



4	<p>Effectiveness: A programme with evidence from at least two high-quality evaluations (RCT/QED) demonstrating positive impacts across populations and environments lasting a year or longer. The evidence may include significant adaptations to meet the needs of different target populations.</p> <p>The evidence must meet the following requirements:</p> <ul style="list-style-type: none">• The intervention has demonstrated consistent significant positive child outcomes in two rigorous evaluations (RCT/QED) meeting all criteria required for Level 3.• At least one evaluation uses a form of measurement that is independent of the study participants (and also independent of those who deliver the programme). In other words, self-reports (through the use of validated instruments) might be used, but there is also assessment information independent of the study participants (e.g. an independent observer, administrative data, etc.)• There is evidence of a long-term outcome of 12 months or more from at least one of these studies. <p>To achieve a 4+ rating:</p> <ul style="list-style-type: none">• All of the criteria for level 4 must be met.• At least one of the effectiveness evaluations will have been conducted independently of the programme developer.• The intervention must have evidence of improving EIF child outcomes from three or more high-quality evaluations (RCT/QED) conducted within real world settings.
3	<p>Efficacy: A programme with evidence from at least one rigorously conducted RCT/QED demonstrating a statistically significant positive impact on at least one child outcome.</p> <p>The evidence must meet the following requirements:</p> <ul style="list-style-type: none">• The evaluation must meet the requirements for a Level 2.• Participants are randomly assigned to the treatment and control groups through the use of methods appropriate for the circumstances and target population, OR sufficiently rigorous quasi-experimental methods (e.g. regression discontinuity, propensity score matching) are used to generate an appropriately comparable sample through non-random methods.• Assignment to the treatment and comparison group is at the appropriate level (e.g. individual, family, school, community).• An ‘intent-to-treat’ design is used, meaning that all participants recruited to the intervention participate in the pre/post measurement, regardless of whether or how much of the intervention they receive, even if they drop out of the intervention (this does not include dropping out of the study – which is then regarded as missing data).• The treatment and comparison conditions are thoroughly described.• The intervention is delivered with acceptable levels of fidelity in the evaluation study.• The comparison condition provides an appropriate counterfactual to the treatment group.• 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).

- Risks for contamination of the comparison group and other confounding factors are taken into account and controlled for in the analysis if possible.
- Participants are blind to their assignment to the treatment or comparison group. (Only a binding criteria if feasible.)
- The study should report on overall and differential attrition (or clearly present sample size information such that this can be readily calculated).
- If overall study attrition is greater than 10%, then study authors must report differences between the study drop-outs and completers, as well as perform analyses demonstrating that study attrition did not undermine the equivalence of the study groups (and adjusting for this if differences are identified).
- Measurement is blind to group assignment.
- There is consistent and equivalent measurement of the treatment and control groups at all points when measurement takes place.
- Statistical models control for baseline differences between the treatment and comparison groups in outcome measures and demographic characteristics that might be apparent after recruitment.
- The treatment condition is modelled at the level of assignment (or deviations from that strategy are justified statistically).
- Appropriate methods are used and reported for the treatment of missing data.
- The findings are of sufficient magnitude to justify further analysis. (Not yet assessed in pure cost-effectiveness terms.)

To achieve a **3+** rating:

- The programme will have obtained evidence of a significant positive child outcome through an efficacy study, but may also have additional consistent positive evidence from other evaluations (occurring under ideal circumstances or real world settings) that do not meet this criteria, thus keeping it from receiving an assessment of 4 or higher.

2 Preliminary: A programme with evidence of improving a child outcome from a study involving at least 20 participants, representing 60% of the sample using validated instruments.

The evidence must meet the following requirements:

- Participants complete the same set of measures once shortly before participating in the programme and once again immediately afterwards.
- 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.
- The sample is sufficiently large to test for the desired impact. A minimum of 20 participants complete the measures at both time points within each study group (e.g. a minimum of 20 participants in pre/post study not involving a comparison group or a minimum of 20 participants in the treatment group AND comparison group).
- The study has clear processes for determining and reporting drop-out and dose.
- For pre/post studies, overall study attrition is not higher than 40% (with at least 60% of the sample retained). For comparison group studies, overall study attrition is not higher than 65% (with at least 35% of the sample retained).
- The measures are appropriate for the intervention's anticipated outcomes and population.
- The measures are valid and reliable. This means that the measures are standardised and validated independently of the study and the methods for standardisation are published. Administrative data and observational measures might also be used to measure programme impact, but there is sufficient information to determine their validity for doing this.
- Measurement is independent of any measures used as part of the treatment.

	<ul style="list-style-type: none"> • The methods used to analyse results are appropriate given the data being analysed (categorical, ordinal, ratio/parametric or non-parametric, etc.) and the purpose of the analysis. • There are no harmful effects. • There is evidence of a statistically significant positive impact ($p < .05$) on at least one EIF outcome. • The intervention's model clearly identifies and justifies its primary and secondary outcomes and there is a statistically significant main effect of improving at least one or more of these outcomes, depending on the number of outcomes measured. • There is consistency amongst the findings, resulting in few mixed results within the study. • Subgroup analysis is used to verify for whom the intervention is effective and the conditions under which the effectiveness is found. (Statistically significant findings within subgroups are not treated as a replacement for a main effect.) <p>To achieve a 2+ rating:</p> <ul style="list-style-type: none"> • The programme will have observed a significant positive child outcome in an evaluation meeting all of the criteria for a Level 2 evaluation, but also involving a treatment and comparison group. • 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).
NL2	<p>No evidence at Level 2: Programmes that do not yet have Level 2 evidence for a variety of methodological reasons.</p> <p>Programmes falling into this category are typically at earlier stages of their development, doing important foundational work.</p> <p>This work includes confirmation of the programme's core assumptions and logic model and verification of its primary child outcomes:</p> <ul style="list-style-type: none"> • Confirmation. The programme is confirming key elements of its logic model and their grounding in scientific evidence. • Verification. Key elements of the logic model have been confirmed and their feasibility for improving child outcomes is being verified.
NE	<p>A programme with evidence from at least one rigorously conducted RCT/QED that is also the most rigorous impact evaluation demonstrating no effect on child outcomes.</p> <p>The evidence must meet the following requirements:</p> <ul style="list-style-type: none"> • The evaluation must meet the requirements for a Level 3. • However, the evaluation will fail to demonstrate any statistically significant positive benefits for parents or children. <p>In these instances, a rating of 'no effect' (NE) is applied to suggest that a rigorously conducted evaluation has failed to confirm positive benefits for parents or children.</p>