

Appendix 1. Standards for evaluating the quality of PROM development

Check the COSMIN website to see if the quality of the PROM development was already rated in another review

Ratings: V= very good; A = adequate; D = doubtful; I = inadequate; N= not applicable

1a. PROM design

	Upton and Upton 2006			Gerrish et al. 2006			Carrner et al. 2010			Kerfmann et al. 2014			Harris-Haywood et al. 2014			Dagrez et al. 2016			Kostemak et al. 2017			Januseviciene et al. 2019			Rusafa-Martinez et al. 2020			Lemetti et al. 2021			Garcia-Salvador et al. 2021			Brathwaite et al. 2022		
	Rater 1	Rater 2	Consensus	Rater 1	Rater 2	Consensus	Rater 1	Rater 2	Consensus	Rater 1	Rater 2	Consensus	Rater 1	Rater 2	Consensus	Rater 1	Rater 2	Consensus	Rater 1	Rater 2	Consensus	Rater 1	Rater 2	Consensus	Rater 1	Rater 2	Consensus	Rater 1	Rater 2	Consensus	Rater 1	Rater 2	Consensus	Rater 1	Rater 2	Consensus
1	General design requirements																																			
1	Is a clear description provided of the construct to be measured?																																			
2	Is the origin of the construct clear: was a theory, conceptual framework or disease model used or clear rationale provided to define the construct to be measured?																																			
3	Is a clear description provided of the target population for which the PROM was developed?																																			
4	Is a clear description provided of the content of use (i.e. discriminative, evaluative purpose and/or predictive)																																			
5	Was the PROM development study performed in a sample representing the target population for which the PROM was developed?																																			
6	Concept elicitation (relevance and comprehensiveness)																																			
6	Was an appropriate qualitative data collection method used to identify relevant items for a new PROM?																																			
7	Were skilled group moderators/interviewers used?																																			
8	Were the group meetings or interviews based on an appropriate topic or interview guide?																																			
9	Were the group meetings or interviews recorded and transcribed verbatim?																																			
10	Was an appropriate approach used to analyse the data?																																			
11	Was at least part of the data coded independently?																																			
12	Was data collection continued until saturation was reached?																																			
13	For quantitative studies: was the sample size appropriate?																																			
	SUBTOTAL QUALITY CONCEPT ELICITATION STUDY (lowest score of items 6-13)																																			
	TOTAL QUALITY OF THE PROM DESIGN (lowest score of items 1-13)																																			

1b. Cognitive interview study or other pilot test

	Upton and Upton 2006			Gerrish et al. 2006			Carrner et al. 2010			Kerfmann et al. 2014			Harris-Haywood et al. 2014			Dagrez et al. 2016			Kostemak et al. 2017			Januseviciene et al. 2019			Rusafa-Martinez et al. 2020			Lemetti et al. 2021			Garcia-Salvador et al. 2021			Brathwaite et al. 2022		
	Rater 1	Rater 2	Consensus	Rater 1	Rater 2	Consensus	Rater 1	Rater 2	Consensus	Rater 1	Rater 2	Consensus	Rater 1	Rater 2	Consensus	Rater 1	Rater 2	Consensus	Rater 1	Rater 2	Consensus	Rater 1	Rater 2	Consensus	Rater 1	Rater 2	Consensus	Rater 1	Rater 2	Consensus	Rater 1	Rater 2	Consensus	Rater 1	Rater 2	Consensus
14	Was a cognitive interview study or other pilot test performed? (V OR I) if NO, skip items 15-35																																			
15	General design requirements																																			
15	Was the cognitive interview study or other pilot test performed in a sample representing the target population?																																			
16	Comprehensibility																																			
16	Were patients asked about the comprehensibility of the PROM? (if NO or not clear, skip items 17-25)																																			
17	Were all items tested in their final form?																																			
18	Was an appropriate qualitative method used to assess the comprehensibility of the PROM instructions, items, response options and recall period?																																			
19	Was each item tested in an appropriate number of patients?																																			
20	Were skilled interviewers used?																																			
21	Were the interviews based on an appropriate interview guide?																																			
22	Were the interviews recorded and transcribed verbatim?																																			
23	Was an appropriate approach used to analyse the data?																																			
24	Were at least two researchers involved in the analysis?																																			
25	Were problems regarding the comprehensibility of the PROM instructions, items, response options and recall period appropriately addressed by adapting the PROM?																																			
	SUBTOTAL QUALITY OF COMPREHENSIBILITY STUDY (lowest score of items 15-25)																																			
	Comprehensiveness																																			
26	Were patients asked about the comprehensiveness of the PROM? (if NO or not clear, skip items 27-35)																																			
27	Was the final set of items tested?																																			
28	Was an appropriate method used for assessing the comprehensiveness of the PROM?																																			
29	Was each item tested in an appropriate number of patients?																																			
30	Were skilled interviewers used?																																			
31	Were the interviews based on an appropriate interview guide?																																			
32	Were the interviews recorded and transcribed verbatim?																																			
33	Was an appropriate approach used to analyse the data?																																			
34	Were at least two researchers involved in the analysis?																																			
35	Were problems regarding the comprehensiveness of the PROM appropriately addressed by adapting the PROM?																																			
	SUBTOTAL QUALITY OF COMPREHENSIVENESS STUDY (lowest score of items 25, 26-35)																																			
	TOTAL QUALITY OF THE PILOT STUDY (lowest score of items 14-35)																																			
	TOTAL QUALITY OF THE PROM DEVELOPMENT STUDY (lowest score of items 1-35)																																			