

Validity of measurement instruments used in PT research

*Mohammed TA, Omar Ph.D. PT, PGDCR-CLT
Rehabilitation Health Science Department*

Momarar@ksu.edu.sa

A Word on Embedded Assessment

Discusses ways of establishing if an outcome measure (OM) is valid.

Described various types of validity and how they can be reported and their relevance to clinical practice.

What is validity?

- Valid=faithful= true
- **Validity** is the degree to which the instrument measures what is intend to measures.

Examples

- Bathroom scale to measuring weight and height
- Goniometer that measures joint position in degree (ROM)
- Thermometer that measure temperature



VALIDITY

Establishing validity involves answers to the following :

- Is there a **consensus (agreement)** that the scale measures what it is supposed to measure?
- Does the measure correlate with other measures of the same concept?
- Does the behavior expected from the measure predict actual observed behavior?



VALIDITY

Important Things to
Remember About Validity

Validity refers to the decisions we make, and not to the test itself or to the measurement

Not all or none phenomena Never totally absent or absolutely perfect. **Ranges from 0-1.**

Validity can never be finally determined It is specific to each administration of the test

VALIDITY

Evaluation of validity
may results in new
OM insurgents

1-Similar in length to the original but with different scoring or instructions

2-Shortened for ease of use.

3-Validated for use with a different group or a different method

The process of validation is repetitive and responsive to new calls placed on the OM in clinical practice.

Relationship Between Reliability & Validity

Reliability is a necessary, but not sufficient, for validity,
Validity is necessary but not sufficient for generalizability



Reliable
Not valid



Valid
Not reliable



No valid
Not reliable



Valid
reliable

Types of Validity

```
graph TD; A[Types of Validity] --> B[Translation]; A --> C[Construct]; A --> D[Criterion]; B --> E[Face]; E --> F[Content]; C --> G[Convergent]; C --> H[Discriminate]; D --> I[Predictive]; I --> J[Concurrent];
```

Translation

Face

Content

Construct

Convergent

Discriminate

Criterion

Predictive

Concurrent



Face and content validity

Face validity is a subjective assessment of the degree to which an OMs instruments appears to measure what is designed to measure

- ❖ **Simplest & most subjective form of validity**
- ❖ **Least scientific definition of validity**
- ❖ **Based on subjective judgment and difficult to quantify.**

In face validity, you look at the operationalization and see whether “on its face” seems like a good translation of the constructed validity.

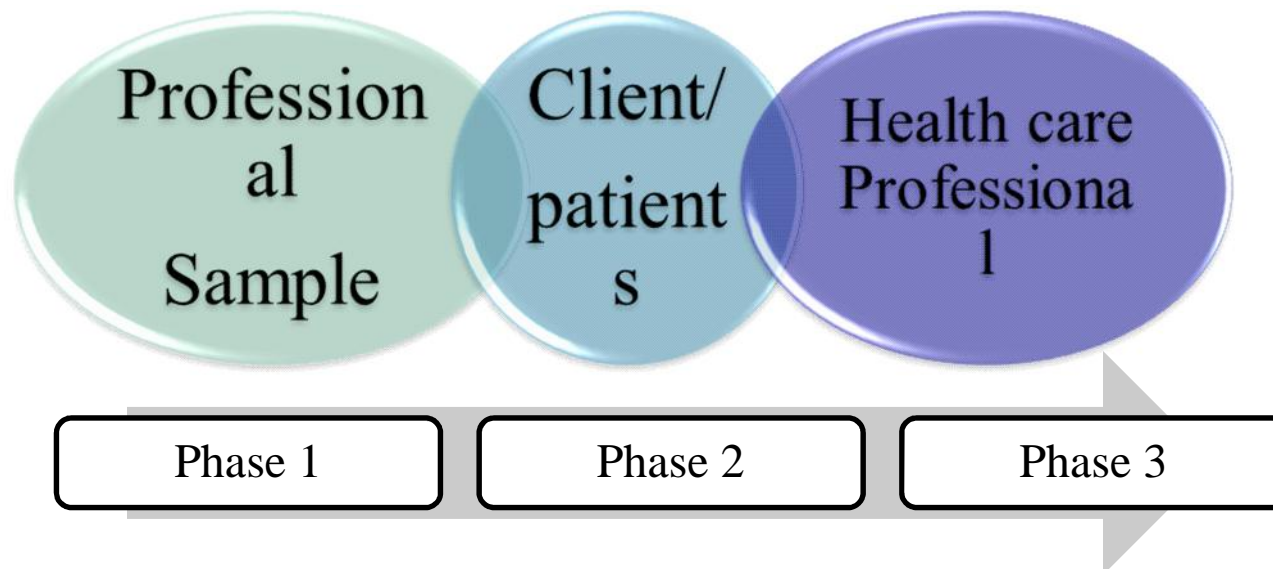
For example

- Hydrostatic weighing
- Lower extremity function scale
- **Dose this instrument appear to be appropriate choice to measure the variables ?**

Face and content validity

Content validity is the extent that, the test items actually represent all the aspect of the domain/construct of interest that is being measured.

- **Asher (1996)** notes that content validity is
 - **Descriptive rather than statistically determined**
 - **Considered a weaker form of validity compared to other types of validity, but deeper than face validity**



Criterion validity

- ❖ The degree of correlation of a measure with other reference standard measures of the same domain/construct.
- ❖ It is established through comparing a new or untested measurement tool against an accepted measurement technique, often known as a “gold standard”.
- ❖ Scientific evidence supports the accuracy and validity of the gold standard method.
- ❖ If the new measure agrees with the gold standard measure, the new measure by association must be a valid technique.

Criterion validity

Concurrent validity

Compares the measure's results to the "gold standard's results that is obtained at approximately the same point in time.

MMT & dynamometer
Exercise tolerance & VO_2max

Criterion validity

Predictive Validity

The ability of a OMs to predict scores on a criterion measured in the future.

Berg Balance Test to predicate falls over the following 6 weeks

Criterion validity

❖ Criterion-related and predictive validity can be expressed in statistical terms known as “Correlation Coefficients”.

1-Pearson-product moment correlation (PPMCC)

2-Spearman's rank order correlation

3-Kendall's rank order correlation

4-Phi coefficient ().

Criterion validity

If a good criterion measure exists, why create another instrument?

The new instrument is less expensive, less invasive or carries less risk.

In rehabilitation

- The consideration of criterion validity becomes less absolute when a gold standard does not exist.
- Criterion validity may still be assessed by substituting other well known and OMs as a substitute for the gold standard. accepted

Criterion validity

Table 5.1 Examples of concurrent and predictive validity

Author	New scale	Validity	Compared with a cognate scale
Smith (1994)	Elderly Mobility Scale (EMS)	Concurrent	Barthel Index (BI) Functional Independence Measure (FIM)
Prosser & Canby (1997)			Barthel Index
Collen et al (1991)	Rivermead Mobility Index (RMI)	Concurrent	BI, Functional Ambulation Categories (FAC), gait speed, 6-min distance test (SMDT)
Benaim et al (1999)	Postural Assessment Scale for Stroke (PASS)	Concurrent and predictive	FIM-transfers, locomotion, total concurrent measure and at 90 days after stroke
Mao et al (2002)			BI, Berg Balance Scale (BBS)

Construct validity

A construct is an idea or concept, and in the context of measurement properties, construct validity is a process, ‘part science and largely art form’ which draws the developer and later the reviewer along an evolving set of hypotheses about the new instrument.

Several methods to test's construct validity:

1. known-groups method assesses
2. Hypothesis/theory testing
3. Factor analysis,

Construct Validity

```
graph TD; A[Construct Validity] --> B[Convergent Validity]; A --> C[Discriminate validity];
```

Convergent Validity

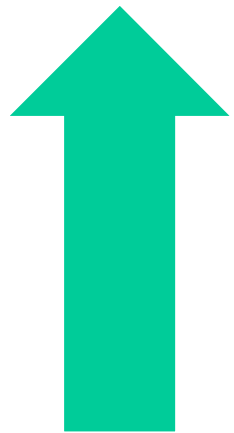
- This form of validity examines the extent to which a measure's results agree with the results of another measure that is believed to be assessing the same domain.
- If the comparison measure is "gold standard" known as criterion validity, if "not gold standard" convergent validity.
- Measures inter-correlated

Discriminate validity

A method of construct validity that reflects the degree to which an instrument can distinguish between or among different concepts or constructs.

Construct validity

- **Factor analysis: content and construct validity**



In the content validity:, the items within the instrument, or sub-scales within a multidimensional inventory are examined to identify how they fit into one or more themes.

In the construct validity indicating the associations between scales measuring similar constructs and lack of associations with scales measuring different concepts