

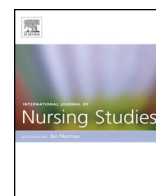


ELSEVIER

Contents lists available at [ScienceDirect](#)

## International Journal of Nursing Studies

journal homepage: [www.elsevier.com/ijns](http://www.elsevier.com/ijns)



# Assessing measurement in health: Beyond reliability and validity

Denise F. Polit <sup>a,b,\*</sup>

<sup>a</sup> Humanalysis, Inc., 75 Clinton Street, Saratoga Springs, NY 12866, USA

<sup>b</sup> Centre for Health Practice Innovation, Griffith University, 170 Kessels Road, Nathan, Brisbane, QLD 4111, Australia

### ARTICLE INFO

#### Article history:

Received 2 April 2015

Received in revised form 30 June 2015

Accepted 3 July 2015

#### Keywords:

Change scores

COSMIN

Instrument development

Measurement

Measurement error

Psychometrics

Reliability

Responsiveness

Scale development

Validity

### ABSTRACT

**Background:** Psychometric concepts have undergone a transformation in health fields, as articulated in a consensus report by an international panel of health measurement experts: COSMIN, the Consensus-based Standards for the selection of health Measurement Instruments.

**Objectives:** The aims of this paper are to describe emerging ideas relating to the development and testing of new measures in health fields, to present a revised measurement taxonomy that builds upon COSMIN, and to explore the extent to which the new measurement concepts have played a role in psychometric assessments in nursing. **Design:** A descriptive analysis of a sample of psychometric papers published in three major nursing journals was undertaken.

**Methods:** A new measurement taxonomy is presented and explained. A sample of 105 studies, representing a consecutive sample of psychometric studies published in the *International Journal of Nursing Studies*, *Nursing Research*, and *Research in Nursing & Health* between 2010 and 2014 was reviewed to ascertain the extent to which psychometric assessments in nursing map onto the new taxonomy.

**Results:** Most nursing studies reviewed adhered to traditional concepts of psychometric assessment, which focus on reliability and validity. The studies in the sample rarely involved assessments of longitudinal measurement aspects, namely the reliability and validity of change scores (responsiveness).

**Conclusions:** Many constructs of interest to nurse researchers are amenable to change—and these constructs are frequently the target of nursing interventions designed to foster change. Future psychometric work by nurse researchers would benefit from assessments of the psychometric adequacy of change scores.

© 2015 Elsevier Ltd. All rights reserved.

### What is already known about the topic?

- Nurse researchers have developed and evaluated hundreds of new scales to measure constructs of importance in clinical and research settings.

- Traditional psychometric assessment procedures, often adopted by nurse researchers, emphasize the measurement properties of reliability and validity.
- A Delphi survey of measurement experts in medicine and clinical epidemiology led to the creation of an expanded, transformational view of key measurement properties for health-related measurements, as articulated in COSMIN (the **C**onsensus-based **S**tandards for the selection of health **M**easurement **I**nstruments).

\* Correspondence to: Humanalysis, Inc., 75 Clinton Street, Saratoga Springs, NY 12866, USA. Tel.: +1 518 587 3994.

E-mail address: [denisefpolit@gmail.com](mailto:denisefpolit@gmail.com)

<http://dx.doi.org/10.1016/j.ijnurstu.2015.07.002>

0020-7489/© 2015 Elsevier Ltd. All rights reserved.

## What this paper adds

- Building on the COSMIN work, this paper introduces a new measurement taxonomy that includes two cross-sectional domains (reliability and validity) and two longitudinal domains (reliability of change scores and responsiveness—the validity of change scores).
- Based on an analysis of 105 psychometric papers published in three leading nursing research journals over a 5-year period, it was concluded that new measurement concepts, especially those relating to change scores, have not yet received much attention by nurse researchers.

Measurement concepts have evolved considerably in the past few decades, and new guidance from an expert panel on measurement in health has emerged. The purpose of this paper is to highlight major evolving measurement concepts of relevance to nurse researchers, to present a new measurement taxonomy, and to explore the extent to which psychometric work in nursing journals maps onto current guidelines for rigorous assessment of scales.

### 1. The evolution of measurement properties

In many fields in which measures of human attributes are developed and tested, classic ideas established by psychometricians decades ago have prevailed. The two measurement properties that have been the focus of standard psychometric assessment are *reliability* and *validity*. Nurse researchers, who have developed and evaluated hundreds of new scales, have largely followed the guidance of prominent psychometricians. Classic books such as the one by [Nunnally and Bernstein \(1994\)](#) are often cited by nurse researchers, as are books on scale development written by psychometricians such as [DeVellis \(2012\)](#) and [Streiner and Norman \(2008\)](#).

In medicine, however, measurement concepts have evolved beyond what is traditional. Indeed, several new ideas reflect a revolt against some aspects of classic psychometrics. The “revolution” has, however, experienced some turbulence, with a great proliferation of terms, definitions, and operationalizations of emerging measurement properties. In an effort to bring order to the turmoil that characterized measurement contributions in medicine, a working group in the Netherlands undertook a Delphi study with the goal of arriving at a consensus among an international panel of measurement experts. Their purpose was to identify and define critical measurement properties for health researchers, to array those properties in a taxonomy, and to create checklists for evaluating measurement papers.

The Delphi study resulted in the creation of COSMIN, the **C**onsensus-based **S**tandards for the selection of health **M**easurement **I**nstruments ([Mokkink et al., 2010a,b](#)). The new measurement ideas articulated in COSMIN await more widespread adoption in health fields other than medicine and clinical epidemiology.

The COSMIN taxonomy, available on the COSMIN website ([www.cosmin.nl](http://www.cosmin.nl)) includes three key measurement properties: reliability, validity, and responsiveness. More recently, [Polit and Yang \(2016\)](#) have proposed a

slight modification of the COSMIN taxonomy that more specifically incorporates a time element. As shown in [Fig. 1](#), the standard properties of reliability and validity are depicted as being relevant for assessments of cross-sectional (point-in-time) measurements. For longitudinal measurement—that is, for measuring *change* in a construct—the two relevant measurement properties in this taxonomy are the reliability of change scores and responsiveness. In both the COSMIN and the Polit-Yang taxonomies, *interpretation* of scores and change scores is an important aspect of measurement rigor.

The current paper reviews some of the new measurement ideas reflected in these two taxonomies. To explore the extent to which new measurement concepts have penetrated into the nursing literature, a content analysis of a small sample of psychometric papers published in nursing journals was undertaken.

*The sample of psychometric nursing papers.* A sample of 105 papers that reported the psychometric testing of a scale was analyzed. A consecutive sample of papers published between 2010 and 2014 was drawn from three general nursing research journals that frequently publish papers on scale development and testing: *International Journal of Nursing Studies* ( $N=58$ ), *Nursing Research* ( $N=19$ ), and *Research in Nursing & Health* ( $N=28$ ). To be eligible for this analysis, the study had to involve one of the following: the development and psychometric assessment of a new instrument; the translation of an instrument into another language and an evaluation of the translated version; a psychometric evaluation of an existing instrument for a new population; or an evaluation of an instrument adaptation (e.g., the testing of a short form). Papers were excluded if they focused on a narrow psychometric question (e.g., a content validity effort or a factor analysis only), if the focus was to compare multiple scales, or if the paper was a systematic review. Papers were coded for the types of psychometric assessments that were undertaken. Intercoder reliability was assessed for a subsample of papers ( $N=25$ ), and was found to be high ( $\kappa = .91$ ).

### 2. The reliability domain

In both the COSMIN and Polit-Yang taxonomies, the reliability domain encompasses three components: reliability, internal consistency, and measurement error. Reliability can be defined as the degree to which “... scores for people *who have not changed* are the same for repeated measurements, under several situations” ([Polit and Yang, 2016](#), p. 25), including repetition on different occasions (test–retest reliability and intra-rater reliability), by different persons (inter-rater reliability), or in the form of different replicates (items) on a multi-item instrument (internal consistency).

*Internal consistency.* Internal consistency concerns the degree to which the items on a scale are measuring the same underlying construct. Nurse researchers have tended to follow the psychometric tradition of emphasizing internal consistency as the most important aspect of reliability, and typically rely on Cronbach’s alpha as the measurement parameter to be estimated. In the sample of

Download English Version:

<https://daneshyari.com/en/article/7515696>

Download Persian Version:

<https://daneshyari.com/article/7515696>

[Daneshyari.com](https://daneshyari.com)