

Structural Validity, Reliability, Measurement Invariance, and Construct Validity of the Personal Growth Initiative Scale-II: A Systematic Review Using COSMIN Methodology.

International Journal of Applied Positive Psychology

*Katleen Verdoodt, Jennifer Reijnders, Natascha de Hoog, Nele Jacobs, and Marianne Simons

*Faculty of Psychology, Open Universiteit, Heerlen, The Netherlands

Katleen.verdoodt@ou.nl

Online Resource 3

Criteria for Good Measurement Properties (Mokkink et al., 2024)

Measurement property	Rating	Criteria
Content validity	+	Included items are relevant for the construct, target population, and context of use, and response options and recall period are appropriate <i>AND</i> No key concepts are missing <i>AND</i> PROM items and response options are appropriately worded and PROM instructions, items and response options understood by the population of interest as intended
		?
	-	Not enough information reported Included items are not relevant for the construct or target population <i>OR</i> Key concepts are missing <i>OR</i> PROM items and response options are not appropriately worded or not understood by the population of interest as intended
Structural validity	+	CTT: EFA/PCA: factor loadings of each item on its factor ≥ 0.30 <i>AND</i> Maximum 10% of the items have factor loadings ≥ 0.30 on multiple factors <i>AND</i> Explained variance $\geq 50\%$ and structure is in line with the theory about the construct to be measured <i>OR</i> results on scree plot or Kaiser criterion (Eigenvalues > 1) are in line with the theory about the construct to be measured CFA: CFI or TLI or comparable measure > 0.95 <i>OR</i> RMSEA < 0.06 <i>OR</i> SRMR < 0.08 IRT/Rasch: No violation of <u>unidimensionality</u> : CFI or TLI or comparable measure > 0.95 <i>OR</i> RMSEA < 0.06 <i>OR</i> SRMR < 0.08 <i>AND</i> No violation of <u>local independence</u> : residual correlations among the items after controlling for dominant factor < 0.20 <i>OR</i> Q3s < 0.37 <i>AND</i>

		No violation of <u>monotonicity</u> : adequate looking graphs <i>OR</i> item scalability >0.30 <i>AND</i> Adequate <u>model fit</u> : IRT: $\chi^2 > 0.01$ Rasch: infit and outfit mean squares ≥ 0.5 and ≤ 1.5 <i>OR</i> Z-standardized values >-2 and <2
	?	Not enough information reported
	-	Criteria for '+' not met
Internal consistency	+	At least low evidence for sufficient unidimensionality <i>AND</i> Cronbach's alpha ≥ 0.70
	?	Criteria for "at least low evidence for sufficient unidimensionality" not met <i>OR</i> Evidence for insufficient unidimensionality <i>OR</i> Not enough information reported
	-	At least low quality evidence for sufficient unidimensionality <i>AND</i> Cronbach's alpha <0.70
Cross-cultural validity\ measurement invariance	+	No important differences found between group factors (such as age, gender, language) in multiple group factor analysis <i>OR</i> no important DIF for group factors (McFadden's $R^2 < 0.02$)
	?	Not enough information reported
	-	Important differences between group factors <i>OR</i> DIF was found
Reliability	+	ICC or (weighted) kappa or Pearson/Spearman correlation ≥ 0.70
	?	Not enough information reported
	-	ICC or (weighted) kappa or Pearson/Spearman correlation <0.70
Measurement error	+	SDC or LoA <MIC
	?	MIC not defined <i>OR</i> not enough information reported
	-	SDC or LoA > MIC
Criterion validity	+	Correlation with gold standard ≥ 0.70 <i>OR</i> AUC ≥ 0.70
	?	Not enough information reported
	-	Correlation with gold standard <0.70 <i>OR</i> AUC <0.70
Hypothesis testing for construct validity	+	$\geq 75\%$ of the results is in accordance with predefined hypotheses
	?	No relevant results were found
	-	$\geq 75\%$ of the results deviates from predefined hypotheses
Responsiveness	+	$\geq 75\%$ of the results is in accordance with predefined hypotheses <i>OR</i> AUC ≥ 0.70
	?	No relevant results were found
	-	$\geq 75\%$ of the results deviates from predefined hypotheses <i>OR</i> AUC <0.70

AUC = area under the curve, CFA = confirmatory factor analysis, CFI = comparative fit index, CTT = classical test theory, DIF = differential item functioning, EFA = exploratory factor analysis, ICC = intraclass correlation coefficient, IRT = item response theory, LoA = limits of agreement, MIC = minimal important change, PCA = principal component analyses, 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