Instrument Validity & Reliability

Prof. Ashraf Zaghloul

Prof. Ashraf Zaghloul

RELIABILITY AND VALIDITY OF MEASURES

When we engage in research, we trust that the **measures** we use to define a variable **actually capture the meaning of the concept** and that they do so **consistently** باستمرار

Prof. Ashraf Zaghloul

Validity صلاحية

Validity is the most critical criterion and *indicates the degree* to which an instrument measures what it is supposed to measure

EXAMPLE

If we are interested in **job satisfaction**, *does our questionnaire or survey of job satisfaction actually measure the extent to which workers are content with their work*, or do the questionnaire items really reflect other concepts like "acceptance of work responsibilities"?

How satisfied are you with our services?

- ☐ Very Satisfied
- ☐ Satisfied
- ☐ Neutral
- ☐ Dissatisfied
- ☐ Very Dissatisfied
- ☐ No Opinion

Prof. Ashraf Zaghloul

The **4** sub-forms of validity

(1) Face validity

(2) Content validity

(3) Predictive/ concurrent criterion-related validity

(4) Construct validity

Prof. Ashraf Zaghloul

Types of Validity

Face validity: صحة الوجه

The extent to which a measure has a ring of plausibility about it.

In other words, does the measure appear to measure what it is intended to measure?

Prof. Ashraf Zaghloul

Face validity is determined by the degree to which a **participant or test taker** looks at an assessment and says to himself/herself.

"This seems to clearly be an assessment of X," when X is the construct you are hoping to measure.

Face validity is an important consideration during the test development process because when people do not believe that a measurement procedure is face valid, they may not respond honestly or fully

Content validity: صحة المحتوى

Does the measure **adequately represent all the content domains of the concept it is intended to measure**?

(Does a measure of academic achievement contain measures of math, reading, writing, science, etc., or merely reading?)

EXAMPLE

If you want to determine a person's proficiency in mathematics, you will ask the individual to solve math problems

In most cases, we determine the content validity of a test by having EXPERTS on the subject review the test.

Therefore, you might ask several professors of mathematics whether a math test adequately assesses the student's ability to understand and apply specific mathematical concepts

It is very easy to confuse the two.

The major difference is that **face validity** deals with the layperson's or test taker's perceptions, while **content validity** is evaluated based on the opinions of experts

Prof. Ashraf Zaghloul

Content Validity Index

1- experts rate the item or statements in terms of relevant =1 / irrelevant = 0

2- item validity = add rates and divided by number of raters = e.g; 0.7, 0.5

3- add instrument proportions by items and divide by number of items

4- CVI = good 0.8 and above

Prof. Ashraf Zaghloul

Kappa Statistic
(Cohen's kappa) K ϰ

Interrater Test of Significance

The kappa statistic is frequently used to test interrater reliability

Measurement of the extent to which data collectors (raters) assign the same score to the same variable is called interrater reliability

Like most correlation statistics, the kappa can range from −1 to +1

One of the important components of any good measure is that it allows us to accurately predict other behaviors or outcomes

This type of validity is referred to as **criterion-related validity** because we are trying to predict something (i.e., the criterion).

Prof. Ashraf Zaghloul

In some cases, we hope that measurements that we take now will allow us to predict events later.

If a test allows us to make these predictions, then we say that the test has **predictive validity**

Prof. Ashraf Zaghloul

Criterion-related validity

صحة معيار ذات الصلة

Suppose you develop an instrument to determine the suitability of applicants for a profession.

The instrument's validity might be determined by comparing it with another assessment, for example by a psychologist, or with a future observation of how well these applicants have done in the job.

If both assessments are similar, the instrument used to make the assessment at the time of selection is assumed to have higher validity.

Predictive validity
صحة التنبؤية

Refers to the **relationship** between measures that are taken at **different** times

Example

A test may be used to predict another context in the future (future score, future performance)

Concurrent validity متوافق

Refers to the **relationship** between measures taken at the same time

Example, a therapist may use two separate depression scales with a patient to confirm a diagnosis.

EXAMPLE

Assume that a researcher is using an instrument or has developed another measurement strategy to capture the construct of **Depression**

The measure would have **concurrent criterion validity** if the measure indicated depression and the participant met diagnostic criteria for depression at the same time

Prof. Ashraf Zaghloul

The measure would have **predictive criterion validity** if the measure indicated depression and the participant met diagnostic criteria for depression at some point in time in the future

Construct validity

Refers to the extent to which the test or measurement **strategy** measures a theoretical construct or trait

Prof. Ashraf Zaghloul

Construct validity

Although there are numerous approaches for determining construct validity, we will focus on the two most common methods:

Convergent validity (strong direct correlation with another valid instrument)

and

Divergent validity (strong indirect correlation with another valid instrument)

Prof. Ashraf Zaghloul

Construct statistical tests of significance

Correlation coefficients

Factor analysis (principal components analysis, varimax rotation, oblique rotation ….etc;)

Prof. Ashraf Zaghloul

Thank You