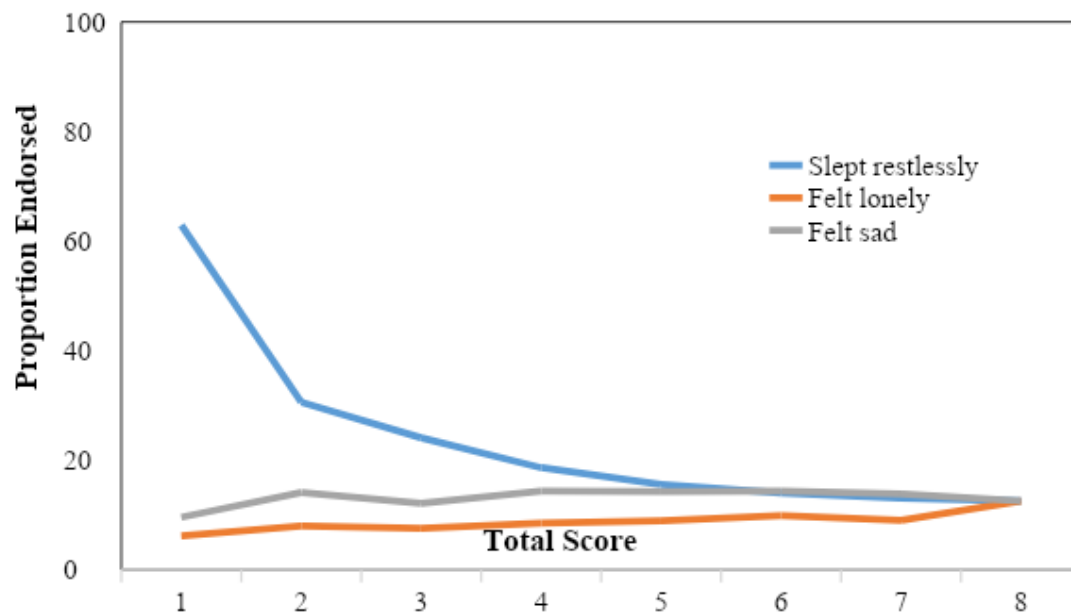
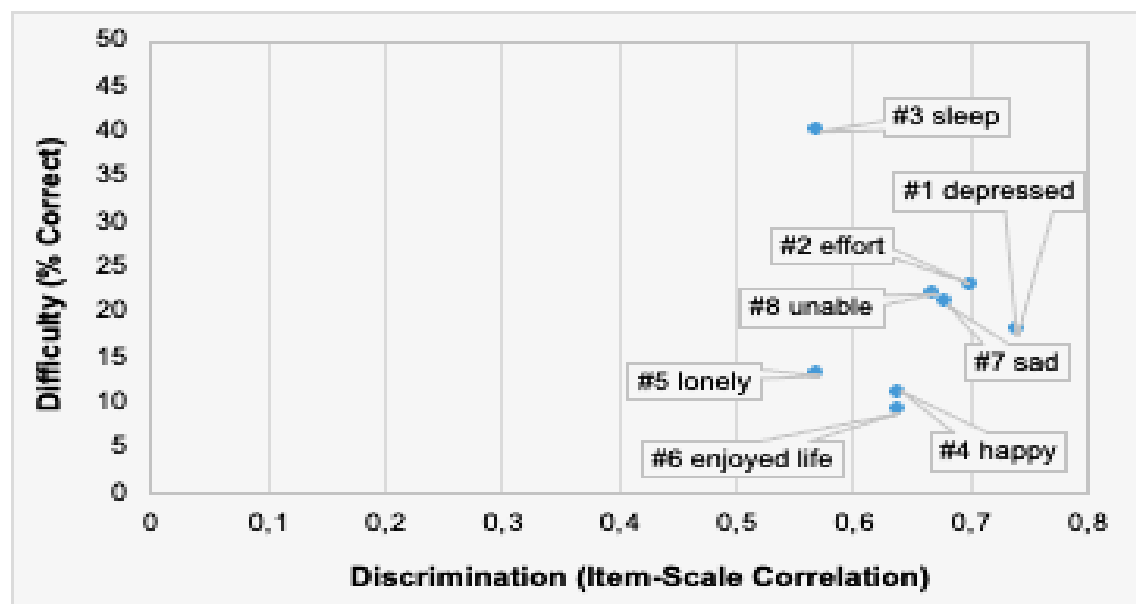
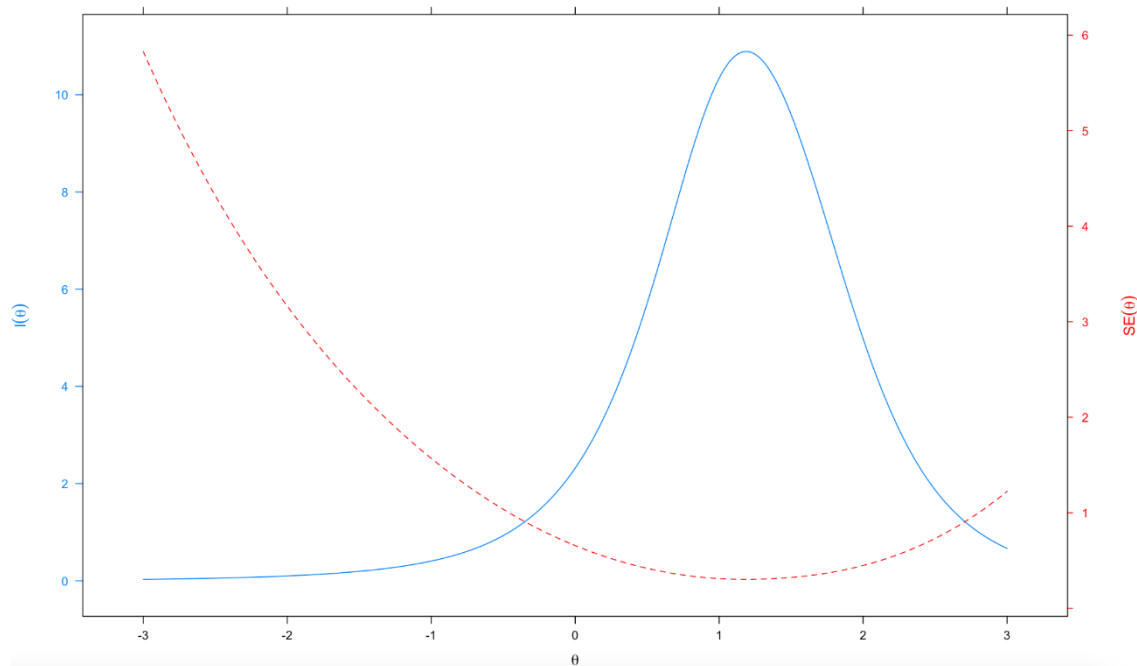


Supplementary Figure S1 Endorsement proportion for slept restlessly (item 3), felt lonely (item 5), and felt sad (item 7), according to the classical test theory approach**Supplementary Figure S2** Difficulty-by-discrimination graph according to the classical test theory approach

Discrimination (x-axis): the higher the score, the more discriminant the item. Difficulty (y-axis): the lower the score, the lower the odds of endorsing the item.

Supplementary Figure S3 CES-D 8 test information curve

CES-D = Center for Epidemiologic Studies Depression; SE = standard error.
In blue: amount of information yielded at each θ level. In red: SE at each θ level.

Supplementary Table S1 McFadden's ΔR^2 values for items flagged for DIF

Item Number	Item content	ΔR^2 (McFadden)
Item #3	Slept restlessly	0.0071
Item #5	Felt lonely	0.0027
Item #7	Felt sad	0.0118
Item #8	Could not get going	0.0015

DIF = Differential Item Functioning.

ΔR^2 values ≥ 0.01 are generally interpreted as small DIF effects and $\Delta R^2 \geq 0.02$ as moderate DIF effects (Jodoin MG, Gierl MJ. Evaluating type I error and power rates using an effect size measure with the logistic regression procedure for DIF detection. Appl Meas Educ 2001;14:329-49).