

Supplemental Material S1. Methodological quality rating scales.

PEDro*		NIH Pre-Post*		SCED*	
Category	Item	Category	Item	Category	Item
Eligibility	1. Criteria specified (not included in score)	Study question	1. Clearly specified	Clinical history	1. Age, sex, etiology, and severity specified (not included in score)
Participants	2. Random allocation	Eligibility	2. Criteria specified	Target behaviors	2. Precise and repeatable measures specified
	3. Concealed allocation	Participants	3. Representative of clinical population of interest	Design	3. Must be A-B-A or multiple baselines
Intervention	4. Similar groups at baseline on prognostic indicators		4. All meeting eligibility criteria enrolled		4. Sufficient baseline sampling
Blinding	5. Blinded participants	Sample size	5. Sufficiently large (adequate statistical power)		5. Sufficient treatment phase sampling
	6. Blinded therapists	Intervention	6. Clearly specified		6. Raw data points reported
	7. Blinded assessors	Outcome measures	7. Clearly specified, valid, reliable and assessed consistently across participants	Observer bias	7. Inter-rater reliability for at least target behavior measure
Outcome measures	8. Obtained for >85% of group participants for at least one key outcome	Blinding	8. Blinded assessors	Assessors	8. Independence of assessors from therapists
	9. Analysis by intention to treat	Follow-up rate	9. ≤ 20% loss to follow-up after baseline and loss to follow-up accounted for in analysis	Analyses	9. Use of statistical analyses
	10. Between-group statistical comparison for at least one key outcome	Statistical analysis	10. Use of statistical tests to assess significance of pre-post changes on outcomes	Replication	10. Either across participants, therapists or settings
Treatment effect & variability	11. Point measures and measures of variability	Outcome measures	11. Multiple measurements before and after intervention	Generalization	11. Evidence for generalization
		Group level intervention**	12. Individual-level outcome efforts accounted for in group level intervention		

*PEDro: Physiotherapy Evidence Database scale (Maher et al., 2003); NIH Pre-Post: NIH Quality Assessment Tool for Before-After (Pre-Post) Studies with No Control Group (National Heart, Lung, and Blood Institute); SCED: Single-Case Experimental Design (Tate et al., 2008).

**No studies in this review provided group level intervention. Therefore, the maximum score for any cohort study on the NIH Pre-Post assessment tool was 11.