

Table SI. Definitions of item performance features used in classical test theory

Item	Item performance feature	Definition
1	Item difficulty	Proportion of missing scores among the 721 respondents. Item difficulty was considered high if 10% or more of scores were missing.
2	Response distribution	The proportion of patients who responded to each item with the same response was determined. An item was described as having a poor distribution if > 70% of patients had chosen the same response.
3	Item-test correlation	The Spearman's correlation coefficients (r) of each item with its subscale were calculated. If the r of an item differed >0.1 with the r of the other items in the subscale, it was considered suboptimal.
4	Item-rest correlation	The Spearman's correlation coefficients (r) of each item with the sum of the other items in that subscale were calculated. Suboptimal item-rest correlation was defined as $r < 0.20$.
5	Item discriminant validity	We compared the item-rest correlation coefficients with the correlation coefficients of an item with the other subscales ^a . If the former equalled or was smaller than the latter, an item was defined as having poor discriminant validity.
6	Item complexity	We investigated the factor loadings in a factor analysis for each item. Suboptimal complexity was said to exist if the highest loading of an item was <0.40 or if the difference between the loadings on different factors was <0.10.
7	Internal consistency	For each subscale, the Cronbach's α was calculated. If $\alpha < 0.70$, the internal consistency was considered suboptimal for each subscale's item.
8	Stepwise regression	For each subscale, a forward stepwise regression analysis was performed. If an item entered the model after 90% or more of the variance of that subscale was explained it was considered suboptimal.

^aSubscales derived from the principal component analysis were used.