

Table of quality criteria for outcomes (From Prinsen (2018), based on Terwee (2007) and Prinsen (2016))

Measurement property	Assessment	Criteria
Structural validity	+	CTT: CFA: CFI or TLI or comparable measure >0.95 OR RMSEA <0.06 OR SRMR <0.082 IRT/Rasch: No violation of unidimensionality: CFI or TLI or comparable measure >0.95 OR RMSEA <0.06 OR SRMR <0.08 AND no violation of local independence: residual correlations among the items after controlling for the dominant factor <0.20 OR Q3's < 0.37 AND no violation of monotonicity: adequate looking graphs OR item scalability >0.30 AND adequate model fit: IRT: $\chi^2 > 0.01$ Rasch: infit and outfit mean squares ≥ 0.5 and ≤ 1.5 OR Z standardized values > -2 and <2
	?	CTT: Not all information for '+' reported IRT/Rasch: Model fit not reported
	-	Criteria for '+' not met
Internal consistency	+	At least low evidence ⁴ for sufficient structural validity AND Cronbach's alpha(s) ≥ 0.70 for each unidimensional scale or Subscale.
	?	Criteria for "At least low evidence for sufficient structural validity" not met
	-	At least low evidence for sufficient structural validity AND Cronbach's alpha(s) < 0.70 for each unidimensional scale or subscale
Reliability	+	ICC or weighted Kappa ≥ 0.70
	?	ICC or weighted Kappa not reported
	-	ICC or weighted Kappa < 0.70
Measurement error	+	SDC or LoA < MIC
	?	MIC not defined
	-	SDC or LoA > MIC
Hypotheses testing for construct validity	+	The result is in accordance with the hypothesis
	?	No hypothesis defined (by the review team)
	-	The result is not in accordance with the hypothesis
Cross-cultural validity\measurement invariance	+	No important differences found between group factors (such as age, gender, language) in multiple group factor analysis OR no important DIF for group factors (McFadden's R2 < 0.02)
	?	No multiple group factor analysis OR DIF analysis performed
	-	Important differences between group factors OR DIF was found
Criterion validity	+	Correlation with gold standard ≥ 0.70 OR AUC ≥ 0.70
	?	Not all information for '+' reported
	-	Correlation with gold standard < 0.70 OR AUC < 0.70
Responsiveness	+	The result is in accordance with the hypothesis OR AUC ≥ 0.70
	?	No hypothesis defined (by the review team)
	-	The result is not in accordance with the hypothesis OR AUC <0.70

AUC: area under the curve, CFA: confirmatory factor analysis, CFI: comparative fit index, CTT: classical test theory, DIF: differential item functioning, ICC: intraclass correlation coefficient, IRT: item response theory, LoA: limits of agreement, MIC: minimal important change, RMSEA: Root Mean Square Error of Approximation, SEM: Standard Error of Measurement, SDC: smallest detectable change, SRMR: Standardized Root Mean Residuals, TLI = Tucker-Lewis Index. "+" = sufficient; "-" = insufficient; "?" = indeterminate.