ILYAS BASHIR ID-16972 (MS-HRM)
Submitted to: Dr Farooq Jan
Assignment: Reliability and validity types with explanation.

Explain Reliability with types

Reliability tells us how consistently a method measures 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.

Four main types of reliability

1 Test-retest, the same test for many times
2 Inter-rater. Same test by different peoples.
3 Parallel forms. different versions equivalently designed
4 Internal Consistency. The individual item of test.

Now explain these four types one by one as under:

1. **Test retest reliability:**

It measures the consistency of results when you repeat the same test on the same sample at a different point in time. Using it when you

are measuring something that you expect to stay constant in sample.

As because color blindness is a trait that does not change over time

Importance:

Many factors can influene your results at different points in time for example respondents might experience different moods, or external conditions might effect their ability to respond accurately.

Measuring it:

You conduct the same test on the same group of people at two different points in time. Then you calculate the correlation between the two sets of results.

Example:

A questionare to measure the IQ of a participants group. You administer the test two months apart to the same group of people, but the results are significantly different, so the test retest reliability of the IQ questionare is low.

Improving it:

when designing tests, try to formulate questions etc in a way that would not be influenced by the mood or concentration of participants, when planning your methods of data collection try to minimize external influence and samples

may be tested under the same conditions participants changes can be expected to occur over time and these can be noted.

2 Interrater liability :

It measures the degree of aggrement between different peoples observing the same thing. You use it when data is collected by researchers assigning ratings, score to one or more variables.

It is an observational study where a team of researchers collect data on classroom behaviours.

Importance :

Peoples are subjective, so different observers and phenomena naturally differs. Reliable research aims to minimize subjectively as much as possible so that a different researcher could replicate the same result. When designing criteria for data collection, make sure that different people will rate the same variable consistently.

Measuring it :

Different researchers conduct the same observation on the same sample. Then you calculate the correlation between their different sets of results. If all the researchers give similar ratings, the test has high Interrater reliability.

Example:

A team of researchers observe the progress of wounds healing in patients. To record the stages of healing, rating scales are used, with a set of criteria to assess various aspects of wounds. The results of different researchers assessing the same set of patients are compared, and there is a strong correlation between all sets of result so the test has high interrater liability.

Improving:

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

3 Parallel forms reliability:

It 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.

Importance:

If you want to use

multiple different versions of test, you first need to make sure that all the sets of questionare measurment give reliable results.

In educational assessements it is often necessary to create different versions of tests to ensure that students donot have acess to the questions in advance.

Parallel forms reliability means that, if the same students take two different version of a reading comprehension test, they should get similar results in both tests.

How to measure parallel reliability:

The most common way to measure parallel form reliability is to produce a large set of questions to evaluate the same thing then randomly into two set of questions. The same group of respondent answer both sets, and you calculate the correlation between the results. High correlation between the two indicates high parallel forms reliability.

Example:

A set of questions is formulated to measure financial risk in a group of respondents. The questions are randomly divided into two sets and the respondents are randomly divided

into two groups both group takes both tests improving.

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

4 Internal Consistency:

The correlation between multiple items in a test that are intended to measure the same construct. You can calculate internal consistency without repeating the tests or involving other researchers, so it is a good way of assessing reliability when you only have one data set.

Importance:

When a set of questions or ratings that will be combined into an overall score you have to make sure that all of the items really do reflect the same thing. If responses to different items contradict one another the test might be unreliable.

Example:

A group of respondents are presented with a set of statements designed to measure optimistic and pessimistic mind-sets. They must rate their agreement with each statement on a scale from 1 to 5. If the test is internally consistent.

an optimistic respondents should generally give high ratings to optimism indicators and low ratings to pessimism indicators.

The correlation is calculated between all the responses to the optimistic statements, but the correlation is very weak. This suggests that the test has low internal consistency.

Improving :

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

Explain vaillibility with types.:

Validity tells us how accurately a method measures something. If a method measures what it claims to measure, and the results closely corresponds to real-world values, then it can be ~~construcetd~~ considered valid, there are four main types of validibility.

1 Constouct validity.

Does the test measures the concept that it is intended to measure.

2 Content validity:

It is the test fully

representative of what it aims to measure?

3 Face validity:

Does the content of the test appear to be suitable to its aims?

4 Criterion validity.

Does the results corresponds to a different test of the same thing.

Now we will discuss it in detail.

1 Construct validity:

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

Construct?

A construct refers to a concept that can not be directly observed, but can be measured by observing other indicators that are associated with it. For example: There is no objective, observable entity called depression that we can measure directly. But based on existing psychological research and theory, we can measure depression based on a collection of symptoms and indicators, such as low self-confidence and low energy levels.

Construct validity ?

It is about ensuring that the method of measurement matches the construct you want to measure. If you develop a questionare to diagnose depression you need to know, does the questionare really measures the construct of depression? or it is actually measuring the respondent mood, self-esteem or some other construct?

To achieve construct validity you have to ensure that your indicators and measurements are carefully developed based on relevant existing knowledge. The questionare must include only relevant questions that measures known indicators of depressions.

Other types of validity described below can all be considered as forms of evidence for construct validity.

2 Content validity :

This 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 aim to measure. If some aspects are missing from the measurement, the validity

is threatened.

For example.

A mathematics teacher develops an end-of-semester algebra test for class. The test should cover every form of algebra that was taught in the class. If some types of algebra are left out, then the results may not be an accurate indication of students understanding of the subjects. Similarly if she includes questions that are not related to algebra, the results are no longer a valid measure of algebra knowledge.

3 Face validity:

Face validity considers how suitable the content of a test seems to be on the surface. It is similar to content validity, but face validity is a more informal and subjective assessment.

Example:

You create a survey to measure the regularity of people's dietary habits. You review the survey items, which ask questions about every meal of the day and snacks eaten in between for every day of week.

On its surface, the survey seems like a good representation of what you want to test, so you consider it to have high face validity. As face validity is subjective measure, its often considered the weakest form of validity. However, it can be useful in the initial stages of developing a method.

4 Criterion validity :

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

what is a criterion ?

The criterion is an external measurment of the same thing, it is usually an established or widely-used test that is already considered valid.

what is criterion validity :

To evaluate criterion validity, you calculate the correlation between the results of your measurment and the results of the criterion measurment. If there is a high correlation, this gives a good

indication that your test is measuring what it intends to measure.

Example:

A university professor create a new test to measure applicants English writing ability. To asses how well the test really does measure students writing ability, she finds an existing test that is considered a valid measurement of English writing ability, and compares the results when the same group of students take both tests. If the outcomes are very similar, the new test has a high criterion validity.