Appendix 2. What Works Centre for Wellbeing Quality Checklist: Quantitative evidence of intervention effectiveness

The checklist below is from the quality checklist for quantitative evidence of intervention effectiveness. In a previous review, WWCW developed a scoring system to provide an indication of overall level of confidence in the design, conduct and reporting of the study. The 10 elements of the checklist can be scored either 1 (yes) or 0 (no, can't tell or N/A). The total score can be used to assign each study an overall level of confidence of low (0-2), moderate (3-6) or high (7-10).

Question	Element	Response options
Was the evidence well-designed?	Fidelity: • The extent to which the intervention was delivered with fidelity is clear – i.e., if there is a specific intervention which is being evaluated, this has been well reproduced.	Yes (1) No (0) Can't tell (0) N/A (0)
	Measurement: • The measures are appropriate for the intervention's anticipated outcomes and population. • Participants completed the same set of measures once shortly before participating in the intervention and once again immediately afterwards. • An 'intent-to-treat' design was used, meaning that all participants recruited to the intervention participated in the pre/post measurement, regardless of whether or how much of the intervention they received, even if they dropped out of the intervention (this does not include dropping out of the study - which may then be regarded as missing data).	Yes (1) No (0) Can't tell (0) N/A (0)

Question	Element	Response options
	Counterfactual: • Assignment to the treatment and comparison group was at the appropriate level (e.g., individual, family, school, community). • The comparison condition provides an appropriate counterfactual to the treatment group. Consider: • Participants were randomly assigned to the treatment and control group through the use of methods appropriate for the circumstances and target population OR sufficiently rigorous quasi-experimental methods (regression discontinuity, propensity score matching) were used to generate an appropriately comparable sample through non-random methods. • The treatment and comparison conditions are thoroughly described.	Yes (1) No (0) Can't tell (0) N/A (0)
Was the study carried out appropriately? Including appropriate sample	Representative: • The sample is representative of the intervention's target population in terms of age, demographics and level of need. The sample characteristics are clearly stated. • There is baseline equivalence between the treatment and comparison group participants on key demographic variables of interest to the study and baseline measures of outcomes (when feasible).	Yes (1) No (0) Can't tell (0) N/A (0)
	Sample size: • The sample size is sufficiently large to test for the desired impact. This depends most importantly on the effect size, however a suggestion could be, for example, that a minimum of 20 participants have completed the measures at both time points within each study group.	Yes (1) No (0) Can't tell (0) N/A (0)
	Attrition: • A minimum of 35% of the participants completed pre/post measures. Overall study attrition is not higher than 65%. • The study had clear processes for determining and reporting drop-out and dose. Differences between study drop-outs and	Yes (1) No (0) Can't tell (0) N/A (0)

Question	Element	Response options
	completers were reported if attrition was greater than 10%. The study assessed and reported on overall and differential attrition.	
	Equivalence: • Risks for contamination of the comparison group and other confounding factors have been taken into account and controlled for in the analysis if possible. • Participants were blind as to their assignment to the treatment and comparison group. There was consistent and equivalent measurement of the treatment and control groups at all points when measurement took place.	Yes (1) No (0) Can't tell (0) N/A (0)
	Measures: • The measures used were valid and reliable. This means that the measure was standardised and validated independently of the study, and that the methods for standardisation were published. Administrative data and observational measures may also have been used to measure programme impact, but sufficient • Information was given to determine their validity for doing this. • Measurement was independent of any measures used as part of the treatment. • In addition to any self-reported data (collected through the use of validated instruments), the study also included assessment information independent of the study participants (e.g., an independent observer, administrative data etc)	Yes (1) No (0) Can't tell (0) N/A (0)
Was the analysis appropriate?	• The methods used to analyse results are appropriate given the data being analysed (categorical, ordinal/ratio, parametric/non-parametric, etc.) and the purpose of the analysis. • Appropriate methods have been used and reported for the treatment of missing data.	Yes (1) No (0) Can't tell (0) N/A (0)

Question	Element	Response options
Is the evidence consistent?	• Are the findings made explicit? • Is there adequate discussion of the evidence both for and against the researcher's arguments? • Has the researcher discussed the credibility of their findings (e.g., triangulation, respondent validation, more than one analyst)? • Are the findings discussed in relation to the original research question?	Yes (1) No (0) Can't tell (0) N/A (0)