What is 'good evidence'?

What do EIF evidence ratings mean?

EIF recognises several levels of strength of evidence. This is not a rating of the scale of impact but of the degree to which a programme has been shown to have a positive, causal impact on specific child outcomes.

- **Level 4** recognises programmes with evidence of a **long-term positive impact through multiple rigorous evaluations**. At least one of these studies must have evidence of improving a child outcome lasting a year or longer.
- **Level 3** recognises programmes with evidence of a **short-term positive impact from at least one rigorous evaluation** – that is, where a judgment about causality can be made.
- **Level 2** recognises programmes with **preliminary evidence of improving a child outcome**, but where an assumption of causal impact cannot be drawn.
- **NL2** – 'not level 2' – distinguishes programmes for which the most robust evaluation evidence does not meet the level 2 threshold for a child outcome.
- **NE** – 'found not to be effective in at least one rigorously conducted study' — is reserved for programmes where a rigorous programme evaluation (equivalent to a level 3) has found no evidence of improving one of our child outcomes or providing significant benefits to other participants. This rating should not be interpreted to mean that the programme will never work, but it does suggest that the programme will need to adapt and improve its model, learning from the evaluation.

For further information please visit the EIF Guidebook.

What makes evidence rigorous?

There are many factors that makes a study rigorous. The following checklist presents some key factors to look for in judging whether studies provides good evidence of impact.

Features of good evidence	Explanation	Checklist
Validated measurement	If a measurement is 'validated' it means it has been shown to reliably measure what it is designed to measure. This is important so that you can determine if an improvement in outcomes for those receiving a treatment represents a real improvement for those people.	✓/✗
Representative sample	It is important for the subjects in an evaluation to be similar in key ways to the real world in which a programme is implemented. If a study's sample is highly specific and unlikely to be found outside the context of the study, you might not be able to generalise the results to future implementations of the programme.	✓/✗
Statistically significant effect	If the outcomes of the people receiving an intervention have improved, it is important to determine if the effect was not simply due to chance. If an effect is statistically significant at the 95% level and above, this means it is unlikely that the difference in outcomes was due to random variation.	✓/✗

Adequate counterfactual	It is often not enough to observe that the outcomes of people receiving the treatment have improved – there may be that there are factors other than the treatment that caused the improvement. Therefore it is important that a study provides an adequate counterfactual to show that the improvement would not have occurred if people hadn't received the treatment. If it shows this then you can conclude the treatment caused the improvement.	✓/✗
Sample size	It is important that the sample size of a study is not too small. One of the reasons for this is that a study may not be able to detect an effect if it does not have enough study participants. EIF requires that a study should have at least 20 participants at each measurement point in order for the study to provide at least preliminary evidence of impact (level 2).	✓/✗
Validated measurement	If a measurement is 'validated' it means it has been shown to reliably measure what it is designed to measure. This is important so that you can determine if an improvement in outcomes for those receiving a treatment represents a real improvement for those people.	✓/✗