Handbook

GUIDEBOOK HANDBOOK V 1.0

Procedures and standards for assessment of evidence for interventions

Applicable to interventions added to the Guidebook **before April 2026**

Foundations
What Works Centre for Children & Families
foundations.org.uk

INTRODUCTION

The Foundations Guidebook is a free online resource that contains evidence on what works to support children, young people, and families.

This technical guide provides more detail on the evidence standards and processes underpinning the Guidebook, as well as further information about the different content available.

Finding out what works best for children and families is not easy. There is a lot of evidence available, but it can be hard to know how reliable it is. This means it can be difficult to decide on the right kind of support for children and families.

The Guidebook aims to change this. It is free and easy to access. It is designed to help local leaders, commissioners and practitioners, and researchers and policymakers, to use evidence when they make decisions about how best to support children and families.

—

Visit the Foundations website: foundations.org.uk
Visit the Foundations Guidebook: foundations.org.uk/toolkit/guidebook

This Handbook *(Guidebook Handbook v 1.0* – previously the *Technical Guide*) reflects the processes and standards by which **interventions published before April 2026** were assessed.

Please see *Guidebook Handbook v 2.0* (2026) for information on the **revised** processes and standards **applicable from April 2026**: https://foundations.org.uk/wp-content/uploads/2026/04/guidebook-handbook-v2.0-2026-procedures-and-standards.pdf

CONTENTS

About the Guidebook

What is the Foundations Guidebook?

The Foundations Guidebook is a new, updated version of the Early Intervention Foundation's (EIF) Guidebook.[1]

It contains information on existing research on more than 130 interventions with at least preliminary evidence of improving children and families' outcomes.

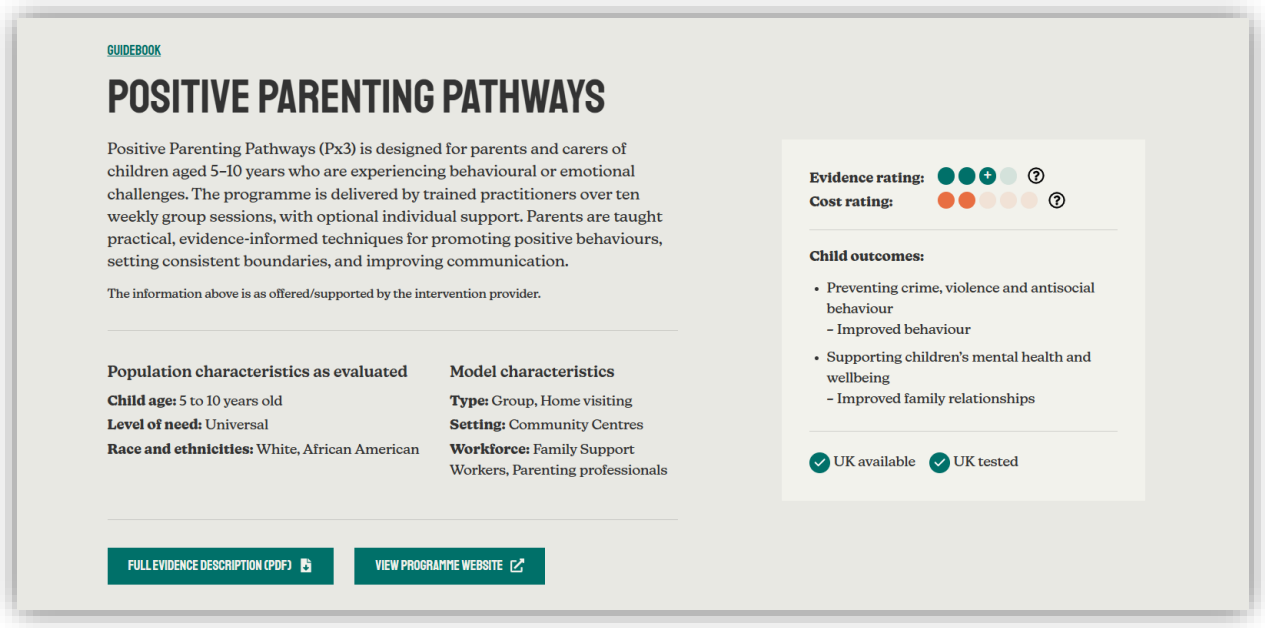
The Guidebook provides robust evidence to support decision-making across the spectrum of early intervention, services for children and families, and children's social care.

What is the Foundations Guidebook's approach to race and ethnicity?

The Guidebook aims to present information about race and ethnicity transparently and sensitively, reflecting how these characteristics are reported in the studies underpinning the Guidebook evidence rating, while aligning with current best practice on reporting and language.

In the intervention summary (shown on the following page), all racial and ethnic terms describing the participants in studies are included under the heading 'Race and ethnicities'. These are listed alphabetically by group. The Guidebook preserves the terms used in studies wherever possible, whether broad (e.g. 'Black') or specific (e.g. 'African American'), except in cases where terms are now outdated or offensive. In such cases, terms are translated to align with government and institutional style guidance (e.g. the UK Office for National Statistics' race and ethnicity categories, US, Australian, and New Zealand governmental guidelines, and Foundations' own internal style guide).

[1] Foundations was created from the merger of EIF and What Works for Children's Social Care in December 2022.


GUIDEBOOK

POSITIVE PARENTING PATHWAYS

Positive Parenting Pathways (Px3) is designed for parents and carers of children aged 5–10 years who are experiencing behavioural or emotional challenges. The programme is delivered by trained practitioners over ten weekly group sessions, with optional individual support. Parents are taught practical, evidence-informed techniques for promoting positive behaviours, setting consistent boundaries, and improving communication.

The information above is as offered/supported by the intervention provider.

Population characteristics as evaluated

Child age: 5 to 10 years old
Level of need: Universal
Race and ethnicities: White, African American

Model characteristics

Type: Group, Home visiting
Setting: Community Centres
Workforce: Family Support Workers, Parenting professionals

FULL EVIDENCE DESCRIPTION (PDF)
VIEW PROGRAMME WEBSITE

Evidence rating:
Cost rating:

Child outcomes:

- Preventing crime, violence and antisocial behaviour
  - Improved behaviour
- Supporting children's mental health and wellbeing
  - Improved family relationships

UK available
UK tested

Terms that indicate nationality (e.g. 'Finnish'), regions (e.g. 'Mediterranean'), or ambiguous cultural and linguistic descriptors are excluded from the intervention summary, particularly when study context does not clarify whether a term refers to ethnicity or nationality.

In study summaries, we list all demographic characteristics of study participants (including ethnicity, race, nationality, language, and related descriptors) as recorded in the studies, with percentages if available. The only changes made in this section are replacements of outdated or offensive language. Where studies do not report race or ethnicity, the guidebook will state this clearly (e.g. 'not reported').

Our approach ensures that evidence is presented in a way that is respectful, clear, and consistent, while supporting commissioners and local area leaders to understand who interventions have worked for, and where evidence may be lacking for specific populations. It is important to note, however, that the intervention may have been delivered and evaluated successfully with other populations; only those populations involved in the studies which underpin the Guidebook evidence rating are mentioned. Also, unless specified, the intervention may have been designed to be implemented with a wide variety of different communities.

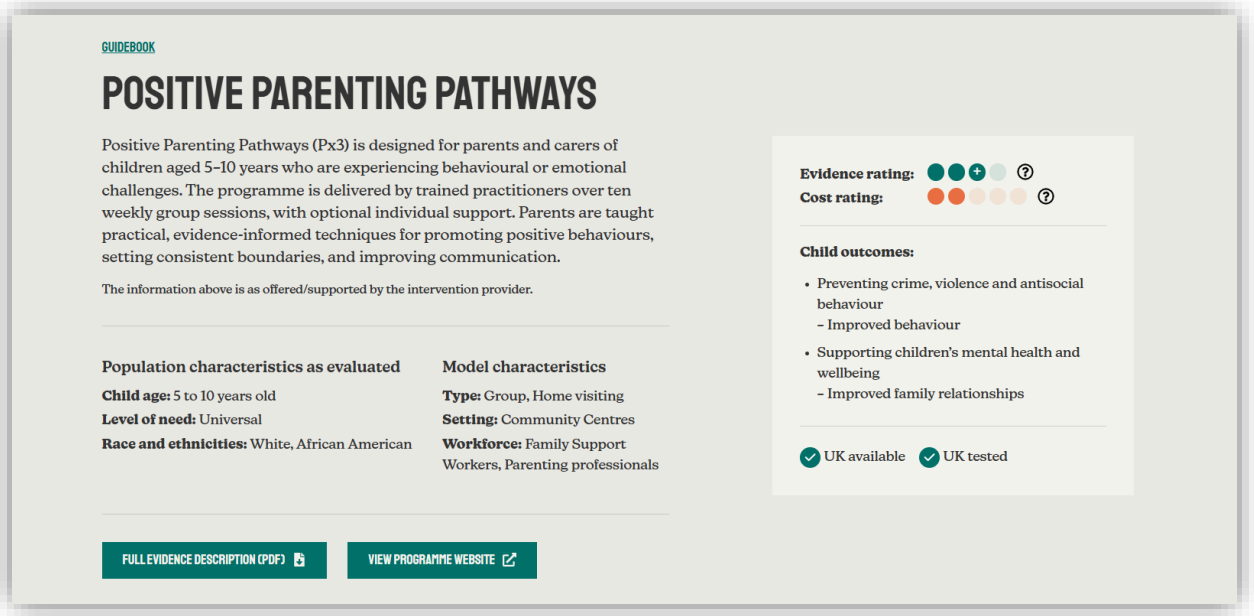
We recognise that this is just a starting point; our approach to race and ethnicity is still in development and will continue to evolve as we engage with new evidence, feedback, and best practice.

Which child outcomes and other outcomes appear on the Guidebook?

The Guidebook assesses the strength of evidence for interventions across a set of seven clearly defined child outcome domains. These reflect key priorities for improving the lives of children and young people:

1. Supporting children's mental health and wellbeing
2. Preventing child maltreatment
3. Enhancing school achievement and employment
4. Preventing crime, violence and antisocial behaviour
5. Preventing substance abuse
6. Preventing risky sexual behaviour and teen pregnancy
7. Preventing obesity and promoting physical healthy development.

Positive outcomes are grouped under these seven domains, for example: 'Supporting children's mental health and wellbeing' includes among other outcomes 'improved emotional wellbeing', 'improved social behaviour', and 'reduced anxiety'. These plain language outcomes are presented under the seven outcome domains in the intervention summary. They help users compare evidence across different interventions using a consistent framework.



The study summaries also include more detail, with the specific outcomes evidenced in each study. In the full evidence description, downloadable in a PDF, all measures used in a study and all outcomes – whether positive, null, or negative – are listed. This allows users to see the exact tools and methods used to evaluate the impact of each intervention.

Importantly, only child outcomes within the seven outcome domains are used to determine an intervention's evidence rating. These outcomes have been assessed to be reliable during the Guidebook's assessment process.

The Guidebook includes information about parent and other outcomes, such as improved positive parenting, reduced parental stress, improved parental mental health, or improved teaching strategies. Parent and other outcomes are listed in study summaries. They are only included when reported in studies which also included child outcomes and which have been assessed for the Guidebook. While parent and other outcomes can provide valuable context, they do not contribute to the evidence rating and have not been assessed with the same level of scrutiny. If a study only reports parent outcomes, with no accompanying child outcomes, it is not included on the Guidebook at all.

The terminology used to describe outcomes may vary between research and practice. In the intervention summary, outcomes are presented in terms more familiar to practitioners and local leaders. However, within the full evidence description, outcomes are described using the language found in the original studies. This may sometimes include terms that are outdated or considered no longer appropriate; these reflect the original wording by the study authors and do not represent Foundations' preferred terminology.

How are interventions selected for inclusion on the Guidebook?

Interventions are selected for inclusion on the Guidebook through a combination of approaches designed to identify interventions with the potential to improve outcomes for children and families. These have included themed assessment rounds focused on specific topics, such as socio-emotional learning and Early Years, open calls where intervention providers can submit their interventions for consideration, and in-house searches carried out by the Foundations Guidebook team. Our methods may evolve over time to reflect emerging priorities and sector needs.

We only include interventions on the Guidebook – manualised, repeatable activities, and models of delivery with a beginning and an end. Currently, we do not include system-level programmes or models, which may include ways of working or the offer of multiple interventions.

We assess both UK-based and international interventions. One of the Guidebook's goals is to identify promising interventions that could be potentially transported to the UK context, even if they are not currently available here.

What kinds of evidence are included on the Guidebook?

Our assessments focus on impact evaluations – studies that help us understand whether an intervention has had a measurable, positive effect on child outcomes. These typically include:

- Randomised controlled trials (RCTs)
- Quasi-experimental designs (QEDs)

- Some well-designed pre–post studies.

We do not include qualitative evidence as part of our formal assessment process, as this type of evidence does not allow us to draw strong conclusions about whether an intervention has caused change. Qualitative evidence is of course useful for other purposes, and we have an organisational commitment to undertake implementation and process evaluations as part of our own funded evaluations and to synthesise qualitative evidence to inform our Practice Guides. For the Foundations Guidebook, we base our assessments on the most robust and relevant impact evaluations available at the time. At the moment, we only include evidence from studies which took place in developing countries, because the findings from these are most relevant and applicable to the UK context.

How does the assessment process work?

We take a targeted and staged approach to ensure that we focus on the most relevant and high-quality evidence. This process has three main stages: searching, triage, and assessment.

Stage 1: Searching

After identifying an intervention to assess, we conduct comprehensive searches to identify relevant evaluation studies. This includes using Google Scholar, academic databases, and other clearinghouse websites. We look for studies that relate directly to the intervention in question.

Each study is initially screened to determine whether it meets the minimum criteria for inclusion. For example, we look for:

- Relevant child outcome measures
- A comparison group or pre–post design
- Sufficient sample size.

If a study meets the basic standards, it moves to the triage stage.

Stage 2: Triage

Triage involves a provisional review of the quality of each study. Reviewers read through each study and assign a screening rating based on our evidence criteria for each Guidebook rating level. This helps us to identify studies with the potential for higher-level ratings, rule out lower-quality studies that are unlikely to affect the intervention's overall rating, and prioritise studies to take to full assessment. This step ensures that we focus the detailed review process on the most promising studies.

Stage 3: Assessment

In the final stage, we conduct a full assessment of the highest-priority studies. This involves a detailed appraisal of the study design and findings, based on our evidence standards. It is conducted by the Foundations evidence synthesis team, with external review by a panel of experts.

The Guidebook rating is based on the most robust study or studies available. Our goal is to ensure that the rating reflects the best available evidence on impact, while also checking for any high-quality studies that may indicate no effect or mixed findings.

What are theories of change?

A theory of change describes the evidence-based assumptions behind an intervention, the need for the intervention, and how the intervention works to achieve its intended positive outcomes for children and young people. Theories of change are useful for those commissioning and implementing interventions, because they help achieve a shared understanding of the intervention, its aims and rationale, and point to the existing evidence base that underpins the need for and design of the intervention. They are useful for evaluators as they can shape what outcomes are being evaluated (the research questions) and provide explanations for any effects which are observed. Reflecting on a theory of change may also highlight gaps in assumptions or uncertainties in plans to implement an intervention.

On the Foundations Guidebook, an intervention's theory of change is based on information provided by the intervention developer or provider, together with descriptions of the intervention found in studies and on the intervention website. For parenting interventions' theories of change, interventions which are based on similar theories of child development or which work in a similar way are now described in a similar way too, making it easier to see how 'families' of interventions are similar or differ.

The updated Foundations Guidebook also makes the theory of change more transparent, by highlighting different aspects of it: the science-based assumption, who the intervention targets, what the intervention does, and what the short-, medium-, and long-term outcomes are intended to be. While the theories of change on the Guidebook include established scientific theories, this does not necessarily mean that the theories of change themselves have been rigorously tested for particular interventions.

What is new on the Foundations Guidebook?

Following the merger of the Early Intervention Foundation and What Works for Children's Social Care to form Foundations, the EIF Guidebook has been updated to become the Foundations Guidebook.

The Guidebook has a longstanding history of supporting evidence-informed decision-making for children and families. The original EIF Guidebook was launched in 2014 and included information on about 50 interventions available in the UK. At that time, the evidence ratings were based on reviews by other international evidence clearinghouses. From 2016, EIF carried out its own independent assessments, and any interventions from the original Guidebook that have since been reassessed by EIF are now included in the current Guidebook with updated information, provided they meet our minimum threshold for inclusion.

In 2022, a significant milestone was reached with the merger of EIF and What Works for Children's Social Care. This brought together two What Works centres to create Foundations: a single, unified What Works centre that covers the full spectrum of support – from early intervention and prevention to targeted services for families at risk of poor outcomes, and support for children with a social worker, children in care, and care leavers. The Foundations Guidebook is part of this broader vision, aiming to provide accessible and robust evidence to support better decision-making across early intervention and children's services.

The Foundations Guidebook is now more accessible and easier to search, making it easier to use for decision-makers, including commissioners, local area leaders, and national policymakers. We've improved the layout and search function so that you can quickly access details about individual interventions.

The majority of intervention entries are the same on the Foundations Guidebook as they were on the EIF Guidebook. This includes the evidence and cost ratings, improvement indices, studies which feature to underpin the evidence rating, and most of the implementation information. There are some areas where the content and how it is presented has changed, including:

- The full evidence description is a downloadable PDF. It provides more information on individual studies underpinning the intervention's evidence rating, including about the study design, sample retention, and full results of the study.
- Information on the ethnicity and race of participants in each study is included. In the intervention summary on the Foundation Guidebook website, only information about race and ethnicity are reported; in the full evidence description, information relating to race, ethnicity, nationality, language, and other characteristics are reported, following the descriptions in the studies.
- Parent outcomes which feature alongside child outcomes in studies contributing to the evidence rating are reported more consistently in the study summaries.
- The evidenced outcomes of interventions remain the same, except that the language used to describe these outcomes in the intervention summary on the Guidebook website has been made more accessible and consistent across interventions. Additionally, for some interventions, outcomes have been added or removed when they were accidentally omitted or included in previous assessment rounds.
- The theory of change is presented in a new easy-to-read format, and for parenting interventions these have been enriched with information from the intervention's studies and other information like the intervention provider's website.

Foundations' evidence standards

What are Foundations' evidence standards?

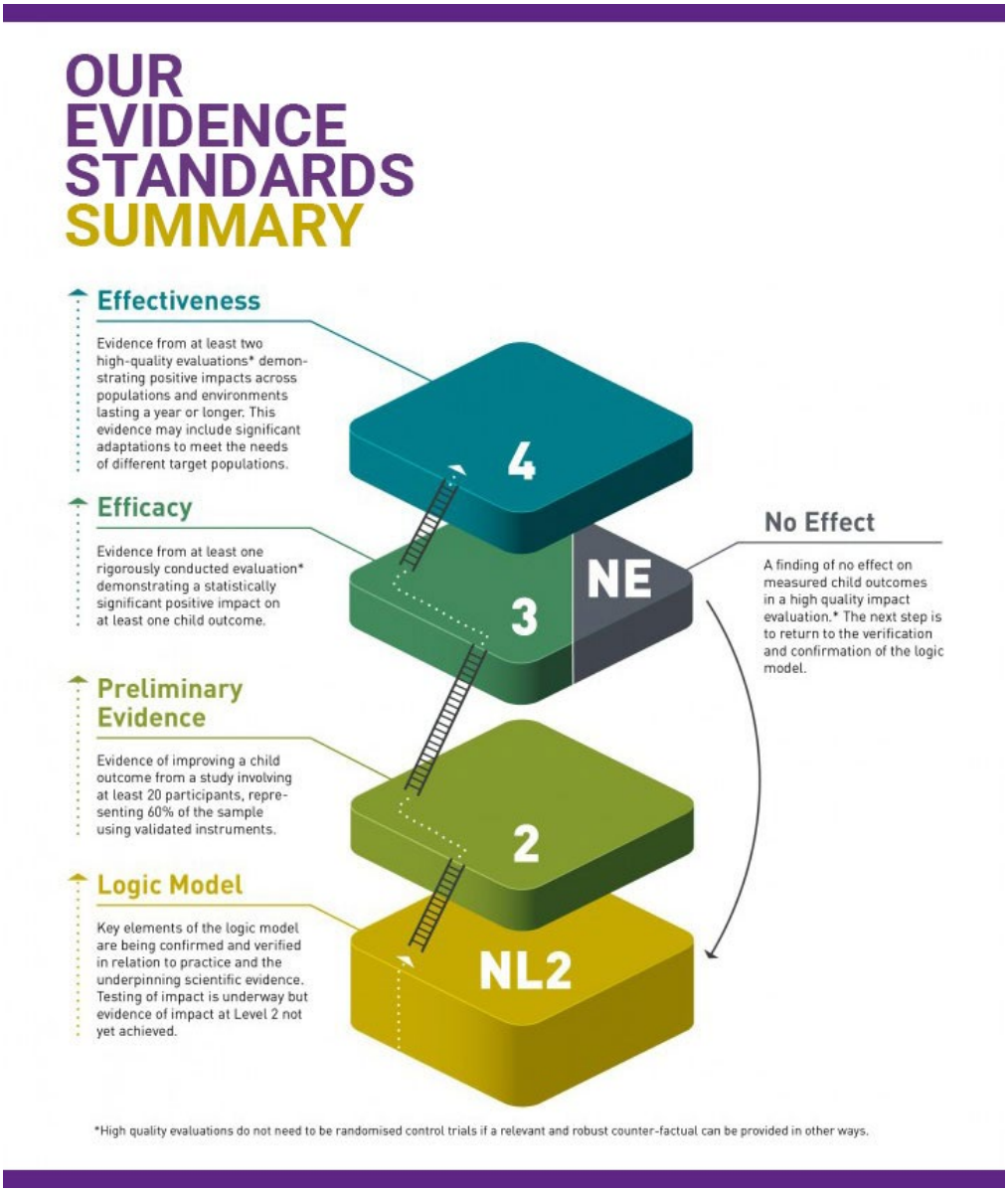
Foundations' evidence standards distinguish levels of the strength of evidence on a six-point scale: an intervention may be given a rating of Level 2, Level 3, or Level 4, with 'plus' levels providing

more incremental steps in strength of evidence (Levels 2+, 3+, and 4+). These levels provide an indication of how confident we can be that there is causal evidence – that an intervention caused an improvement in a child outcome. The standards encompass many features of a study's design, assessing its internal and external validity.

The term 'evidence-based' is frequently applied to interventions with a Level 3 evidence rating or higher, because this is the point at which there is sufficient confidence that a causal relationship can be assumed. The term 'preliminary' is applied to interventions at Level 2 to indicate that causal assumptions are not yet possible.

Interventions that are 'Not Level 2' (or rated NL2) are not included on the Guidebook; interventions at this level may have other important kinds of evidence, such as a feasibility or implementation study. The rating of NE (no effect) is for interventions with a robust evaluation which has not found evidence of improving any child outcomes.

Currently, the Guidebook provides information based on evidence of impact gathered through a formal evaluation process and does not make judgements about the adequacy or correctness of logic models and intervention designs.



What do the different evidence ratings mean?

Level 4: Effectiveness

Level 4 recognises interventions with evidence of a long-term positive impact through multiple rigorous evaluation studies. At least one of these studies must have evidence of improving a child outcome lasting a year or longer. The evidence may include significant adaptations to meet the needs of different target populations.

The evidence must meet the following requirements:

- The intervention has demonstrated consistent significant positive child outcomes in two rigorous studies meeting all criteria required for Level 3.
- At least one study uses a form of measurement that is independent of the study participants (and also independent of those who deliver the intervention). 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, or administrative data).
- In at least one study there is evidence of a long-term outcome of 12 months or more.

Level 4+

To achieve a 4+ rating:

- All of the criteria for Level 4 must be met.
- At least one of the Level 4 studies will have been conducted independently of the programme developer.
- The intervention must have evidence of improving Foundations' child outcomes from three or more rigorously conducted studies meeting all criteria required for Level 3 and conducted within real-world settings.

Level 3: Efficacy

Level 3 recognises interventions with evidence of a short-term positive impact from at least one rigorous evaluation study – that is, where a judgement about causality can be made. The evaluation should demonstrate a statistically significant positive impact on at least one child outcome.

The evidence must meet the requirements for Level 2, and additionally the following requirements:

-
- 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 or 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- and 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.
- 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).
- For quasi-experimental designs, 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. If there are relevant baseline differences with an effect size of $d > 0.05$ and ≤ 0.25, these differences must be adjusted for in analyses; if there are differences of $d > 0.25$, then the study cannot satisfy the requirement for Level 3.
- For RCTs, where a truly random, rigorous randomisation method is used and <10% attrition, baseline equivalence is not assessed. Where there is >10% attrition, relevant baseline differences with an effect size of $d > 0.05$ and ≤ 0.25 must be adjusted for in analyses; if there are differences of $d > 0.25$, then the study cannot satisfy the requirement for Level 3.
- 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.)
- 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.
- 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.

Level 3+

To achieve a 3+ rating:

- The intervention will have obtained evidence of a significant positive child outcome through a Level 3 efficacy study but also has additional consistent positive evidence from a Level 2 or Level 2+ study with a comparison group design (occurring under ideal circumstances or real-world settings).

Level 2: Preliminary evidence

Level 2 recognises interventions with preliminary evidence of improving a child outcome, but where a conclusion of causal impact cannot be drawn.

The evidence must meet the following requirements:

- Participants complete the same set of measures once shortly before participating in the intervention 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 timepoints 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 intervention 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.
- 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 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.

Level 2+

To achieve a 2+ rating:

- There is a significant positive child outcome in an evaluation study meeting all the criteria for a Level 2 study 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, meeting the Level 3 baseline equivalence requirements.

Not Level 2 (NL2)

The intervention is judged to not meet the Level 2 threshold for a variety of methodological reasons, including the representativeness of the sample and the validity and objectivity of the methods used to measure child impact in one of our seven child outcomes. The Guidebook does not include information on interventions rated below Level 2.

Interventions falling into this category are typically at earlier stages of their development, with important foundational work being carried out. This might include developing a theory of change or logic model, or carrying out feasibility, implementation, or pilot evaluation studies.

No Effect (NE)

The intervention has a rigorous study, meeting Level 3 criteria, which is also the most rigorous impact evaluation, and which has not found evidence of improving a child outcome.

The evidence must meet the following requirements:

- The study must meet the requirements for Level 3.
- It fails to confirm any statistically significant benefits with respect to at least one Foundations child outcome.

This rating should not be interpreted to mean that the intervention will never work, but it does suggest that there are key aspects of the intervention's logic model which require respecification and re-evaluation.

What are mixed findings?

Sometimes, different studies underpinning an intervention's evidence rating may have mixed findings: that is, there are studies suggesting positive impact alongside studies that on balance indicate no effect or negative impact. These are indicated by an asterisk.

- **Level 3+* rating:** If an intervention has strong evidence of impact from a single robust study at Level 3, but also has strong evidence of not having achieved impact from another robust study at Level 3, the intervention will receive a Level 3+* rating.
- **Level 4+* rating:** If an intervention has strong evidence of impact from multiple robust studies, meeting Level 4 requirements, but also has strong evidence of not having achieved impact from other robust studies, the intervention will receive a Level 4+* rating.

How were the evidence standards created?

Foundations' evidence standards were developed in consort with other What Works centres to assess interventions in terms of their impact and cost. They are broadly similar to the Maryland Scale and other critical appraisal systems that recognise stages of development, and were formally approved by our evidence panel during the set-up phase of the Early Intervention Foundation. Foundations' evidence standards are the same as those used on the EIF Guidebook.

The approach of these evidence standards differs from that taken by some other evidence synthesis organisations (such as Cochrane and NICE), which make use of meta-analytic methods to synthesise findings from multiple interventions with similar aims and objectives. These alternative methods result in an aggregate score or statement to provide a robust estimate of the quality of evidence for a given practice type.

Similarly, Foundations' Practice Guides bring together evidence on several interventions or approaches within a practice, in their recommendation standards. Within the Foundations Toolkit, the Practice Guides and Guidebook offer a complementary set of tools, allowing users to access guidance on evidence-based practices, while at the same time identifying interventions which could support the implementation of these practices.

How have the evidence standards changed over time?

Foundations' evidence standards have been continually updated to maintain alignment with other gold standard assessment and reporting tools, such as Version 2 of the Cochrane Risk-of-Bias Tool for randomised trials (RoB 2) and Consolidated Standards of Reporting Trials (CONSORT). We also learn from the evidence standards of other clearinghouses, such as the Title IV-E Prevention Services Clearinghouse, What Works Clearinghouse, California Evidence-Based Clearinghouse, Home Visiting Evidence of Effectiveness (HomVEE), and Blueprints. As we update our evidence standards, it takes time to reassess interventions in line with new standards. This explains some of the variations in ratings across interventions. We include the publication date of each Guidebook entry on our website to indicate when they were assessed.

NL2 rating

Previous versions of the Foundations evidence standards included lower ratings of 0 and 1, where 1 indicated that an intervention has a logic model and testable features of the kind that might be used to evaluate its impact at some point in the future. In our view, this was not a helpful distinction that can help commissioners and others to understand what the evidence was saying about likely or possible impact. Therefore, these interventions were later excluded and are now rated as NL2.

Introduction of plus ratings

After the initial creation of the EIF Guidebook, plus ratings (Level 2+, 3+, and 4+) were added, to provide more granularity to the strength of evidence ratings. Over time, their requirements were also formalised. For example, initially, a Level 2+ rating could be given to a study which was a comparison group study but did not quite meet the Level 3 requirements; subsequently, it was also required that the baseline equivalence requirements of Level 3 were met to receive Level 2+ (but some other Level 3 requirements were not met). A Level 3+ intervention rating, on the other hand, requires just a Level 3 rated study and a Level 2 rated study which is a comparison group study.

Attrition and baseline equivalence

In 2021, Foundations' evidence assessment criteria were recalibrated so that they were consistent with recent global shifts in evidence assessment and reporting standards. These changes were discussed and approved by the Technical Advisory Group, which included a panel of external members with expertise in experimental design and statistical equivalence.

Following this recalibration, when appraising evaluation studies with a comparison group (RCTs and QEDs):

- Baseline equivalence is not assessed in RCT designs where a series of requirements for randomisation are met, and where attrition is low.
- Baseline equivalence is always assessed in RCT designs where a series of requirements for randomisation are not met (or unknown), and where attrition is high.
- Baseline equivalence is always assessed in QED designs, regardless of any design features and levels of attrition.

You can find our baseline equivalence guidance here.

Understanding impact on the Guidebook

On the Foundations Guidebook, impact is the size of the improvements that an intervention has generated for children and young people. In other words, by how much has an intervention improved child outcomes – has it made a big change or a small one? Impact is not the same as strength of evidence, which helps us to understand how confident we can be that the change is caused by the intervention.

How does the Guidebook report impact?

The Guidebook reports impacts of interventions in two ways. For some studies rated at least Level 3, it provides an improvement index score. This can be found in the evidence summary section of the intervention page. With Level 3 studies, we can be sufficiently confident that the impact scores are reliable and meaningful. If impact information is not reported, this is either because a) the study and intervention in question did not receive an evidence rating of Level 3 or above, or b) insufficient information is reported in the original evaluation studies. Some assessment rounds also did not calculate the improvement index for outcomes.

In the full evidence descriptions of all studies, the effect sizes of child and parent outcomes are also presented in an outcomes table, as they are reported in the studies underpinning the evidence rating. These are reported using the unit of effect size used in the original studies (e.g. Cohen's d, Hedge's g, Odds Ratio or Risk Ratio).

Why does the Guidebook report impact in two different ways?

Effects as they were originally measured in the study tell us something useful about the nature of the improvement that an intervention has generated: a 20% reduction in smoking is easily understood, and we can learn what a 5-point improvement on the Problem Behaviour Scale means in practical terms, even if we are initially unfamiliar with the scale.

However, there are limitations to this information. In particular, effects described as they are originally measured will not always be directly comparable. For example, child behaviour problems might be measured in one study on a scale of 1–5 using the Problem Behaviour Scale, and in another study on a scale of 1–12 using the Externalising Problems Inventory. A three-point change on one of these scales may mean something very different from a three-point change on the other scale, and it may be unclear, at a glance, which effect is larger or more meaningful.

Improvement index scores tell us something useful about the relative size of improvement, compared with improvements measured using other scales. This is because the improvement index score is based on a standardised measure of the size of effects, which allows us to compare the relative size of effects, and to compare effects across interventions that may have evaluated improvements on a given outcome using different scales. It also means that a larger improvement index value always, relatively speaking, indicates a larger effect.

How is an improvement index score interpreted?

The improvement index score is a number between 0 and 50 that captures the magnitude of an effect, and it allows you to compare effects that were originally measured on different scales. This metric is sometimes called ‘percentile growth’ or ‘percentile rank improvement’.

Alongside the improvement index is its interpretation: here we report the effects as they were originally measured in the study. This number describes the difference between the average outcomes of those who have received the intervention, and the average outcomes for those who did not receive it – the difference between these outcomes is the improvement that we can attribute to the intervention. Because this describes effects as they were measured in the original study, this information can range from readily interpretable statistics such as ‘a 20% reduction in smoking’ or ‘a 15-percentage-point reduction in the proportion of participants who have developed a major depressive disorder’, to those which require specialist knowledge of particular measures, such as ‘a 5-point improvement on the Problem Behaviour Scale’. We also report the timepoints at which these improvements were observed, including highlighting improvements that have persisted over a longer period of time.

The improvement index score can be interpreted as an estimate of how much we would expect outcomes for the average participant in the control group to improve if they had received the intervention, relative to other members of the control group. That is, if you ordered all the participants in the control group from lowest to highest – worst to best – on a certain outcome, how much would the average person – the person in the middle of the range – improve if they had

received the intervention? Would they move from the middle into the top 25%, into the top 10%, or into the very top 1%?

For example:

- An improvement index score of 25 means we would expect the average participant in the comparison group who did not receive the intervention (for whom 50% of their peers have better outcomes and 50% have worse outcomes) to improve to the point where they would have better outcomes than 75% of their peers, and worse outcomes than 25% of their peers, if they had received the intervention.
- An improvement index score of 50 means we would expect the average participant in the comparison group who did not receive the intervention to improve to the point where they would have better outcomes than 100% of their peers, and worse outcomes than 0% of their peers, if they had received the intervention. In other words, they would have the very best outcome relative to their peers.
- An improvement index score of 0 means that there is no improvement. The average participant in the comparison group who did not receive the intervention would maintain this ranking if they had received the intervention.

In more technical terms, the improvement index is the difference between the percentile rank corresponding to the mean value of the outcome for the intervention group, and the percentile rank corresponding to the mean value of the outcome for the comparison group distribution.

It is worth noting that negative improvement index scores are possible, in cases where studies find that an intervention was harmful. However, these are not currently listed here, as the Guidebook only includes interventions with some evidence of having a positive impact on child outcomes.

While improvement index scores make comparison across interventions possible, their interpretation still requires an understanding of the intervention and its implementation context. There is no simple answer to what a 'good' or 'bad' improvement index score is. It depends on what you are trying to achieve, for whom, and at what cost – and what other similar interventions have a record of achieving.

Where does the Guidebook report effect sizes?

Effect size information, where available in the original studies, is included in the outcomes table in the full evidence description which is available to download as a PDF.

We do not directly report effect sizes on our website, because the improvement index is likely to be more intuitive than effect sizes expressed in units of standard deviation. Also, the improvement index is limited to values between 0 and 50, which helps to contextualise effects and is more user-friendly than effect sizes, which can take on a greater variety of values.

However, it is possible to get a sense of how an effect size translates into an improvement index score. For example, a Cohen's d of 0.5 is roughly equal to an improvement index score of 19; a Cohen's d of 1 roughly equals 34; 2 roughly equals 48; and 3 roughly equals 50.

What do the outcomes tables in the full evidence description show?

Each intervention page includes a downloadable PDF with a full evidence description. This includes an outcomes table, which shows the full results of studies underpinning the intervention. These tables report all outcome measures, their statistical significance, timepoint, and, where available, the number of participants (n) and effect sizes. Not all studies report sample sizes at every timepoint or provide effect sizes. Outcomes which do not contribute to an intervention's rating (for example, because they were not a statistically significant positive outcome), appear only on the outcomes table, and not in the study or evidence summaries on the intervention webpage.

The full evidence description PDF also includes new information about how studies were conducted, characteristics of those who participated, how many participants stayed in the study, and which measurement tools were used. This helps build a fuller picture of how reliable and relevant the evidence is to the local context.

Cost rating

What is the cost rating?

The Guidebook provides a cost rating for each intervention. It is an estimate of how costly an intervention is per person receiving the intervention, based on the resources required to set up and deliver it. The estimate is based on aspects such as the time it takes to deliver the intervention, how many families it aims to reach, and any staff training or qualifications required, which is provided by the intervention developers. It does **not** represent actual prices or fees.

The estimated cost per person is translated to a scale from 1 to 5, where 1 indicates the least resource-intensive interventions and 5 the most resource-intensive per recipient. The rating helps commissioners and decision-makers compare the likely costs of different interventions, based on the resources they require.

The cost rating is **not** the same as the market price for an intervention. It is an assessment of the resources an intervention requires in order to be delivered fully, and the relative cost per recipient, not what it will cost to buy or commission the service. The actual market price typically includes commercially sensitive information that is not routinely available, so will in practice need to be negotiated between provider and commissioner, and can vary.

What do the cost ratings mean?

Interventions are rated on a scale from 1 to 5:

- **1** = Low cost to set up and deliver, compared with other interventions reviewed (estimated cost per recipient: under £100)

- **2** = Medium-low cost to set up and deliver, compared with other interventions reviewed (£100–£499)
- **3** = Medium cost to set up and deliver, compared with other interventions reviewed (£500–£999)
- **4** = Medium-high cost to set up and deliver, compared with other interventions reviewed (£1,000–£2,000)
- **5** = High cost to set up and deliver, compared with other interventions reviewed (over £2,000)
- **NA** = Not enough information to provide a rating.

Note that the cost rating scale was developed several years ago, and the indicative costs have **not** yet been updated to reflect inflation and current costs.

How is the cost rating produced?

The cost rating is based on a weighted nonlinear least squares regression model, which includes information about the different resources needed to deliver the intervention, and their costs. This information is provided to Foundations by the intervention developers or providers, and includes:

- Training fees and time needed to train practitioners
- Requirements for follow-up or booster training
- Costs of initial and ongoing intervention materials
- Practitioner hours required for delivery
- Qualification levels of practitioners and supervisors
- Internal and/or external supervision needs
- Licensing fees
- Typical group size for delivery.

The model provides an estimate of the resource cost per individual, which is then mapped onto the five-point cost rating scale.