

Measuring Outcomes: Validity, reliability and sensitivity



Dr Adrian MM Schoo

“Not everything that can be
counted counts, and not
everything that counts can be
counted”

(Albert Einstein 1879-1955)

Valid Questions

- Why Measuring Outcomes
- How credible are the measures? Do they measure what we want to measure?
- Does it make a difference who is measuring (e.g., client-clinician, skilled or novice)?
- Are there preferred measures for specific conditions, clients or research?
- What can be predicted from the results?

Psychometric Properties of Standardised Outcome Measures

- Validity
 - Face or Content
 - Criterion (i.e., concurrent or predictive)
 - Construct (e.g., sensitivity to skill level)
- Reliability
 - Internal consistency
 - Test-retest (e.g., intrarater) reliability
 - Inter-observer or interrater reliability
- Sensitivity and responsiveness

Validity

- Does the instrument measure what it is supposed to measure?
- Are results dependent on the level of skill of the assessor?
- What can be inferred about what is measured?
- What might be inferred about other domains from the results of the assessment?

Reliability

- Internal consistency is the consistency of the instrument across its items (Cronbach 1951)
- Test-retest (e.g., intrarater) is the stability of an instrument when no important change has occurred (Nunnally 1979)
- Could even test to see if results are consistent across settings (would be less reliable due to different circumstances, preferences etc.)

Sensitivity and Responsiveness

(Laing 2000)

- Sensitivity is the ability of an instrument to measure change regardless of whether it is meaningful
- Responsiveness is the ability of an instrument to measure meaningful change

When Assessing Psychometric Properties of an Instrument

- Pearson's r or Intraclass Correlation Coefficients
- Are correlations 0.80 or stronger. Correlations with other instruments (concurrent validity) are weaker since they tap into different domains)
- What was the sample and purpose? (instruments are designed for specific populations and specific purposes)
- Was the sample random, convenient or stratified
- Where assessors experienced

Enhancing Reliability of an Instrument

- Ensuring that the scoring instructions are clear and unambiguous
- Ensuring that the test is logical and sensible
- Ensuring that administration of the test is simple and clear
- Ensuring that self-assessment instruments can be read (i.e., older people)