5. Reliability and Validity in a Research Study

Dr. Judith Gomes

Assistant Professor,
Bhilai Institute of Technology, Durg.

Mr. Shrawan Pandey

Assistant Professor,
Bhilai Institute of Technology, Durg.

Dr. Rachana Pandey

Assistant Professor,
S.S.S.S.M.V, Hudco, Bhilai.

5.1 Introduction:

In this chapter you will learn about:

- Reliability Vs Validity
- The Concept of Reliability
- Types of Reliability
- The concept of Validity
- Types of Validity
- How to ensure validity and reliability in your research
- Reliability and Validity in a thesis

Keywords: Reliability, validity, tests, instrument, measurement, research, consistency, construct validity, content validity, face validity.

5.2 Reliability Vs Validity:

In every research, it is important for us to attempt to establish the quality of results. The concepts of Reliability and validity are used to extract the research quality. Combinedly they evaluate the best way of measuring tests and techniques result. They provide the measurement consistency and accuracy, i.e; Reliability measure a consistency and Validity measures an accuracy. It is used especially in quantitative research to design research methodology and its analysis.

Reliability Vs Validity		
Basis	**Reliability**	**Validity**
Meaning	The extent at which reproduction of the results can be done in same conditions.	The extent at which the measurement of results can be done when what they are supposed to measure.

Reliability Vs Validity		
Basis	**Reliability**	**Validity**
Assess	By going through the result consistency within a time period, with different observers, and throughout the parts of the test itself.	By going through the wellness of the correspondence of results in comparison to theories and results of others concept.
Relatedness	Reliable measurement is not always valid.	But a valid measurement is generally reliable.

The concepts of reliability and validity:

Both Reliability and validity are related to each other, but they mean different things. A measurement can be reliable without being valid. However, if a measurement is valid, it is usually also reliable.

5.3 The Concept of Reliability:

We use the word 'reliable' very often in our lives. When we say that a person is reliable, what do we mean? We infer that s/he is dependable, consistent, predictable, stable and honest. The concept of reliability in relation to a research instrument has a similar meaning: if a research tool is consistent and stable, hence predictable and accurate, it is said to be reliable. The greater the degree of consistency and stability in an instrument, the greater its reliability. Therefore, 'a scale or test is reliable to the extent that repeat measurements made by it under constant conditions will give the same result' (Moser & Kalton 1989: 353).

Reliability refers to how consistently a method measure something. If the same result can be consistently achieved by using the same methods under the same circumstances, the measurement is considered reliable.

The concept of reliability can be looked at from two sides:

a. How reliable is an instrument?
b. How unreliable is it?

The first question focuses on the ability of an instrument to produce consistent measurements. When you collect the same set of information more than once using the same instrument and get the same or similar results under the same or similar conditions, an instrument is considered to be reliable.

The second question focuses on the degree of inconsistency in the measurements made by an instrument – that is, the extent of difference in the measurements when you collect the same set of information more than once, using the same instrument under the same or similar conditions. Hence, the degree of inconsistency in the different measurements is an indication of the extent of its inaccuracy. This 'error' is a reflection of an instrument's unreliability. Therefore, reliability is the degree of accuracy or precision in the measurements made by a research instrument. The lower the degree of 'error' in an instrument, the higher the reliability.

Reliability tells you how consistently a method measure something. When you apply the same method to the same sample under the same conditions, you should get the same results. If not, the method of measurement may be unreliable.

5.4 Types of Reliability:

Types of reliability can be estimated through various statistical methods. There are a number of ways of determining the reliability of an instrument and these can be classified as either external or internal consistency procedures.

Types of Reliability	
Test-retest	Measures the consistency of the same test over time.
Interrater	Measures the consistency of the same test conducted by different researchers
Parallel forms	Measures the consistency of two different tests that measures the same thing.
Internal consistency	Measures the consistency of the individual items by using a multi-item test.

	What does it assess?	**Example**
Test-retest	The consistency of a measure **across time**: do you get the same results when you repeat the measurement?	A group of participants complete a questionnaire designed to measure personality traits. If they repeat the questionnaire days, weeks or months apart and give the same answers, this indicates high test-retest reliability.
Interrater	The consistency of a measure **across raters or observers**: do you get the same results when different people conduct the same measurement?	Based on an assessment criteria checklist, five examiners submit substantially different results for the same student project. This indicates that the assessment checklist has low inter-rater reliability (for example, because the criteria are too subjective).
Internal consistency	The consistency of **the measurement itself**: do you get the same results from different parts of a test that are designed to measure the same thing?	You design a questionnaire to measure self-esteem. If you randomly split the results into two halves, there should be a <u>strong correlation</u> between the two sets of results. If the two results are very different, this indicates low internal consistency.

5.4.1 Test-Retest Reliability:

Test-retest reliability measures the consistency of results when you repeat the same test on the same sample at a different point in time. You use it when you are measuring something that you expect to stay constant in your sample.

Improving test-retest reliability

- When designing tests or questionnaires, try to formulate questions, statements and tasks in a way that won't be influenced by the mood or concentration of participants.
- When planning your methods of data collection, try to minimize the influence of external factors, and make sure all samples are tested under the same conditions.
- Remember that changes can be expected to occur in the participants over time and take these into account.

5.4.2 Interrater Reliability:

Interrater Reliability (also called interobserver reliability) measures the degree of agreement between different people observing or assessing the same thing. You use it when data is collected by researchers assigning ratings, scores or categories to one or more variables.

Improving interrater reliability

- Clearly define your variables and the methods that will be used to measure them.
- Develop detailed, objective criteria for how the variables will be rated, counted or categorized.
- If multiple researchers are involved, ensure that they all have exactly the same information and training.

5.4.3 Parallel Forms Reliability:

Parallel forms reliability measures the correlation between two equivalent versions of a test. You use it when you have two different assessment tools or sets of questions designed to measure the same thing.

Improving parallel forms reliability

- Ensure that all questions or test items are based on the same theory and formulated to measure the same thing.

5.4.4 Internal Consistency:

Internal consistency assesses the correlation between multiple items in a test that are intended to measure the same construct. Internal consistency can be calculated without repeating the test or involving other researchers, so it's a good way of assessing reliability when you only have one data set.

Improving internal consistency

- Take care when devising questions or measures: those intended to reflect the same concept should be based on the same theory and carefully formulated.

5.5 The Concept of Validity:

Validity refers to how accurately a method measures what it 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.

As inaccuracies can be introduced into a study at any stage, the concept of validity can be applied to the research process as a whole or to any of its steps: study design, sampling strategy, conclusions drawn, the statistical procedures applied or the measurement procedures used.

Broadly, there are two perspectives on validity:

a. Is the research investigation providing answers to the research questions for which it was undertaken?
b. If so, is it providing these answers using appropriate methods and procedures?

High reliability is one indicator that a measurement is valid. If a method is not reliable, it probably isn't valid.

In terms of measurement procedures, therefore, validity is the ability of an instrument to measure what it is designed to measure: 'Validity is defined as the degree to which the researcher has measured what he has set out to measure' (Smith 1991: 106).

According to Kerlinger, 'the commonest definition of validity is epitomised by the question: Are we measuring what we think we are measuring?' (1973: 457). Babbie writes, 'validity refers to the extent to which an empirical measure adequately reflects the real meaning of the concept under consideration' (1989: 133).

However, reliability on its own is not enough to ensure validity. Even if a test is reliable, it may not accurately reflect the real situation.

Validity is harder to assess than reliability, but it is even more important. To obtain useful results, the methods you use to collect your data must be valid: the research must be measuring what it claims to measure. This ensures that your discussion of the data and the conclusions you draw are also valid.

5.6 Types of Validity:

The validity of a measurement can be estimated based on three main types of evidence. Each type can be evaluated through expert judgement or statistical methods.

Types of Validity	
Construct	Measures the concept of existing theory and knowledge
Content	Measures the concept from all aspects
Face	Measures the concept of aim of the test
Criterion	Measures the concept from other valid measures

Type of validity	What does it assess?	Example
Construct	The adherence of a measure to **existing theory and knowledge** of the concept being measured.	A self-esteem questionnaire could be assessed by measuring other traits known or assumed to be related to the concept of self-esteem (such as social skills and optimism). Strong correlation between the scores for self-esteem and associated traits would indicate high construct validity.
Content	The extent to which the measurement **covers all aspects** of the concept being measured.	A test that aims to measure a class of students' level of Spanish contains reading, writing and speaking components, but no listening component. Experts agree that listening comprehension is an essential aspect of language ability, so the test lacks content validity for measuring the overall level of ability in Spanish.
Face	The extent to which the result of a measurement has logical link between items and objectives.	A study is conducted to identify health injuries due to smoking. If the results predict that smoking is injurious to health, this will indicate the obvious link between the conditions.
Criterion	The extent to which the result of a measure corresponds to **other valid measures** of the same concept.	A survey is conducted to measure the political opinions of voters in a region. If the results accurately predict the later outcome of an election in that region, this indicates that the survey has high criterion validity.

5.6.1. Construct Validity:

Construct validity evaluates whether a measurement tool really represents the thing we are interested in measuring. It's central to establishing the overall validity of a method.

A construct refers to a concept or characteristic that can't be directly observed but can be measured by observing other indicators that are associated with it. It is a more sophisticated technique for establishing the validity of an instrument.

It is based upon statistical procedures. It is determined by ascertaining the contribution of each construct to the total variance observed in a phenomenon.

Constructs can be characteristics of individuals, such as intelligence, obesity, job satisfaction, or depression; they can also be broader concepts applied to organizations or social groups, such as gender equality, corporate social responsibility, or freedom of speech. One of the main disadvantages of construct validity is that you need to know about the required statistical procedures. The other types of validity described below can all be considered as forms of evidence for construct validity.

5.6.2 Content Validity:

Content validity assesses whether a test is representative of all aspects of the construct. To produce valid results, the content of a test, survey or measurement method must cover all relevant parts of the subject it aims to measure. If some aspects are missing from the measurement (or if irrelevant aspects are included), the validity is threatened.

The judgement that an instrument is measuring what it is supposed to is primarily based upon the logical link between the questions and the objectives of the study. Hence, one of the main advantages of this type of validity is that it is easy to apply. Each question or item on the research instrument must have a logical link with an objective. It is important that the items and questions cover the full range of the issue or attitude being measured. Assessment of the items of an instrument in this respect is called content validity. In addition, the coverage of the issue or attitude should be balanced; that is, each aspect should have similar and adequate representation in the questions or items. Content validity is also judged on the basis of the extent to which statements or questions represent the issue they are supposed to measure, as judged by you as a researcher, your readership and experts in the field.

5.6.3 Face Validity:

Face validity considers how suitable the content of a test seems to be on the surface. It's similar to content validity, but face validity is a more informal and subjective assessment. As face validity is a subjective measure, it's often considered the weakest form of validity. However, it can be useful in the initial stages of developing a method. The judgement that an instrument is measuring what it is supposed to is primarily based upon the logical link between the questions and the objectives of the study. Hence, one of the main advantages of this type of validity is that it is easy to apply. Each question or item on the research instrument must have a logical link with an objective. Establishment of this link is called face validity.

5.6.4 Criterion Validity:

Criterion validity evaluates how closely the results of your test correspond to the results of a different test.

The criterion is an external measurement of the same thing. It is usually an established or widely-used test that is already considered valid. To evaluate criterion validity, you calculate the correlation between the results of your measurement and the results of the criterion measurement. If there is a high correlation, this gives a good indication that your test is measuring what it intends to measure. 'In situations where a scale is developed as an indicator of some observable criterion, the scale's validity can be investigated by seeing how good an indicator it is' (Moser & Kalton 1989: 356).

How to ensure validity and reliability in your research

The reliability and validity of your results depends on creating a strong research design, choosing appropriate methods and samples, and conducting the research carefully and consistently.

Ensuring validity

If you use scores or ratings to measure variations in something (such as psychological traits, levels of ability or physical properties), it's important that your results reflect the real variations as accurately as possible. Validity should be considered in the very earliest stages of your research, when you decide how you will collect your data.

- Choose appropriate methods of measurement.

Ensure that your method and measurement technique are high quality and targeted to measure exactly what you want to know. They should be thoroughly researched and based on existing knowledge. For example, to collect data on a personality trait, you could use a standardized questionnaire that is considered reliable and valid. If you develop your own questionnaire, it should be based on established theory or findings of previous studies, and the questions should be carefully and precisely worded.

- Use appropriate sampling methods to select your subjects.

To produce valid generalizable results, clearly define the population you are researching (e.g., people from a specific age range, geographical location, or profession). Ensure that you have enough participants and that they are representative of the population.

Ensuring Reliability:

Reliability should be considered throughout the data collection process. When you use a tool or technique to collect data, it's important that the results are precise, stable and reproducible.

- Apply your methods consistently.

Plan your method carefully to make sure you carry out the same steps in the same way for each measurement. This is especially important if multiple researchers are involved.

For example, if you are conducting interviews or observations, clearly define how specific behaviours or responses will be counted, and make sure questions are phrased the same way each time.

- Standardize the conditions of your research.

When you collect your data, keep the circumstances as consistent as possible to reduce the influence of external factors that might create variation in the results.

For example, in an experimental setup, make sure all participants are given the same information and tested under the same conditions.

Where to write about reliability and validity in a thesis It's appropriate to discuss reliability and validity in various sections of your thesis or dissertation. Showing that you have taken them into account in planning your research and interpreting the results makes your work more credible and trustworthy.

Reliability and validity in a thesis	
Section	**Discuss**
Literature review	What have other researchers done to devise and improve methods that are reliable and valid?
Methodology	How did you plan your research to ensure reliability and validity of the measures used? This includes the chosen sample set and size, sample preparation, external conditions and measuring techniques.
Results	If you calculate reliability and validity, state these values alongside your main results.
Discussion	This is the moment to talk about how reliable and valid your results actually were. Were they consistent, and did they reflect true values? If not, why not?
Conclusion	If reliability and validity were a big problem for your findings, it might be helpful to mention this here.

5.7 References:

1. Ranjit Kumar, "Research Methodology: A Step-by-step guide for beginners", Sage Publications, 3rd Edition,
2. https://www.scribbr.com/methodology/types-of-validity/
3. https://www.scribbr.com/methodology/reliability-vs-validity/
4. https://www.scribbr.com/methodology/types-of-reliability/
5. https://opentextbc.ca/researchmethods/chapter/reliability-and-validity-of-measurement/
6. https://www.simplypsychology.org/validity.html