

Appendix 4. Approach for grading the quality of the evidence

Grade factor	Downgrading	Definition
Risk of bias	0	Multiple studies of at least adequate quality <i>OR</i> one study of very good quality
	-1	Only one study of adequate quality <i>OR</i> multiple studies of doubtful quality
	-2	Only one study of doubtful quality <i>OR</i> multiple studies of inadequate quality
	-3	Only one study of inadequate quality
Imprecision (not for content validity, structural validity, and cross-cultural validity\ measurement invariance)	0	Total sample size of all studies >100
	-1	Total sample size of all studies 50-100
	-2	Total sample size of all studies <50
Inconsistency	0	Results are consistent <i>OR</i> results are summarized and rated per subset of studies, and subsequently graded
	-1	Overall rating based on the majority of consistent results
Indirectness	0	Does not occur; definitions for construct and/or target population have been stated in the inclusion criteria

0: high, -1: moderate, -2: low, -3: very low; Per protocol of the COSMIN guideline for systematic reviews: the quality of evidence for internal consistency cannot be higher than the quality of evidence for structural validity[23]