Online Appendix 4: Modified COSMIN with Scoring

Question	Scoring
Internal Consistency	
1. Was internal consistency assessed?	NS
2. Was Cronbach's alpha calculated? If no, what statistic was calculated to determine internal consistency?	Yes = 1 No = 0
3. Did the authors report whether any data as missing?	Yes = 1 No = 0
4. If data was missing did the authors describe how this was handled in the analysis?	NS
5. Was the total sample size $\geq$ 30?	Yes = 1 No = 0
6. If the sample size was less than 30, was a power analysis done to determine number of subjects needed?	NS
7. If a sample size power analysis was done, was the size requirement met?	Yes = 1 No = 0
8. What was the sample size?	NS
9. Input internal consistency value(s).	NS
10. Were there any important flaws in the design or methods of the study (for the internal consistency calculation)? If yes, please describe.	Yes = 0 No = 1
Maximum points for the section (Note: counted question 5 OR 7)	**4**
Reliability	
11. Did the study evaluate the reliability of the measure?	NS
12. Did the authors report whether any data as missing?	Yes = 1 No = 0
13. If data was missing did the authors describe how this was handled in the analysis?	NS
14. Was the total sample size $\geq$ 30?	Yes = 1 No = 0
15. If the sample size was less than 30, was a power analysis done to determine number of subjects needed?	NS
16. If a sample size power analysis was done, was the size requirement met?	Yes = 1 No = 0
17. Were at least two measurements taken and reported?	Yes = 1 No = 0
18. Was the time interval between measurements stated? If yes, what was the time interval?	Yes = 1 No = 0
19. Were the test conditions (e.g. type of administration, environment, instructions) similar for both measurements?	Yes = 1 No = 0

	Not stated = 0
20. Did the article state that each administration was done without knowledge of the other set of scores?	Yes = 1 No = 0
21. Are you reasonably confident that the patients were stable between test administrations? If no, please describe.	Yes = 1 No = 0
22. Were there any important flaws in the design or methods of the study that would affct the reliability analysis? If yes, please describe.	Yes = 0 No = 1
23. For ratio or interval level data, was an intraclass correlation coefficient calculated?	Yes = 1 No = 0 NA = 1
24. For dichotomous/nominal/ordinal scores, was kappa calculated?	Yes = 1 No = 0 NA = 1
25. For each type of reliability value calculated, enter the type and value.	NS
Maximum points for the section (Note: counted question 14 OR 16)	**10**
Responsiveness / Interpretability	
26. Were any of the following values assessed: standard error of measurement (SEM), smallest real difference (SRD), minimal detectable change (MDC), minimal important change (MIC), and minimal important difference (MID)? Questions 27 – 43 pertain to these psychometric variables.	NS
27. Did the authors report whether any data as missing?	Yes = 1 No = 0
28. If data was missing did the authors describe how this was handled in the analysis?	NS
29. Was the total sample size $\geq$ 30?	Yes = 1 No = 0
30. If the sample size was less than 30, was a power analysis done to determine number of subjects needed?	NS
31. If a sample size power analysis was done, was the size requirement met?	Yes = 1 No = 0
32. Were at least two measurements taken and reported?	Yes = 1 No = 0
33. Was the time interval stated?	Yes = 1 No = 0
34. Were the test conditions (e.g. type of administration, environment, and instructions) similar for both measurements?	Yes = 1 No = 0
35. Did the article state that each administration was done without knowledge of the other set of scores?	Yes = 1 No = 0
36. Are you reasonably confident that the patients were stable between test administrations?	Yes = 1 No = 0
37. Was more than one subgroup (e.g., diagnosis, condition, acuity or impairment level) included in the sample?	NS
38. Was data reported for each subgroup?	Yes = 1 No = 0

	NA = 1
39. Which of the following values were assessed? (Listed all values) Report where the values are found in the manuscript.	NS
40. Was the distribution of the (total) scores in the study sample described?	Yes = 1 No = 0
41. Were there any important flaws related to interpretability or responsiveness in the design or methods of the study? If yes, please describe the flaws.	Yes = 0 No = 1
42. If a longitudinal design was used, was the time interval stated? If yes, please report the interval.	Yes = 1 No = 0 NA = 1
43. If a longitudinal design was used, if anything occurred in the interim period (e.g. intervention, other relevant events), was it adequately described?	Yes = 1 No = 0 NA = 1
Maximum points for the section questions 27 - 43 (Note: counted question 29 OR 31)	**12**
44. Were any of the following values assessed: MCID, floor effect, and ceiling effect? Questions 45 – 55 pertain to these psychometric variables.	
45. Was more than one subgroup (e.g., diagnosis, condition, acuity or impairment level) included in the sample?	NS
46. Was data reported for each subgroup?	Yes = 1 No = 0 NA = 1
47. Which of the following values were assessed? (Listed all values) Report where the values are found in the manuscript.	NS
48. Was the distribution of the (total) scores in the study sample described?	Yes = 1 No = 0
49. Did the authors report whether any data as missing?	Yes = 1 No = 0
50. Was the total sample size $\geq$ 30?	Yes = 1 No = 0
51. If the sample size was less than 30, was a power analysis done to determine number of subjects needed?	NS
52. If a sample size power analysis was done, was the size requirement met?	Yes = 1 No = 0
53. What was the sample size?	NS
54. If a longitudinal design was used, was the time interval stated?	Yes = 1 No = 0 NA = 1
55. If a longitudinal design was used, if anything occurred in the interim period (e.g. intervention, other relevant events), was it adequately described?	Yes = 1 No = 0 NA = 1
Maximum points for the section – questions 45 - 55 (Note: counted question 50 OR 52)	**6**
Generalizability	
56. Was the sample in which the instrument was evaluated adequately described in terms of subject disease/condition or normative sample (select as many as apply): brain injury,	NS

multiple sclerosis, Parkinson's disease, spinal cord injury, stroke, vestibular dysfunction, normative sample, other (please describe).	
57. What best describes the patient population studied: Acute/Subacute; Chronic Progressive; Chronic Stable? (select as many as apply)	NS
58. Was the setting(s) in which the study took place (e.g., acute care, in-patient rehabilitation, out-patient, community) described. If yes, please describe.	Yes = 1 No = 0
59. Was the age of the subject reported (median or mean age, with standard deviation or range) described? If yes, please describe.	Yes = 1 No = 0
60. Was the distribution of sex described? If yes, please describe (e.g., 45% male; 55% female).	Yes = 1 No = 0
Maximum points for the section	**3**
General Methodology*	
61. Was the method (e.g. convenience, consecutive, or random) used to select patients adequately described?	Yes = 1 No = 0
62. Was the total sample size $\geq$ 30?	Yes = 1 No = 0
63. If the sample size was less than 30, was a power analysis done to determine number of subjects needed?	NS
64. If a sample size power analysis was done, was the size requirement met?	Yes = 1 No = 0
65. What was the sample size?	NS
66. If more than one group was included in the sample, was data reported by subgroups?	Yes = 1 No = 0 NA = 1
67. Did the authors report whether any data as missing?	Yes = 1 No = 0
68. If data was missing did the authors describe how this was handled in the analysis?	NS
69. Was there a description of how the raters were training? If yes, please describe.	Yes = 1 No = 0
70. Were the raters (select as many as apply): research scientists, research clinicians, research assistants, practicing clinicians, students, not reported, other (please describe).	NS
71. Is there anything else you would like to tell us about this study?	NS
Maximum points for the section (Note: counted question 50 OR 52)	**5**

NA – not applicable; NS – not scored

*Questions in the general methodology section were asked only once to avoid redundancy. After CATs were scored, the methodology questions were added to the relevant boxes to calculate a total score for each section.