

**Table S2.** Analyses and tests with reference values considered in Rasch analyses

<b>Analyses</b>	<b>Tests</b>	<b>Reference value</b>	<b>Software</b>
Data-model fit	Total item $\chi^2$ with $p$ -values	Insignificant total item $\chi^2$ indicates good data-model fit	RUMM2030Plus
Targeting	Person-item location	The distribution of person location estimates is compared to the distribution of the item location estimates. Mean person location values close to 0 indicate good targeting between the person's ability and the item difficulty <sup>a</sup>	RUMM2030Plus
Dimensionality	PCA/ $t$ -test procedure	Proportions of significant $t$ -tests >5 % indicate multidimensionality <sup>b</sup>	RUMM2030Plus
Response dependence	Residual correlations	>0.3 indicates response dependence <sup>c</sup>	RUMM2030Plus
Item fit	$\chi^2$ statistics with $p$ -values	$P$ -values <5 % Bonferroni-adjusted are deemed as misfit	RUMM2030Plus
	fit residuals	Values below/above -2.5/ 2.5 indicate over-/under-discrimination <sup>a</sup>	RUMM2030Plus
	Infit weighted mean square (MNSQ)	Infit values between 0.7 and 1.3 was considered as sufficient. Values lower than 0.7 indicate over-discrimination, whereas values higher above 1.3 indicate under-discrimination <sup>d</sup>	ConQuest5
Differential item functioning (DIF)	Analysis of variance (two-way ANOVA)	High F-values with probability <5% Bonferroni-adjusted indicates DIF <sup>e</sup>	RUMM2030Plus
Ordering of response categories	Ordering of thresholds	Inspecting the thresholds to determine whether they are in correct order <sup>e</sup>	RUMM2030Plus
Reliability	Person separation index (PSI)	>0.85 at individual level >0.65 at group level <sup>f</sup>	RUMM2030Plus

Suggestions for references (a selection) for the different analyses: <sup>a</sup>[45], <sup>b</sup>[6-48], <sup>c</sup>[55], <sup>d</sup>[54], <sup>e</sup>[56-58], <sup>f</sup>[40].