or *“how valid is the interpretation I propose for the instrument?”*

Validity Test

- Validity test is used to determine the level of validity of the instrument (questionnaire) for collecting data
- Product Moment Correlation: A statistical tool usually used to test the validity of and instrument (construct validity)
- The level of validity is measured by the coefficient of correlation between the scores of each indicator / item questions (Xi) with a total score / factor (X). It can be determined by any computer software; MINITAB, SPSS
- Using SPSS: Analyze, Correlate, Bivariate, Variables, Total Variable

What are some ways to improve validity?

- Make sure your goals and objectives are clearly defined and operationalized.
- Match your assessment measure to your goals and objectives.
- Get students involved; have the students look over the assessment for troublesome wording, or other difficulties.
- If possible, compare your measure with other measures, or data that may be available.

Reliability

Reliability of the instrument can be tested in two ways,

- External reliability test; measure or criteria are outside the instrument
- Internal reliability test; measure or criteria based on data obtained from the instrument

External reliability is tested by using parallel technique (double test double trial)

To test the internal reliability can be done in several ways Alpha formula is one among them

Types of Reliability

- **Test-retest reliability (external)**
- **Parallel forms reliability** (external)
- **Inter-rater reliability**
- **Internal consistency reliability**
  - **Average inter-item correlation**
  - **Split-half reliability** (external)

Test-retest reliability

- Test-retest reliability is a measure of reliability obtained by administering the same test twice over a period of time to a group of individuals. The scores from Time 1 and Time 2 can then be correlated in order to evaluate the test for stability over time
- The degree to which scores are consistent over time. It indicates score variation that occurs from testing session to testing session as a result of errors of measurement.

Parallel forms reliability

- **Parallel forms reliability** is a measure of reliability obtained by administering different versions of an assessment tool (both versions must contain items that probe the same construct, skill, knowledge base, etc.) to the same group of individuals. The scores from the two versions can then be correlated in order to evaluate the consistency of results across alternate versions.
- **Parallel-forms reliability** is gauged by comparing two different tests that were created using the same content
- The two tests should then be administered to the same subjects at the same time.
- The scores on the two test forms are correlated in order to determine how similarly the two test forms function

Inter-rater reliability

- Inter-rater reliability is a measure of reliability used to assess the degree to which different judges or raters agree in their assessment decisions. Inter-rater reliability is useful because human observers will not necessarily interpret answers the same way; raters may disagree as to how well certain responses or material demonstrate knowledge of the construct or skill being assessed
- By using SPSS, to run this analysis in the menus, specify Analyze>Descriptive Statistics>Crosstabs, specify one rater as the row variable, the other as the column variable, click on the Statistics button, check the box for Kappa, click Continue and then OK.
- RELIABILITY offers a set of intra class correlation coefficients (ICCs) designed for two or more raters rating objects, normally on an interval scale. Specify Analyze>Scale>Reliability Analysis. Specify the raters as the variables, click on Statistics, check the box for Intra class correlation coefficient, choose the desired model, click Continue, then OK.

Internal consistency reliability

- Internal consistency reliability is a measure of reliability used to evaluate the degree to which different test items that probe the same construct produce similar results
  - Average inter-item correlation; It is obtained by taking all of the items on a test that probe the same construct, determining the correlation coefficient for each *pair* of items, and finally taking the average of all of these correlation coefficients. This final step yields the average inter-item correlation
  - Split-half reliability; It is begun by "splitting in half" all items of a test that are intended to probe the same area of knowledge in order to form two "sets" of items. The *entire* test is administered to a group of individuals, the total score for each "set" is computed, and finally the split-half reliability is obtained by determining the correlation between the two total "set" scores.

Split-half reliability

- The process of obtaining split-half reliability is begun by “splitting in half” all items of a test that are intended to probe the same area of knowledge.
- The *entire* test is administered to a group of individuals, the total score for each “set” is computed, and finally the split-half reliability is obtained by determining the correlation between the two total “set” scores.
- Gutman Split-Half Coefficient as a reliability coefficient derived from correlation coefficient between the two total “set” scores, by using the formula:

$$r = \frac{2r_{12}}{1 + r_{12}}$$

- Split-half reliability can be calculated by using SPSS

- Cronbach's alpha was developed by Cronbach's alpha coefficient (1951) as a general measure of the internal consistency of multi-item scale
- Cronbach's alpha is calculated by using the formula :

$$\alpha = \left(\frac{k}{k-1}\right)\left(1 - \frac{\sum \sigma b^2}{(\sigma^2 t)}\right)$$

- If cronbach's alpha > $r_{table(\alpha, n-1)}$, considered the instrument reliable
- Items which have "corrected item-total correlation" <$r_{table(\alpha, n-1)}$ considered unreliable
- Using Cronbach's alpha to test reliability can be performed by SPSS as the following: *Analyze, Scale, Reliability Analysis, select items, statistics, descriptive for, item, scale, scale if item deleted, continue, OK*

Some Methods to Calculate Internal Reliability besides Alfa Cronbach

- Spearman Brown (split half)
- *Flanagant*
- *Rulon*
- KR 20
- KR 21
- Analyisis of Variance Hyot)

Zulkarnain Lubis