
Article:
Statham, V., Emerson, L.-M. and Rowse, G. orcid.org/0000-0003-3292-4008 (2019) A systematic review of self-report measures of paranoia. Psychological Assessment, 31 (2). pp. 139-158. ISSN 1040-3590

https://doi.org/10.1037/pas0000645

Appendix B

Methodological appraisal scoring

Table A1

Summary of COSMIN criteria adapted from

Internal Consistency	Excellent	Good	Fair	Poor
Design requirements				
1. Was the percentage of missing items given?	10,22	1,2,3,4,5,6,7,8,9,16, 21,24,25,26		
2. Was there a description of how missing items were handled?	22,25	10	1,2,3,4,5,6,7,8,9,16, 24,26,21	
3. Was the sample size included in the internal consistency analysis adequate?	1,3,4,5,8,9(non-clinical),10,16,21,22, 25,24,26 (non-clinical)	7(non-clinical),9(clinical),25	2,6,26(clinical)	7(clinical)
4. Was the unidimensionality of the scale checked? Using factor analysis or Item Response Theory?	1,4,5,8,10, 24,25,26	16		2,3,6,7,9,21,22
5. Was the sample size included in the unidimensionality analysis adequate?	1,4,5,8,10,16, 24			25, 26
6. Was internal consistency calculated for each (unidimensional) (sub)scale separately?	1,4,5,8,10,16,24,25 26			2,3,6,7,9,21,22
7. Were there any important flaws in the design of methods of the study?	1,2,3,4,5,7,8,9,10,16, 21,22,24,25,26		6	
Statistical methods				
8. Was Cronbach's alpha calculated?	1,2,3,4,5,6,7,8,9,10 (persecution),16,21,22, 24,25,26			10(deservedness)

(continued)

Reliability (test-retest)	**Excellent**	**Good**	**Fair**	**Poor**
1. Was the percentage of missing items given?		1,5,6,9,21,25		
2. Was there a description of how missing items were handled?	25		1,5,6,9,21	
3. Was the sample size used in the analysis adequate?	1,25	5	9,21	6
4. Were at least 2 measurements available?	1,5,6,9,21,25			
5. Were the administrations independent?		1,5,6,9,21,25		
6. Was the time interval stated?	1,5,6,9,25		21	
7. Were patients stable in the interim period on the construct to be measured?		1,5,6,9,21,25		
8. Was the time interval appropriate?	25		1,5,6,9	
9. Were the test conditions similar for both measurements? E.g. type of administration, environment, instructions		1,6,9,21,25	5	
10. Were there any important flaws in the design or methods of the study?	1,5,9,21,25		6	
Statistical methods				
11. Was an intraclass correlation coefficient calculated?	9,21,25		1,5,6	
Content Validity [a]	**Excellent**	**Good**	**Fair**	**Poor**
1. Was there an assessment of whether all items refer to relevant aspects of the construct to be measured?	7,11,13,25	1,5,10,21,24		22
2. Was there an assessment of whether all items are relevant for the study population? (age, gender, country, setting) i.e. by piloting items	1,5,26			2,3,4,6,7,9,10,11, 12,13,14,15,20,22, 21,24,25
3. Was there an assessment of whether all items are relevant for the purpose of the measurement instrument? (discriminative, evaluative, and/or predictive)	7,11,13,21	1,5,10,22,25		24

(continued)

Content Validity [a]	**Excellent**	**Good**	**Fair**	**Poor**
4. Was there an assessment of whether all items together comprehensively reflect the construct to be measured?	1,24,25		5,7,10,11,13,22	
5. Were there any important flaws in the design of the study?	1,2,3,4,5,7,9,10,11,12, 13,14,15,22,21,24,25, 26	21	6,20	
Structural Validity	**Excellent**	**Good**	**Fair**	**Poor**
1. Was the percentage of missing items given?	10	1,4,5,8,11/13,12,14,15, 16,20,23,24,25,26		
2. Was there a description of how missing items were handled?	11/13, 25	10	1,4,5,8,12,14,15,16, 20, 23,24,26	
3. Was the sample size used in the analysis adequate?	1,4,5,8,10,12,14,15,16, 20,23, 24	11/13		25,26
4. Were there any important flaws in the design of the study	1,4,5,8,10,11/13,12,14,15 16, 23,24,25,26		20	
Statistical methods				
5. Was exploratory or confirmatory factor analyses performed?	1,4,5,8,10,11/13,12,14,15 16,20, 23,24,25,26			
Hypothesis testing (convergent & divergent validity)	**Excellent**	**Good**	**Fair**	**Poor**
1. Was the percentage of missing items given?	10,22	1,2,3,6,7,9,18,21,24,25, 26		
2. Was there a description of how missing items were handled?	22,25	10	1,2,3,6,7,9,21,24,26, 18	
3. Was the sample size used in the analysis adequate?	1,3,9,10,21, 22,25(non-clinical),26(non-clinical)	6,7(non-clinical),25 (clinical),18	2,24,26(clinical)	7(clinical)
4. Were hypotheses regarding correlations or mean formulated a priori (i.e. before data collection)?		2	1,3,6,7,9,10,21,22,24, 25,26	
5. Was the expected direction of correlation or mean differences formulated a priori	1,2,24,25	3,6,7,9,10, 21,22,26,18		

Hypothesis testing (convergent & divergent validity)	**Excellent**	**Good**	**Fair**	**Poor**
6. Was the expected absolute or relative magnitude of correlations or mean differences included in the hypotheses?		1,2,3,6,7,9,10, 21,22, 25,26,18		
7. Convergent validity: adequate description of the comparator instrument?	2,3,9,10,21,24,25,26,18		1,7,22	6
8. Convergent validity: were measurement properties of the comparator instrument adequately described?	18	3,10, 25,26	2,6,7,21,22,24	1,9
9. Were there any important flaws in the design of the study?	1,2,3,7,9,10,22,24,25, 26		6	
Statistical methods				
10. Were design and statistical methods adequate for the hypotheses to be tested?		1,2,3,6,7,9,10,21, 22,24,25,26,18		
Cross-cultural validity [a]	**Excellent**	**Good**	**Fair**	**Poor**
1. Was the percentage of missing items given?	10	2,3,4,9,12,14,15,17,18, 19,20,23,24,26		
2. Was there a description of how missing items were handled?		10	2,3,4,9,12,14,15,17, 18,19,20, 23,24,26	
3. Was the sample size used in the analysis adequate?	3,4,9,10,12,14,15,17, 18,19,20,23,24			2,26
4. Were both the original language in which the HR-PRO instrument was developed and the language in which the instrument was translated described?	2,4,9,10,12,14,15,16, 17,18,19,20,23,26			
5. Was the expertise of people involved in translation adequately described? E.g. expertise in the disease, expertise in the construct, expertise in the language		10,20,26	9,12,14,15	
6. Did the translators work independently from each other?	10,26	9,12,14,20	15	
7. Were items translated backward and forward?	26	14	9,12,14,15,20	
8. Was there adequate description of how differences between original and translated versions were resolved?		9,10,12,14,15,20,26		

(continued)

Cross-cultural validity [a]	Excellent	Good	Fair	Poor
9. Was the translation reviewed by a committee? [a]		9,10,12,14,15,20,26		
10. Was the instrument pre-tested to check interpretation, cultural relevance of the translation, and ease of comprehension? [a]	26			9,10,12,14,15,20
11. Was the sample used in the pre-test adequately described?			26	
12. Were all samples similar for all characteristics except for language and/or cultural background?			9,12,14,15,20,26	2,3,4,10,17
13. Were there any important flaws in the design of the study	2,3,4,9,10,12,14,15,17, 18,19,23,24,26		20	
Statistical methods				
14. Was confirmatory factor analyses performed?	4,10,12,14,15, 23,24,26			2,3,9,17,18,19,20

Responsiveness	Excellent	Good	Fair	Poor
1. Was the percentage of missing items given?		24,25		
2. Was there a description of how missing items were handled?	25		24	
3. Was the sample size used in the analysis adequate?	24		25	
4. Was a longitudinal design with at least 2 measurements used?	24,25			
5. Was the time interval stated?	25			24
6. If anything happened in the interim was this described?	24		25	
7. Did a proportion of the patients change?	24	25		
8. Were hypotheses about changes in scores formulated a priori (i.e. before data collection)?			24,25	
9. Was the expected direction of correlations or mean differences of the change scores of HR-PRO instruments included in these hypotheses?	24	25		

(continued)

Responsiveness	Excellent	Good	Fair	Poor
10. Were the expected absolute or relative magnitude of correlations or mean differences of the change scores of HR-PRO instruments included in these hypotheses?		25,24		
11. Was an adequate description provided of the comparator instrument(s)?	25			
12. Were the measurement properties of the comparator instrument(s) adequately described?				25
13. Were there any important flaws in the design or methods of the study?	24,25			
Statistical methods				
14. Were design and statistical methods adequate for the hypotheses to be tested?	24,25			
Design requirements for comparison to gold standard				
15. Can the criterion for change be considered as a reasonable gold standard?		25		
16. Were there any important flaws in the design or methods of the study?	25			
17. For continuous scores: were correlations between change scores, or the area under the Receiver Operator Curve (ROC) curve calculated?	25			
18. For dichotomous scales: Were sensitivity and specificity (changed versus not changed) determined?				

Note. [a] could not rate methodologies of some studies on items in these sections, as texts were not available in English; ratings based upon a subset of items. Item 2 of content validity section was rated when questionnaires were validated with new populations (e.g. new language, culture, clinical sample). Numbers are used within the table to refer to the different studies within the review. Table 2 within the main manuscript identifies the corresponding number for each study.