Validity of Measurement Instruments Used in PT Research

Mohammed TA, Omar Ph.D. PT, PGDCR-CLT
Rehabilitation Health Science Department
Momarar@ksu.edu.sa

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

Lecture Outline:

- Definition of Terms

- Types of Validity

  - **Face validity**

  - **Content validity**

  - **Criterion-related validity**

    - **Predictive validity**

    - **Concurrent validity**

  - **Construct validity**

    - Convergent validity

    - Discriminant validity

What is Validity?

Valid=faithful= true

Validity is the degree to which the instrument measures what is intend to measures *Or* The soundness or appropriateness of a test or instrument in measuring what it is designed to measure"

- Bathroom scale
- Goniometer)
- Thermometer
- Hand held dynamometer

VALIDITY

Is there a consensus (agreement) that the scale measures what it is supposed to measure?

– Does the behavior expected from the measure predict actual observed behavior?

To Establishing **validity**

– Does the measure correlate with other measures of the same concept?

– Not all or none phenomena

– Validity refers to the decisions we make, based on instrument scores and not to the test itself or to the measurement

Relationship Between Reliability & Validity

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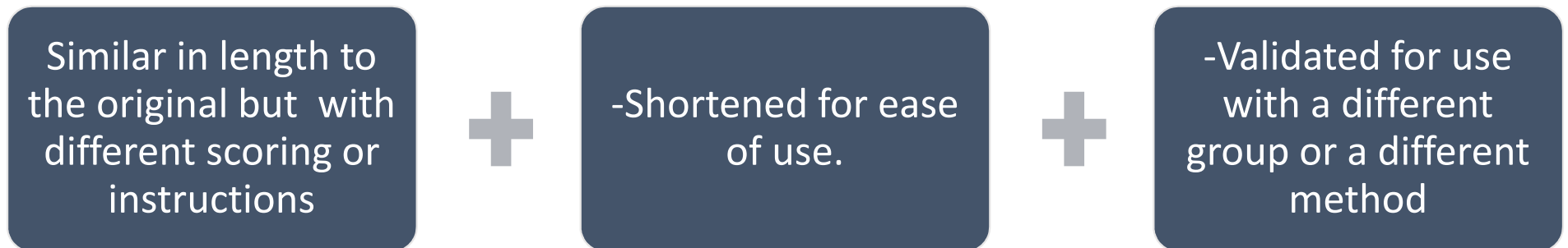

> Reliable
> Not valid

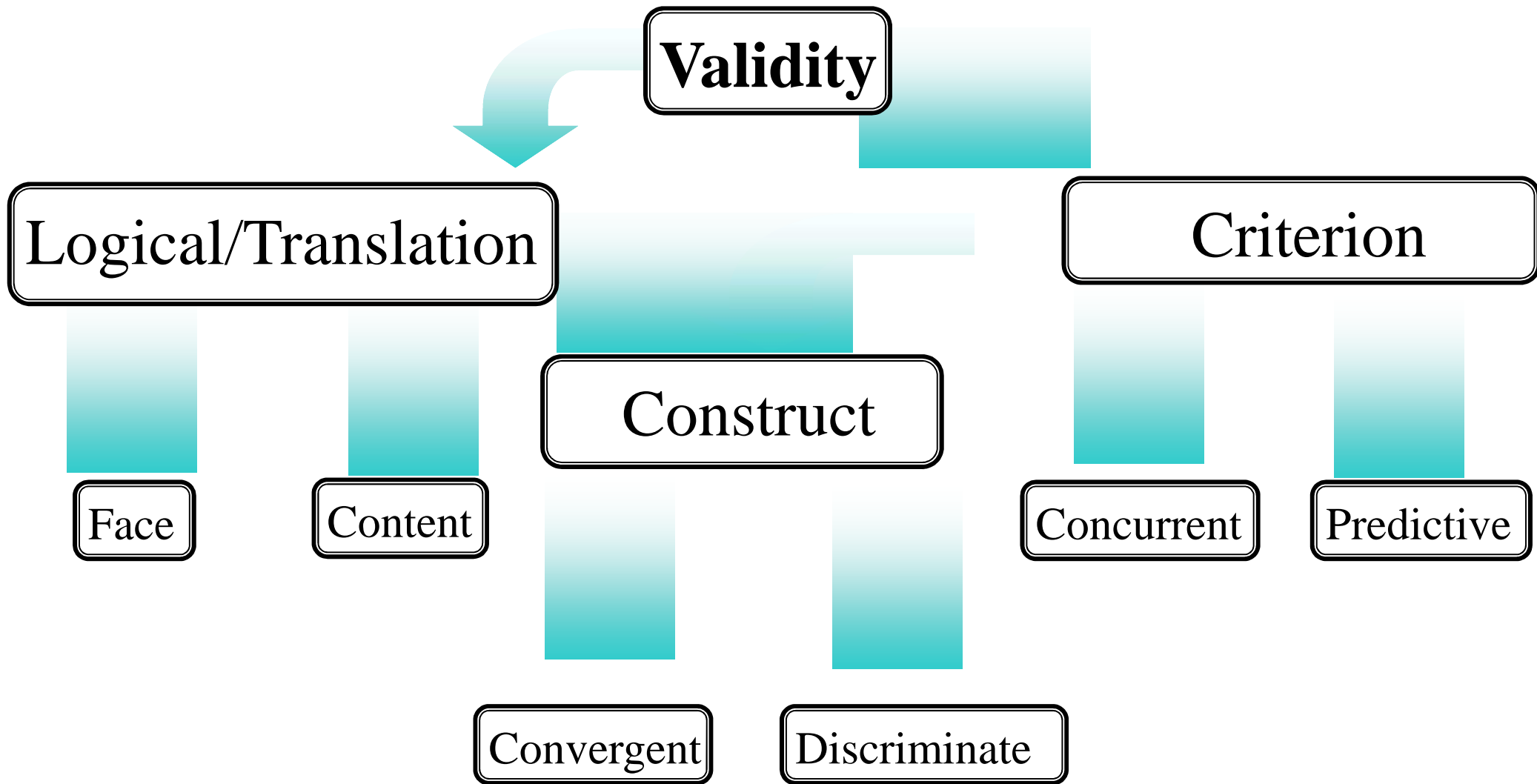> Valid
> Not reliable

> No valid
> Not reliable

> Valid
> reliable

VALIDITY

Evaluation of **validity** may results in development of new **OM** instrument

Similar in length to the original but with different scoring or instructions	**+**	-Shortened for ease of use.	**+**	-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.

Validity

Logical/Translation

Face

Content

Construct

Convergent

Discriminate

Criterion

Concurrent

Predictive

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.

❖ Most widely used form of validity in developing countries

Relevance
Representativeness

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
  - ❖Weaker form of validity compared to other types of validity, except face validity

Professional Sample	Client/Patients	Health care professional
Phase 1	**Phase 2**	**Phse3**
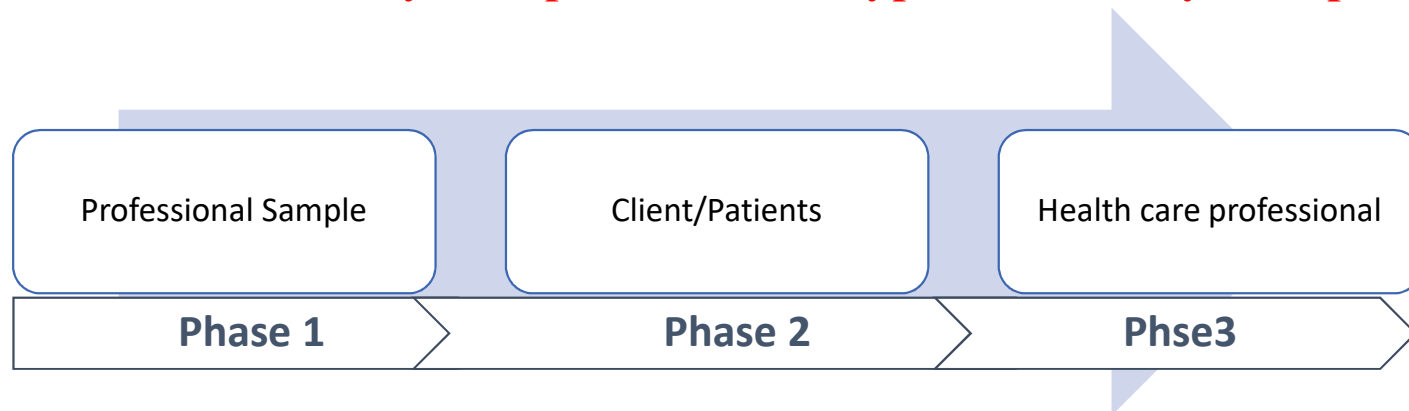
Criterion Validity

Examine the degree of correlation of a measure with other reference standard OMs of the same domain/construct.

If the new measure agrees with the gold standard measure, the new measure by association must be a valid technique

Criterion Validity
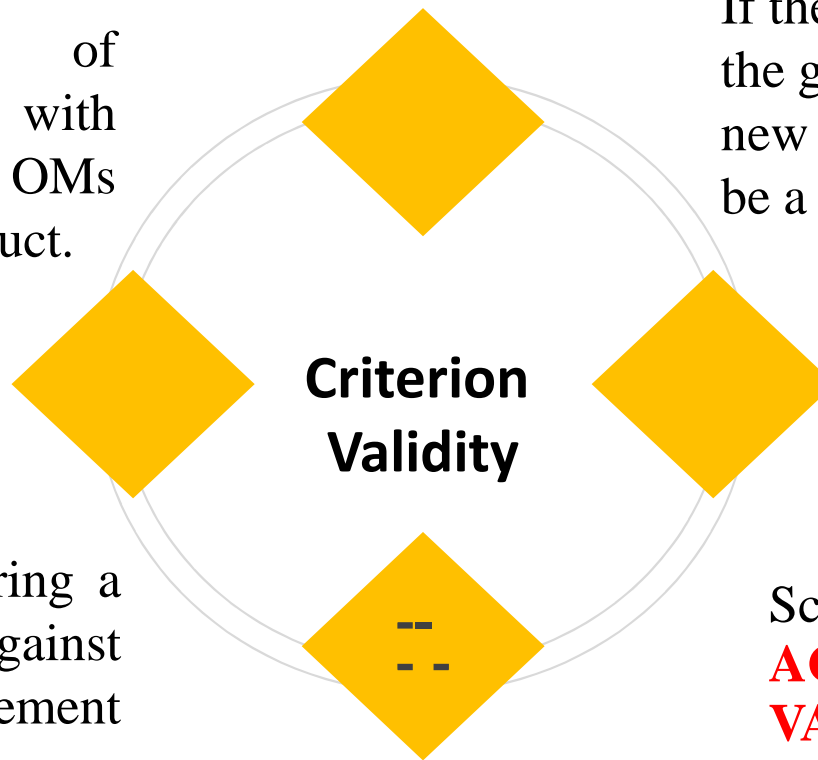
Established through comparing a new or untested OM tool against an accepted measurement technique, often known as a "**GOLD STANDARD**".

Scientific evidence supports the **ACCURACY AND VALIDITY** of the GOLD STANDARD method.

Criterion validity

Predictive Validity

Assesses the ability of the questionnaire/ instrument to forecast future events, behaviour, attitudes or outcomes

Berg Balance Test to predicate falls over the following 6 weeks

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

Exercices tolerance & VO2max

Criterion validity

Predictive Validity

Preoperative 6-minute walk test for predicting postoperative pulmonary complications

Berg Balance Test to predicate falls over the following 6 weeks

Concurrent validity

Grip strength a predictor for total muscle strength

MMT & dynamometer

Exercices tolerance & VO2max

Criterion validity

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

❖Pearson-product moment correlation (PPMCC)

❖Spearman's rank order correlation

❖Kendall's rank order correlation

❖Phi coefficient (   ).

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.

Construct Validity

Construct validity is the degree to which an instrument measures the trait or theoretical construct that it is intended to measure

❖It does not have a criterion for comparison rather it utilizes a hypothetical construct for comparison

❖Most valuable and most difficult measure of validity.

❖Basically, it is a measure of how meaningful the scale or instrument is when it is in practical us

Several methods to test's construct validity:

1. known-groups validity
2. Hypothesis/theory testing
3. Factor analysis validity

Construct Validity

Convergent Validity

- Same concept measured in different ways yields similar results

- This form of validity examine the extent to which a measure's results agrees 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 " no gold standard" convergent validity.

Discriminate validity

- Reflects the degree to which an instrument can distinguish between or among different concept or construct

- One concept is different from other closely related concepts

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

- **The UK version of the Seattle Angina Questionnaire (SAQ-UK)**

- **Reliability and Structural and Construct Validity of the Functional Strength Measurement in Children Aged 4 to 10 Years**

- **Known-Groups Analysis of the Harris Infant Neuromotor Test**

Validity and Reliability of the Chronic Respiratory Disease Questionnaire in Elderly Individuals with Mild to Moderate Non-Cystic Fibrosis Bronchiectasis

Objective

Test –retest Reliability

Internal consistency

convergent and Discriminant validity

Data and sampling

Setting
Primary care setting

Sampling

85 Confirmed diagnosis of non-CF bronchiectasis

4 Instruments to Measure Quality of Life

LCQ

a 19-item, self-administered question- naire of three domains measuring the physical, psychological and social impact of chronic cough, with a total severity score ranging from 3 to 21 and lower scores indicating greater impairment [21, 36].

CRDQ

consists of 20 items and assesses the domains of dyspnoea, fatigue, emotional function and mastery, with a total score ranging from 0 to 100 and higher scores signifying better HRQOL

SGRQ is a self-administered tool which consists of 76 items and comprises three components: symptoms, activity and impact, with a total score ranging from 0 to 100 and higher scores indicating a poorer HRQOL

Hospital Anxiety and Depression Scale (HADS)

Statistical Analysis

A — **Internal consistency**

B — Test-retest reliability

C — Validity

The internal consistency of the CRDQ was measured using Cronbach's α, based on the baseline assessment of the three do- mains and total score in all participants.

For assessment of reliability, the intra-class correlation coefficient (ICC)(2, 1) was measured in a group of 43 participants who had no change in management between baseline and the follow-up visit 9 weeks later

convergent validity was assessed using total scores and specific domains

Discriminant validitywas tested using the CRDQ domains and total score with lung function (spirometry) and 6MWD.

	CRDQ				
	dyspnoea	fatigue	emotional function	mastery	total
FEV$_1$ (% predicted)	−0.08	0.08	0.01	0.024	0.01
FVC (% predicted)	−0.04	0.08	−0.07	−0.003	−0.02
LCQ					
Total	0.00	0.406**	0.561**	0.619**	0.512**
Physical	0.06	0.551**	0.53*	0.669**	0.571**
Psychological	−0.04	0.334*	0.559**	0.556**	0.467**
Social	0.098	0.177	0.170	0.188	0.153
SGRQ					
Total	0.28	−0.541**	−0.471**	−0.685**	−0.493**
Symptoms	0.38*	−0.319*	−0.342*	−0.559**	−0.367*
Activity	0.215	−0.63**	−0.306*	−0.463**	−0.346*
Impact	0.26	−0.383*	−0.521**	−0.684**	−0.482**
HADS					
Anxiety	0.004	−0.381*	−0.550**	−0.442**	−0.459**
Depression	−0.14	−0.43*	−0.529**	−0.602**	−0.555**
6MWD (m)	−0.139	−0.280*	0.176	0.345**	0.189
mMRC	0.112	0.223	−0.138	−0.301*	−0.160

Data are based on Pearson's correlation coefficient or Spearman's rank coefficient. FEV$_1$ = Forced expiratory volume in 1s; FVC = forced vital capacity. * $p < 0.05$; ** $p < 0.01$.

RELIABILITY AND VALIDITY OF A CHINESE VERSION OF THE PEDIATRIC EVALUATION OF DISABILITY INVENTORY IN CHILDREN WITH CEREBRAL PALSY

Objective

Test –retest & Internal consistency

Concurrent validity

Discriminant validity

Data and sampling

Setting

National Taiwan University Hospital and 2 nursery schools in Taipei

Sampling

2 groups: 58 children with CP and 89 normally developing children

23

4 Instruments to Measure Quality of Life

PEDI

The PEDI contains 3 scales that can be used together or separately: a Functional Skill Scale, a Caregiver Assistance Scale, and a Modifica- tions Scale.

+

GMFC

The GMFCS (27) is a 5-level system providing a standardized clas- sification of the patterns of motor disability for children with CP from birth to 12 years of age.

+

The WeeFIMTM (31–33) measures the amount of assistance a child needs in order to perform daily activities

Statistical Analysis

A — Internal consistency

The internal consistency of the CRDQ was measured using Cronbach's α, based on the baseline assessment of the three do- mains and total score in all participants.

B — Test-retest reliability

test-retest reliability, the caregivers of the children with CP were interviewed with the PEDI-C twice, 2 weeks apart

C — Validity

To assess concurrent validity, the therapist interviewed all the caregivers with the WeeFIMTM in the week subsequent to the second PEDI-C interview

For the discriminative validity, 2 occupational thera- pists interviewed all caregivers of the normally developing children.

11/5/2018

Domain of the MS	Items of each domain								
	1	2	3	4	5	6	7	8	Mean
Self-care	0.87	0.91	0.63	0.83	0.95	0.85	1.00	0.97	0.88
Mobility	0.88	0.93	0.63	0.77	0.77	0.72	0.78	–	0.78
Social Function	1.00	1.00	1.00	1.00	1.00	–	–	–	1.00

PEDI-C	WeeFIM™ domains			
	Self-care	Mobility	Cognition	Total
Functional Skills Scale				
Self-care	0.95	0.87	0.90	0.97
Mobility	0.85	0.97	0.76	0.92
Social Function	0.86	0.70	0.92	0.87
Caregiver Assistance Scale				
Self-care	0.96	0.78	0.89	0.93
Mobility	0.84	0.99	0.73	0.92
Social Function	0.88	0.68	0.95	0.88

Covariate	Estimate	Standard error	Wald chi-square	p-value	Odds ratio
CP group logistic regression model I[1]: Functional Skills Scale factor scores (FSSfs) and age					
Intercept	−8.80	1.60	30.31	<0.0001	–
Age	2.71	0.53	26.50	<0.0001	15.01
Age × FSSfs	−1.57	0.31	25.09	<0.0001	0.21
CP group logistic regression model II[2]: Caregiver Assistance Scale factor scores (CASfs) and age					
Intercept	−8.23	1.49	30.44	<0.0001	–
Age	2.29	0.44	26.95	<0.0001	9.87
Age × CASfs	−1.14	0.22	26.37	<0.0001	0.32
CP group logistic regression model III[3]: both combined factor scores (CBfs) and age					
Intercept	−9.59	1.77	29.21	<0.0001	–
Age	2.85	0.55	26.53	<0.0001	17.22
Age × CBfs	−1.49	0.30	25.57	<0.0001	0.23