Psychology Factsheets

www.curriculum-press.co.uk Number 85

Reliability and validity

This Factsheet summarises what reliability and validity are, types of reliability and validity, and how they can be assessed/ensured. Terms in bold are explained in the glossary.

A. What are reliability and validity?

Reliability and validity are aspects of investigation design. They are ways of assessing the quality of research studies. Good quality research has good reliability and good validity.

Psychologists ask: "Are the measures both reliable and valid?"

- Reliability: The reliability of any method of measurement (this can be an interview, questionnaire, experimental test, etc) refers to how *consistently* it measures something. A reliable test gives similar results in similar circumstances. For example, if your kitchen scales gave you different readings each time you weighed out the same amount of sugar for your cake, they would not be reliable (neither would your cakes!).

Is it reliable = Is it repeatable?

- **Validity:** The validity of any method of measurement refers to how *truly/realistically* it measures something. A valid test measures what it is supposed to measure - it does not measure something else! For example, a valid **IQ** test should really be measuring intelligence and not just measuring general knowledge.

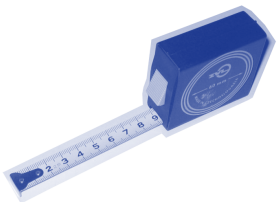
Is it valid? = Does it actually measure what it intends to measure?

Exam Hint:- *Reliability and validity are not the same things! You must know what they both are. Make sure that you can explain what they are in writing.*

~ Jake is a participant in an experimental test. He sees a series of words on a computer screen, one at a time. He is asked to decide if each word is a real word or not (e.g., SWOAD is not a real word, SWORD is a real word). His reaction times are recorded.

~ If this test is reliable, it will measure Jake's performance consistently each time he takes the test.

Remember, a test can be reliable without being valid. For example, I bought a tape measure from my local pound shop. It is not very accurate as the centimetre spacing is actually measuring 11mm, not 10mm as it should. This means that it is not a true (valid) measure of length. However, it does always measure 11mm for a centimetre instead of 10mm every time I use it. This means that it is reliable. So my tape measure is reliable but not valid!



Exam Hint:- *Make sure that you know that reliability and validity have nothing to do with the ethics of a research study.*

Exam Hint:- *There are various types of reliability and validity. Make sure that you understand them and can explain what they are.*

B. Types of reliability

1. Internal reliability

How consistently a method/test measures within itself.
For example, if my 30cm ruler had different size centimetres all the way along its length, it would not be standardised and would not have good internal reliability. Also, internal reliability would be bad if two people simultaneously observing the same participant's behaviour scored it differently.

2. External reliability

How consistently a method/test measures *over time* when repeated. A method/test with good external reliability will give similar scores for the same people under similar conditions.

For example, if my ruler found different lengths each time it measured the same object, it would not have good external reliability.

Exam Hint:- *Make sure that you can describe what internal and external reliability both are and know the difference between them.*

Exam Hint:- *Psychologists assess the reliability of their test/method in several ways. You need to know what these are.*

C. Ways of assessing/ensuring reliability

There are several ways of assessing and ensuring reliability. These involve doing **correlations**. A test/method is reliable if there is a strong positive correlation.

1. The split half method

This assesses internal reliability. Half of the *items* in the method/ test are correlated with the other half of the items. If there is a strong positive correlation, the test/method has good internal reliability.

2. The test-retest method

This assesses external reliability. The test is given to the same participants on two different occasions. The *results* of the test on one occasion are correlated with the results obtained on the later occasion. If there is a strong positive correlation, the test has good external reliability.

3. Inter-rater reliability

This is also called inter-observer or inter-judge reliability. This assesses how much the rating/scoring of participants' behaviour by one observer is consistent with the rating/scoring by another observer. Two observers rate/score participants' behaviour separately but at the same time. Their results are compared. If there is a strong positive correlation, there is good reliability.

D Types of validity

1. Internal validity

How much the findings of a test/method are due to the manipulation of a variable rather than another factor. It is also called experimental validity.

For example, a method/test has good internal validity if the findings do result from the manipulation of the independent variable on the dependent variable.

2. External validity

How much the findings of a test/method can be generalised to different settings. This includes how much the findings can be generalised to other environments (ecological validity) and people (population validity).

E. Ways of assessing/ensuring validity

There are various ways of assessing/ensuring validity. The table below shows some of these ways.

Type of validity	*Does the test/method...*
Face	...actually *look* like it is measuring what it is supposed to be measuring?
Construct	... test the *theory* about the variable that it is supposed to be measuring?
Concurrent	...have a strong positive correlation with an already well-established one that claims to measure the same variable?
Predictive	... predict a person's future performance on a test/method as indicated by its results?

F. The reliability and validity of research methods

Exam Hint:- *You should know that different research methods had different levels of reliability and validity. For example:*

- Laboratory experiments are done in an artificial environment. This may cause the participants to show unnatural behaviour. So, laboratory experiments can lack ecological validity.
- Laboratory experiments are reliable as they tend to have a high level of control and can be replicated producing similar findings.
- As field experiments are done in the participants' own environment they have greater ecological validity than laboratory experiments.
- Participant observations (where the researcher becomes involved in the participants' lives) have higher ecological validity than controlled observations (which are done under conditions set by researcher, such as in a laboratory).
- Case studies can have low reliability due to lack of observer objectivity.

Exam Hint:- *If you are asked to explain what is meant by the term validity/reliability, do not just name a type of validity/ reliability. You do need to describe validity/reliability in enough detail as well as accurately.*

G. The reliability and validity of research studies

Exam Hint: *- Can you comment on the reliability and/or validity of research studies?*

For example:
In 1986, Sue Savage-Rumbaugh and colleagues conducted a case study, observing how chimps could learn language. The chimps communicated with humans using a keyboard of symbols connected to a speech synthesizer. One particular chimp, Kanzi, was unusual as he spontaneously started to use symbols when separated from his mother who used them. The researchers kept a record of Kanzi's utterances for 17 months using (i) an automatic computerised record from the keyboard, (ii) notes from observers and (iii) 4½ hours of videotape. The utterances were categorised (e.g., correct/ incorrect, spontaneous/imitated). They found that Kanzi produced 2540 correct combinations of symbols (e.g. 'chase Kanzi'), most of them being spontaneous.

What might affect the reliability and validity of this study?

- The chimps learnt and showed language use in a natural, outdoors environment, which affects its ecological validity.
- Was the coding of observers reliable? The reliability of the coding (categorising of utterances) was checked independently by two different observers. They scored a sample of the utterances. One observer scored the data in real-time while the other observer scored the videotape. There was 100% agreement.
- There will be other factors affecting the reliability and validity (can you think of any?).

Glossary

Correlation: A form of analysing data which measures the relationship between variables. A positive correlation is where one variable increases as another variable increases. A negative correlation is where one variable decreases as another variable increases.
IQ (intelligence quotient): A measure of intelligence calculated originally by dividing mental age by age in years and multiplying by 100.

Worksheet: Reliability and Validity

Name _____

1. Explain what is meant by the term validity.

---

2. Explain what is meant by the term reliability.

---

3. Draw arrows between the boxes to match the types of validity/reliability to the descriptions.

Type	→	***Description***
Internal validity		How consistently a method/test measures over time when repeated.
External validity		How much the findings of a test/method are due to experimental manipulation of a variable rather than some other factor.
Internal reliability		How consistently a method/test measures within itself.
External reliability		How much the findings of a test/method can be generalised to other settings and groups of people.

4. Describe one way of assessing reliability.

---

5. Describe one way of assessing validity.

Example Exam Question

(a) Explain what is meant by the term validity.

(b) Explain why studies of eyewitness testimony have been said to lack validity.

(a) Make sure that you explain validity and not reliability. Do not just give the name of one type of validity. For example, you might say that validity means whether something measures what it claims to measure. A valid test measures what it is supposed to measure, not something else. If it is not valid, it does not give legitimate findings.

(b) Show that you understand what validity is and that you know about this type of study. For example, you could write about such studies being conducted in a laboratory setting, participants not being emotionally aroused in the same way they would when witnessing a real incident and the tasks not being realistic as participants saw a video, not a real incident.

Acknowledgements: *This Psychology Factsheet was researched and written by* ***Amanda Albon.***