ActaDV Acta Dermato-Venereologica ActaDV Advances in dermatology and venereology

Table SI. Overview of Rasch analysis for ordinal scales

Step	Description	Question	Evaluation method (interpretation)
1	Model selection	Is the distance between response categories (nearly) the same across items (RSM) or do they need to be estimated separately (PCM)?	Compare model fit via the LR test (significance supports the more complicated model), AIC (smaller value indicates a better model), or BIC (smaller value indicates a better model).
2	Evaluation of response categories	1. Do the response categories advance logically (higher response category equals higher impairment)? 2. Are response patterns consistent with the model's predictions?	1. Compare the response categories' mean measure (higher response categories are expected to have higher measures) 2. Unstandardized or standardized fit statistics (fit statistics expected to be within the acceptable range defined; both too unexpected and too expected response patterns are improbable and therefore misfit the model).
3	Evaluation of person and item fit	Do persons respond as expected by the model and are response patterns for items consistent with the model's predictions?	Unstandardized or standardized fit statistics (same as in step 2).
4	Differential item functioning	1. Do items function (nearly) the same across different groups of persons (sex, age, culture)? 2. In case items function differently across groups: are the differences consistent over the complete range of the measured attribute (uniform DIF) or limited to certain levels of the measured attribute (non-uniform DIF)?	1. Compare response patterns across groups and test whether meaningful differences are statistically significant via the Mantel-Haenzel procedure (significance supports DIF). 2. Test whether differences are consistent across the complete range of the measured attribute via ordinal logistic regression (DIF is non-uniform when the group-specific regression lines cross).
5	Model assumptions	1. Is a single attribute being measured? 2. Are items responses independent from each other?	1. Principal component analysis of the residuals (an identified meaningful component violates unidimensionality) 2. Pearson's correlation of the standardized item residuals (larger violation with higher correlation).
6	Targeting	How well is the scale calibrated to the population?	Location of persons and items on the measured attribute is compared to each other (less difference is better).
7	Person separation index	How many levels of the measured attribute can the scale distinguish between reliably?	Test how much larger the person ability variance is than the error variance (larger is better).
8	Scale optimization	Can the scale be improved if changes are made?	Modify different properties of the scale and then evaluate the effect via step 1–7.

DIF: differential item functioning; RSM: rating scale model; PCM: partial credit model; LR: Likelihood ratio test; AIC: Akaike's information criterion; BIC: Schwartz's Bayesian information criterion.