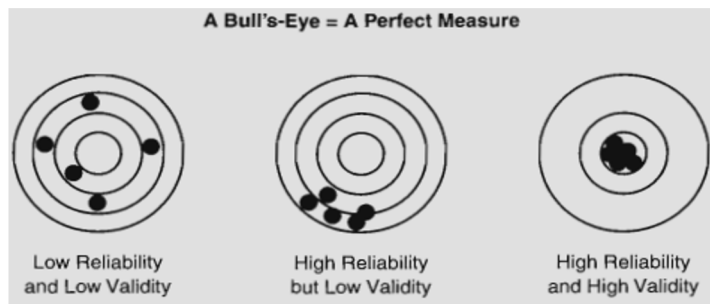


RELIABILITY & VALIDITY IN SOCIAL RESEARCH

Dr. Syed Owais

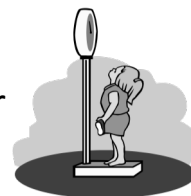
Department of Sociology, University of Peshawar



1

Reliability

- When a *Measurement Procedure* yields **consistent scores when the phenomenon being measured is not changing.**
- Consistency of measurement.
- Reliability means dependability or consistency” (Neuman, 1996: 109)
- **For Example: *Measuring height with reliable bathroom scale***



03/04/2020

Dr. Syed Owais, Department of Sociology,
University of Peshawar

2

2

Importance

- “Reliability is that property of a measuring device [e.g., questionnaire, interview guide, observational schedule] for social phenomena which yields consistent measurements when the phenomena are stable, regardless of who uses it, provided the basic conditions remain the same” (Payne & Payne, 2004: 203).
- “The extent to which a measuring instrument, for example a test to measure intelligence, gives consistent results” (Jupp, 2006: 262).

03/04/2020

Dr. Syed Owais, Department of Sociology,
University of Peshawar

3

3

Types of Reliability Measurement

- **1. Stability Reliability**
- **2. Equivalence Reliability**
- **3. Representative Reliability**

03/04/2020

Dr. Syed Owais, Department of Sociology,
University of Peshawar

4

4

1) Stability Reliability

- It's **reliability across time**. i.e.,
 - Does the measure (the questionnaire, e.g.,) deliver the same answer when applied in different time periods? (Neuman, 2014: 212).
 - Administering, e.g., the same questionnaire at 2 different times. (i.e., Test-retest method)

| | |
|------------------|------------------|
| T ₁ | T ₂ |
| Obs ₁ | Obs ₂ |

03/04/2020

Dr. Syed Owais, Department of Sociology,
University of Peshawar

5

5

1. Stability Reliability

However:

- Only works if phenomenon is unchanging
 - i.e., if a long span of time is involved then changes in respondents circumstances might change his/her responses;
- Respondents answering at T1 (time 1) might influence how they reply at T2 (time 2) ...

03/04/2020

Dr. Syed Owais, Department of Sociology,
University of Peshawar

6

6

2. Representative Reliability

- “Reliability across subpopulations or groups of people”. Addresses the Question:
 - “does the indicators [questionnaire] deliver the same answer when applied to different groups (e.g., Different classes, ethnic groups, sexes, age groups)?”
- e.g., : Asking about age from different age group; some overstate it, others understate it
 - lack of representative reliability!

03/04/2020

Dr. Syed Owais, Department of Sociology,
University of Peshawar

7

7

Cont'd...

Subpopulation analysis method (to determine stability reliability)

- **i.e.** Asking the same question from different categories in a population and then independently cross-checking the information through other means.
- **E.g.** Asking level of education, then checking school records etc. (representative reliability, if there are no differences in the error rate for men and women)

03/04/2020

Dr. Syed Owais, Department of Sociology,
University of Peshawar

8

8

3. Equivalence Reliability

“it applies when researchers use *multiple indicators* – that is when, *multiple specific measures are used in the operationalisation of construct* (i.e. Several items in a questionnaire all measure the same construct).

How to check?

1. Inter-item/split-half (split ½)
2. Interobserver/Intercoder reliability/Agreement: Is every observer scoring the same?

03/04/2020

Dr. Syed Owais, Department of Sociology,
University of Peshawar

9

9

Cont'd...

- Inter-item/split-half (split ½)

When there is a long questionnaire, we use this method to determine equivalence reliability.

- E.g., you have 4 questions to measure the sub-dimension 'leadership' (as part of ATTRIBUTES in the definition of QUALITY EDUCATION)
 - Divide them in half with 2 in each group.
 - allocate each group on a random (or an odd–even basis). The degree of correlation between scores on two halves would then be calculated with the aim to establish whether respondents scoring high on one of the two groups also scored high on the other group of indicators.
 - most researchers use a test of internal/equivalence reliability known as **Cronbach's alpha** (covered in **Statistics** course).

03/04/2020

Dr. Syed Owais, Department of Sociology,
University of Peshawar

10

10

How to Improve Reliability

1. Clearly Conceptualize All Constructs.

- Reliability increases when a single construct or sub-dimension of a construct is measured. This means developing unambiguous, clear theoretical definitions. Constructs/concepts should be specified to eliminate "noise" (i.e., distracting or interfering information) from other constructs.
- Each measure should indicate **one and only one concept.**

03/04/2020

Dr. Syed Owais, Department of Sociology,
University of Peshawar

11

11

2. Increase the Level of Measurement.

- Indicators at higher or more precise levels of measurement are more likely to be reliable than less precise measures because the latter pick up less detailed information.
- If more specific information is measured, then it is less likely that anything other than the construct will be captured.

03/04/2020

Dr. Syed Owais, Department of Sociology,
University of Peshawar

12

12

3. Use Multiple Indicators of a Variable.

- Two indicators of the same construct are better than one & three indicators of the same construct are much better than two indicators and so on!

For example, **three indicators** of the variable, racial-ethnic prejudice. My **first indicator** is an attitude question on a survey. I ask research participants their beliefs and feelings about many different racial and ethnic groups.

For a **second indicator**, I observe research participants from various races and ethnic groups interacting together over the course of three days. I look for those who regularly either (1) avoid eye contact, appear to be tense, and sound cool and distant; or (2) make eye contact, appear relaxed, and sound warm and friendly as they interact with people of their same or with people of a different racial-ethnic group.

03/04/2020

Dr. Syed Owais, Department of Sociology,
University of Peshawar

13

13

Cont'd...

- Last, I **create an experiment**. I ask research participants to read the grade transcripts, resumes, and interview reports on 30 applicants for five jobs youth volunteer coordinator, office manager, janitor, clothing store clerk, and advertising account executive. The applicants have many qualifications, but I secretly manipulate their racial or ethnic group to see whether a research participant decides on the best applicant for the jobs based on an applicant's race and ethnicity.

03/04/2020

Dr. Syed Owais, Department of Sociology,
University of Peshawar

14

14

4. Use Pre-tests, Pilot Studies, and Replication.

- Reliability can be improved by using a pre-test/pilot version of a measure first. Develop one or more draft or preliminary versions of a measure and try them before applying the final version.
 - This takes more time and effort.
- If while search the literature you find measures (i.e., questionnaire etc.) from past research, go ahead and use it.
 - build on and use a previous measure if it is a good one (citing the source!).
 - In addition, you may also want to add new indicators and compare them to the previous measure.

03/04/2020

Dr. Syed Owais, Department of Sociology,
University of Peshawar

15

15

VALIDITY

1. The extent to which measures indicate what they are intended to measure. **OR**
2. The match between the conceptual definition and the operational definition.



03/04/2020

Dr. Syed Owais, Department of Sociology,
University of Peshawar

16

16

4 Types of Measurement Validity

1. Face Validity
2. Content Validity
3. Criterion Validity
 1. Concurrent Validity
 2. Predictive Validity
4. Construct Validity
 1. Convergent Validity
 2. Discriminant Validity

03/04/2020

Dr. Syed Owais, Department of Sociology,
University of Peshawar

17

17

Types of Measurement Validity

Face validity A type of measurement validity in which an indicator “makes sense” as a measure of a construct in the judgment of others, especially in the scientific community.

Content validity A type of measurement validity that requires that a measure represent all aspects of the conceptual definition of a construct.

03/04/2020

Dr. Syed Owais, Department of Sociology,
University of Peshawar

18

18

Types of Measurement Validity

Criterion validity Measurement validity that relies on some independent, outside verification.

Concurrent validity Measurement validity that relies on a preexisting and already accepted measure to verify the indicator of a construct.

Predictive validity Measurement validity that relies on the occurrence of a future event or behavior that is logically consistent to verify the indicator of a construct.

03/04/2020

Dr. Syed Owais, Department of Sociology,
University of Peshawar

19

19

Types of Measurement Validity

Construct validity A type of measurement validity that uses multiple indicators and has two subtypes: how well the indicators of one construct converge or how well the indicators of different constructs diverge.

Convergent validity A type of measurement validity for multiple indicators based on the idea that indicators of one construct will act alike or converge.

Discrete variables Variables in which the attributes can be measured with only a limited number of distinct, separate categories.

03/04/2020

Dr. Syed Owais, Department of Sociology,
University of Peshawar

20

20

4 Types of measurement validity

1. Face Validity:

- The easiest to achieve and the most basic kind of validity.
- 'It is a judgment by the scientific community that the indicator really measures the construct'. i.e., addresses the question:
 - *On the face of it, do people believe that the definition and method of measurement fit?*
(Neuman, 2014: 216)

03/04/2020

Dr. Syed Owais, Department of Sociology,
University of Peshawar

21

21

2. Content Validity

- Addresses the question:
 - *Is the full content of a definition represented in a measure?*
- First, specify fully the entire content in a construct's definition.
- Next, sample from all areas of the definition.
- Finally, develop an indicator that taps all of the parts of the definition.

03/04/2020

Dr. Syed Owais, Department of Sociology,
University of Peshawar

22

22

3. Criterion Validity

- When some **standard or criterion** is used to indicate a construct accurately
- The validity of an indicator is verified by comparing it with another measure of the same construct that is widely accepted.
- **3.1 Concurrent Validity:**
- To have this an indicator must be associated with a pre-existing indicator that is judged to be valid (i.e. it has face validity).
 - E.g., if a new test to measure intelligence is created, it'll be concurrently valid, if it's highly associated with existing IQ tests (assuming the same definition of intelligence is used). This means that most people who score high on the old measure also score high on the new one, and vice versa.
 - The two measures may not be perfectly associated, but if they measure the same or a similar construct, it is logical for them to yield similar results

03/04/2020

Dr. Syed Owais, Department of Sociology,
University of Peshawar

23

23

3.2. predictive validity

- “A criterion validity whereby an indicator predicts future events that are logically related to a construct” is called predictive validity”.
 - Not used for all measures.
- The measure and the action predicted must be distinct from but indicate the same construct.
- E.g., The Scholastic Assessment Test (SAT) that many U.S. High school students take measures scholastic aptitude – the ability of a student to perform in college.
 - If SAT has high predictive validity, then students who get high SAT scores will subsequently do well in college. But if no then, SAT has low predictive validity.

03/04/2020

Dr. Syed Owais, Department of Sociology,
University of Peshawar

24

24

Predictive validity ... cont'd...

- If a measure (especially a questionnaire) is measuring something in terms of “high and low”, e.g., political conservatism, it can be pilot-tested on Another way to test predictive validity is to select a group of people who have specific characteristics and predict how they will score (very high or very low) vis-a-vis the construct. I "validate" the measure with the groups-that is, I pilot-test it by using it on members of the groups. It can then be used as a measure of political conservatism for the general public.

03/04/2020

Dr. Syed Owais, Department of Sociology,
University of Peshawar

25

25

4. Construct validity

- It addresses the question:
 - *if the measure is valid, do the various indicators operate in a consistent manner?*
 - It requires a definition with clearly specified conceptual boundaries.
- 4.1: **Convergent validity**: means that multiple measures of the same construct hang together or operate in similar ways.
- For example, I measure the construct “education”

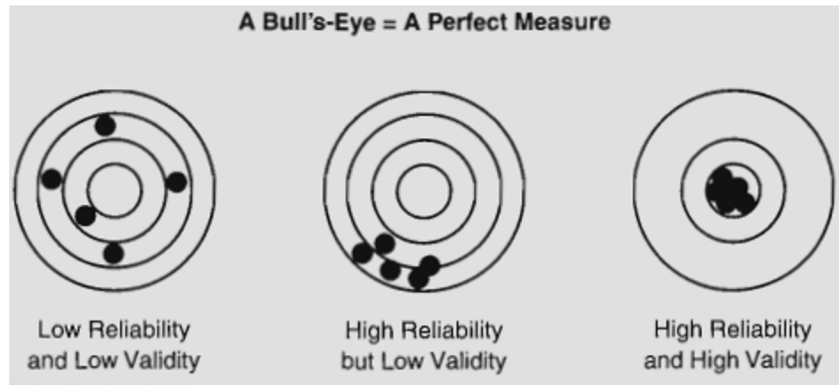
03/04/2020

Dr. Syed Owais, Department of Sociology,
University of Peshawar

26

26

Relationship b/w reliability and validity



03/04/2020

Dr. Syed Owais, Department of Sociology,
University of Peshawar

27

27

References

- Neuman, W. Lawrence (2013). *Social Research Methods: Qualitative and Quantitative Approaches*. Available from: <https://cleavermonkey.files.wordpress.com/2015/10/w-lawrence-neuman-social-research-methods-qualitative-and-quantitative-approaches-pearson-education-limited-2013.pdf>
- Bryman, Alan (4th Ed.) (2012). *Social Research Methods* (3rd ed.). Oxford: Oxford University Press. Available from: <http://en.bookfi.net/book/2266020>

03/04/2020

Dr. Syed Owais, Department of Sociology,
University of Peshawar

28

28