
Article:
Statham, V., Emerson, L.-M. and Rowse, G. orcid.org/0000-0003-3292-4008 (2019) A systematic review of self-report measures of paranoia. Psychological Assessment, 31 (2). pp. 139-158. ISSN 1040-3590

https://doi.org/10.1037/pas0000645

Appendix C

Table B1

COSMIN recommended appraisal criteria for questionnaire properties

Domain	Definition	Appraisal criteria
Reliability		
Internal	Interrelatedness among scale items	P: Cronbach's alpha(s) ≥ .7 I: No Cronbach's alphas, or dimensions undetermined N: Cronbach's alpha(s) < .7
Test-retest	The proportion of the variance in measurements which is due to 'true' differences between patients	P: Intra-class correlation/weighted Kappa ≥ .7, or, Pearson's r ≥ .8 I: Neither of the above analyses reported N: Intra-class correlation/weighted Kappa < .7, or Pearson's r < .8
Validity	Whether scores on an instrument adequately reflect a 'gold standard'. The only real gold standard comparator instrument is another version of the same measure. Other comparators are often wrongly considered as gold standard	
Criterion		P: Criterion variable is a 'gold standard' assessment of the construct. Correlation ≥ .7 I: No information. No arguments that criterion variable is 'gold standard' N: Correlation with gold standard < .7
Content	How well an instrument reflects the construct to be measured	P: Items relevant to construct and target population, and questionnaire is comprehensive I:Not enough information N: Some items irrelevant to construct, or for target population, or questionnaire is not comprehensive
Construct Validity		
Hypothesis testing	How consistent instrument scores are with hypotheses. For example, about relationships with other instruments (convergent/divergent validity) or differences between relevant groups (known-groups validity).	P: Correlations with associated measures ≥ .5, or at least 75% results are in accordance with hypotheses and correlations with related constructs are higher than with unrelated constructs I: No information, or correlations solely with unrelated constructs presented N: Correlations with associated measures < .5, or less than 75% of hypotheses accepted

(continued)

Domain	Definition	Appraisal criteria
Cross-cultural validity	Whether items on a translated or culturally adapted instrument adequately reflect of the performance of items on the original instrument.	P. No difference in factor structure or important differential item functioning (DIF) between versions I: Factor analysis not applied and DIF not assessed N: Differences in factor structure or DIF between languages
Structural validity	Whether scores on an instrument adequately reflect the dimensionality of the construct to be measured.	P: Factors explain at least 50% of the variance I: Explained variance not mentioned, or factor analysis not completed. N: Factors explain < 50% variance
Responsiveness	The ability of an instrument to detect change over time in the construct to be measured.	P: Correlation with changes on instruments measuring the same construct ≥ 0.5 OR at least 75% of results in accordance with hypotheses OR area under the curve ≥ 0.7 and correlations with changes in related constructs are higher than unrelated constructs I: Appropriate analyses not conducted, or solely correlations determined with unrelated constructs. N: Responsiveness analyses conducted but criteria not met

Note. Information in table obtained from Terwee et al. (2011) and Mokkink et al. (2012). P = Positive evidence, I = indeterminate evidence, N = negative evidence