

Supplementary File 1. Definitions of measurement properties

Measurement properties

Validity

The validity of a measure refers to whether an instrument assesses what it is intended to measure, including:

- Content validity - whether a measure adequately assesses all aspects of the phenomenon it seeks to measure (includes face validity). Typically established during scale development by consultation with experts and target population representatives
- Construct validity assesses whether a new instrument measures the underlying construct(s) it is intended to. Factor analysis and related techniques can be used to examine the underlying (latent) dimensions of a new scale and/or its subscales
- Convergent validity – whether scores on the measure correlate with scores on existing measures of the same or related constructs
- Discriminant validity – whether scores on the measure are uncorrelated with scores on existing measures known to assess different constructs
- Known groups validity - the ability of a measure to adequately distinguish between groups who are known to differ on the construct measured
- Criterion validity – the extent to which scores on a measure concur with a gold-standard measure of the same construct
- Cross-cultural validity – the degree to which item performance is consistent in translated or culturally adapted versions
- Ecological validity – the extent to which a measure is accurate in assessing/predicting real-world outcomes/behaviours

Reliability

The reliability of a measure can be defined as the extent to which it is free from measurement error. The following types of reliability were considered:

- Inter-rater reliability (IRR) - the degree of consistency between scores provided by independent raters
- Internal consistency - the degree to which different items on a measure assess the same underlying construct, when intended to do so. Usually quantified as the degree to which item scores on a measure correlate with one another, commonly yielding a Cronbach's alpha score
- Test-retest reliability - the stability of scores on a measure across time points, assuming all other factors remain the same

Responsiveness (sensitivity to change)

The degree to which scores on a measure change over time, in accordance with a known or expected change in the construct (e.g., pre-post treatment symptom change)