HSOR NEWS
The evidence behind your decisions

Health Services & Outcomes Research, **Issue 2013/1**
19 February 2013

RELIABILITY AND VALIDITY IN MEASUREMENT TOOLS

INSIDE THIS ISSUE

INTRODUCTION

Clinicians and researchers use various instruments to measure physical measurement or abstract concepts like cognitive ability to draw inferences on patients. The tools have to fulfil 2 prerequisites: ***reliability*** and ***validity***. In this article, we will explore these and state their importance.

What is *Reliability*?

Reliability is concerned with random errors. A reliable instrument gives the same results when repeated under similar conditions. The random errors occur due to chance and cause measurements to fluctuate in both directions. The 4 common types of reliability are inter-rater, intra-rater, test-retest and internal consistency. Inter-rater and intra-rater reliability look at the random error associated with the person administering the measure. Test-retest reliability looks at the random error associated with administering the test over multiple occasions and internal consistency looks at the inter-relatedness of the items within the measure.

What is *Validity*?

Validity is concerned with systematic errors. If an instrument is valid, it measures what it is intended to measure. If it does not, it would yield errors in the measurement that differs conceptually with what is intended, which gives rise to the systematic nature in which the errors are committed [1]. While there have been much debate over how to classify the various forms of validity [2,3], it has been acknowledged that a unified framework of validity is needed for 2 main reasons [4]. The first is that the distinctions between traditional concepts of validity (content, criterion-related and construct) are unclear. Secondly, the implications of the measurement as a basis for action and the consequences of its intended use need to be considered together with the various forms of validity.

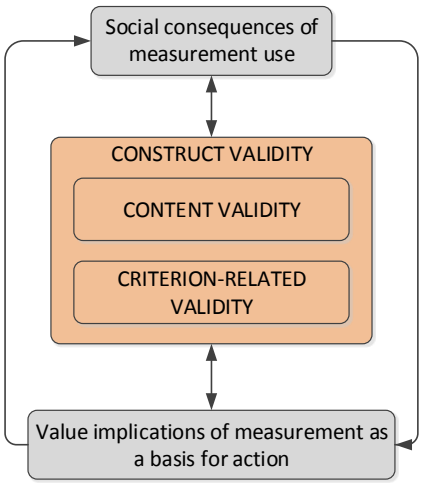

Figure 1 Universal framework of validity

Figure 1 summarises the new unified framework. Of note is that it assumes both content and criterion-related validity under construct validity, providing clarity that they are all inter-related and facets of validity related to the construct of the measure.

Why are they important?

Figure 2 shows the main difference between validity and reliability. If the real measurement is the bull's eye, the black dots are our measurement attempts. Consistently grouped dots like C and D would mean that the measurements are reliable while dots gathered near the bull's eye would mean that the measurements are valid. *Without reliability and validity, we would not know if we are measuring what we claim we are measuring and how precisely are we measuring it.*

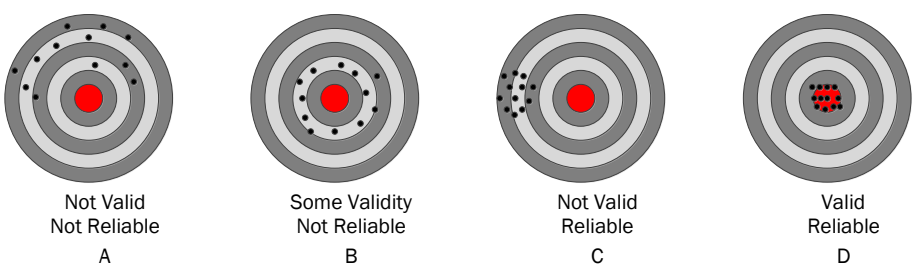

Figure 2 Differences between reliability and validity

Case study: Validation of the Paediatric Hearing Impairment Caregiver Experience assessment tool

A paediatric hearing impairment caregiver experience assessment tool was developed in the United States, but its applicability has not been assessed before in Singapore. However, with data from a prior study that collected the assessment results, we analysed the evidence for local validity. The study enrolled parents of children who have hearing impairment and administered the assessment at the start of the study. Though the study did not start out with the intention of validating the instrument, our analysis provided some insight on the evidence supporting the validity of the instrument.

In this case, structural validity (a component of construct validity) was explored using both exploratory and confirmatory factor analysis methods. Structural validity aims to understand the latent factor structure that underscores the instrument. The original instrument conceptualised that experience of caring for a child with hearing impairment have 8 domains. We found that the data fitted better with a 5 domain model. This provides evidence on the structure of the concept (in this case, caregiving experience) that we are trying to measure.

Given the limitation of the study, we could only compute internal consistency for each of the domains for reliability. In an ideal situation, a more robust design such as the different models of the intra-class correlation (ICC) can be used to compute intra and inter-rater reliability and test-retest reliability. These indices would allow us to build a confidence interval known as the standard error of measurement (SEM) around the point estimate of the measurement. This provides us with a good grasp on the level of uncertainty that we have about the point estimate and to what extent can we trust it as being a good representation of the true measurement.

References

1. Nunnally, J. C. Psychometric Theory. (Mcgraw-Hill College, 1978).
2. Cronbach, L. J. & Meehl, P. E. Construct validity in psychological tests. Psychol Bull 52, 281–302 (1955).
3. Loevinger, J. A systematic approach to the construction and evaluation of tests of ability. Psychological Monographs 61, iii–49 (1947).
4. Messick, S. Constructs and their vicissitudes in educational and psychological measurement. Psychological Bulletin 89, 575–588 (1981).

Our past newsletters are available under "Publications" section at http://www.hsor.nhg.com.sg/

Author

Li Ruijie, MSc (Occupational Therapy)
Senior Research Analyst

Ruijie worked as an occupational therapist prior to joining NHG as a research analyst. He has experience working in both the acute and community healthcare settings in Singapore. Ruijie obtained his Masters in Hong Kong and has since worked on numerous research projects aimed at documenting and improving healthcare related outcomes both in Hong Kong and locally. His research interests include assessing psychometric properties of instruments and using various quantitative and qualitative methods in analysis and interpretation of the data collected in studies.

Feedback and enquiries: hsor@nhg.com.sg