Education Endowment Foundation

Implementation and process evaluation (IPE) for interventions ineducation settings: An introductory handbook

Neil Humphrey, Ann Lendrum, Emma Ashworth, Kirsty Frearson, Robert Buck and Kirstin Kerr

MANCHESTER 1824
The University of Manchester

Authors

Neil Humphrey, Ann Lendrum, Emma Ashworth, Kirsty Frearson, Robert Buck and Kirstin Kerr

Education Endowment Foundation

If you have any enquiries about the EEF's approach to conducting implementation and process evaluations, please get in touch with Anneka Dawson.

e. Anneka.Dawson@eefoundation.org.uk

Contents

1 Introduction

The purpose of this introductory handbook is to provide guidance to Education Endowment Foundation (EEF) evaluators (and, indeed, other researchers) on how to conduct high-quality implementation and process evaluations (IPE) at the various stages of intervention development and testing (e.g. pilot, efficacy, effectiveness). The foundation for this handbook was a synthesis of the IPE literature commissioned by the EEF and written by the authors (available here.)

> *"There is no such thing as a typical [implementation and] process evaluation" (Evans, Scourfield & Murphy, 2015, p.1)*

Our starting point is reflected in the quote above. There is no single, universally agreed way to conduct an IPE. Thus, this handbook presents the reader with advice and guidance on the choices available rather than attempting to impose a single evaluation model. That said, there are certain core *principles* of IPE to which all EEF evaluators are advised to adhere (see section 3).

1.1 EEF's approach to IPE

The aim of all EEF projects is to find out what works to raise the attainment of disadvantaged pupils in English schools. To achieve this, all intervention projects are subject to a quantitative impact evaluation[1] to estimate the effect on children's attainment (and, in some cases, other outcomes). An IPE is also commissioned alongside every impact evaluation to understand how a project is implemented on the ground and the elements of successful delivery. It is crucial that the IPE and impact evaluation are carried out by an independent evaluator so that the evidence produced is robust and unbiased.

In the past, the EEF commissioned a process evaluation for all of its projects, but the size, extent and purpose of the process evaluation varied widely. In the last couple of years there has been a greater emphasis on IPE at EEF and this new guidance demonstrates part of EEF's investment in the importance of this aspect of the evaluation. Although there will be some variability in the extent of the IPE necessary for different interventions, some key elements (outlined in section 3) will now be expected in all projects. However, EEF still wants evaluators to be able to explore different methodologies and also to work together with the delivery team to develop the best approach for that particular project.

This guidance should also be used in combination with the EEF's approach to impact evaluation which is clearly articulated on the EEF's website. This handbook explains that the EEF commissions evaluations that reflect the existing evidence base and stage of development of the intervention. It is important that the impact and IPE approach are complementary and can be interwoven in the final evaluation report to aid in understanding why the intervention is successful or unsuccessful.

When reading this guidance it is important to bear in mind that all of the dimensions do not need to be covered in every evaluation. The balance or focus of IPE should be determined by two main factors which both relate to the research questions and objectives. First, what

[1] With the exception of some pilot stage projects that focus on evidence of promise and as such may not include assessment of impact on attainment.

stage is the evaluation at (pilot, efficacy or effectiveness)? Second, what are the key issues which might explain any variation in intervention effects for this stage of evaluation and for this particular intervention? The IPE is therefore closely linked to the logic model or theory of change for the intervention and which features might be identified and evaluated to understand the trial outcomes more effectively.

For a pilot project, the emphasis is very much on evidence of promise (e.g. is there evidence of expected change happening?), feasibility (e.g. is the approach acceptable to participants?), and readiness for trial (e.g. is it replicable and affordable?). How manageable is the intervention and what appear to be the most important factors in successful implementation? The intervention is unlikely to be powered to detect main effects, so any investigation will be exploratory. The emphasis is on identifying factors at this stage which should be explored more systematically at efficacy level or which might inform the design of an efficacy trial.

At efficacy level, the study may well be powered to find significant associations between aspects of implementation and uptake which correlate with successful outcomes. At this stage the emphasis is on understanding variation more systematically, generating further hypotheses for exploration, and/or to help with guidelines for successful implementation for larger-scale effectiveness trials. A key part of this is rigorous assessment of different aspects of implementation (e.g. fidelity).

At effectiveness level, IPE should explore how the intervention is interpreted or used at larger scale ("in the wild"). This may include exploration of the influence of contextual variability on implementation. The scale of the project is likely to make analyses reasonably robust. The emphasis should be on identifying features of successful interventions (and learning from challenges of unsuccessful interventions) which could guide policy or practice guidelines to improve the chances of success at larger scale.

1.2 Navigating this handbook

You will find some sections of this guidance more useful than others, depending upon which stage of intervention development and testing you are focusing on. The flowchart in Figure 1 summarises the sections which follow. There is a more detailed breakdown of content at the beginning of each section.

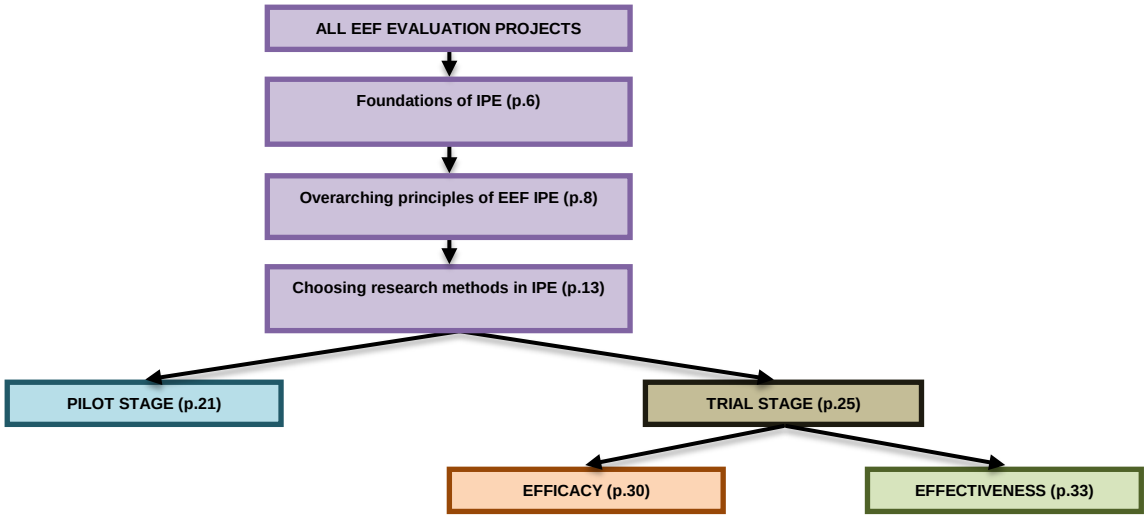


Figure 1. Flow chart of IPE in EEF evaluations.

A brief summary of the EEF IPE guidance

IPE refers to the generation and analysis of data to examine how an intervention is put into practice, how it operates to achieve its intended outcomes, and the factors that influence these processes.

There are 8 generally agreed upon dimensions of implementation (fidelity, dosage, quality, reach, responsiveness, programme differentiation, monitoring of control groups, adaptations) and a number of factors that are believed to affect implementation (preplanning and foundations, the implementation support system, implementation environment, implementer factors, and intervention characteristics).

Key sources of IPE data include researchers (e.g. EEF evaluators), implementers (e.g. teachers), external support staff (e.g. EEF delivery partners) and intervention participants (e.g. students).

The Key principles of IPE we recommend that our evaluators follow are:

1. Evaluators should use the checklist of implementation dimensions to identify how they might assess the full range of those dimensions and prioritise them where necessary.
2. Evaluators should use the checklist of factors affecting implementation to identify how they might assess the full range of those factors, and prioritise them where necessary.
3. The use of mixed methods in IPE should be standard unless there is a sound reason for using a solely quantitative or qualitative approach
4. IPE and impact evaluation should be undertaken using an integrated approach, as opposed to being viewed as completely separate processes
5. The EEF adapted Template for Intervention Description and Replication (TIDieR) framework should be used to ensure that an adequate description of the intervention being evaluated is provided, both in set-up and reporting stages.
6. Evaluators should develop an intervention logic model or theory of change in partnership with the delivery team to inform the evaluation. This should be more thorough than the basic logic model developed in the EEF set up meetings in the majority of cases.
7. Following the project set-up meetings, the evaluation team and delivery partner should meet for an Intervention Delivery and Evaluation Analysis (IDEA) workshop where possible.
8. Detailed descriptive data on implementation should be considered as a bare minimum reporting requirement, without which the validity of findings is undermined.
9. Evaluators should report on the development, piloting and psychometric properties (for example, inter-rater reliability statistics) of quantitative implementation measures in appendices in the EEF report.
10. Evaluators should map out the possible evidence indicators at each stage of the intervention theory and use the pilot evaluation to collect data so as to assess its veracity.
11. Evaluators should seek to assess usual practice in *all* trial schools prior to randomisation and at least once more at an appropriate future point in time (e.g. at post-test).
12. Quantitative usual practice data should be analysed in order to assess any changes that have occurred during the trial that may have implications for the interpretation of trial outcomes (e.g. evidence of compensatory rivalry in control schools, or displacement in intervention schools).
13. Evaluators should examine what adaptations are made, when, why, and the extent to which they are congruent with the intervention theory.
14. All subgroups should be pre-specified and the choice of subgroups should be informed by the intervention theory and literature.
15. Quantitative IPE analysis plans should be specified clearly in the trial protocol.
16. Where quantitative implementation data is available, analyses should be undertaken to examine the relationship between implementation variability and intervention outcomes.
17. Where appropriate and feasible, evaluators should conduct an on-treatment analysis alongside their intention-to-treat analysis. The on-treatment threshold should be articulated and agreed in advance of data collection in order to avoid data mining
18. Evaluators should document clearly the contextual variation that is introduced in effectiveness trials, and where appropriate and feasible they should conduct analyses to explore context-implementation and/or context-implementation-outcomes associations.

2 Foundations of implementation and process evaluation

IPE refers to the generation and analysis of data to examine how an intervention is put into practice, how it operates to achieve its intended outcomes, and the factors that influence these processes.

This is particularly important as research consistently shows that the implementation of interventions is variable across settings, and that this variability impacts upon the achievement of expected outcomes.

> *"[We need to learn] why various programs do or do not work, for whom and under what conditions they work, what is needed to scale up proven programs, and what policy supports are needed to scale them up without losing their effectiveness" (Slavin, 2012, p.xv)*

Put very simply, if randomised controlled trials (RCTs) tell us 'what works', IPE helps us to understand *how* and *why* they work (or, in the case of null impact, why an intervention appears not to have worked).

An understanding of IPE is essential at all stages of intervention development. IPE data can be used formatively (e.g. providing feedback that helps developers refine their intervention), summatively (e.g. helping to explain programme impact, or lack thereof), and for knowledge generation (e.g. improving our understanding of how interventions work).

The primary academic foundation of IPE is implementation science. Implementation science is a multi-disciplinary field whose broad aims are to enhance the effectiveness of interventions by: (a) improving understanding of implementation processes, (b) supporting implementation processes, and (c) evaluating implementation processes.

Theories, models and frameworks underpin high quality IPE. They can be used by EEF evaluators to inform data generation and analysis, improve understanding of what influences implementation processes and outcomes, and help to guide the process of translating research into practice.

2.1 Dimensions of and factors affecting implementation

Implementation is a multidimensional construct. There are eight generally agreed-upon dimensions:

1. **Fidelity/adherence** – the extent to which implementers (e.g. teachers) adhere to the intended treatment model
2. **Dosage** – how much of the intended intervention has been delivered and/or received
3. **Quality** – how well different components of an intervention are delivered
4. **Reach** – the rate and scope of participation
5. **Responsiveness** – the degree to which participants engage with the intervention
6. **Programme differentiation** – the extent to which intervention activities can be distinguished from other, existing practice
7. **Monitoring of control/comparison groups** (in a trial context) – determination of the 'counter-factual' (e.g. that which is taking place in the absence of the intervention)
8. **Adaptation** – the nature and extent of changes made to the intervention

Many evaluations focus solely on a single implementation dimension, such as fidelity or dosage. This limits the utility of IPE and can lead to a Type III error (the inaccurate attribution of cause)[2].

There are also a number of factors that are believed to affect implementation, including:

- **Preplanning and foundations** – e.g. what is the level of need, readiness and capacity for change in the setting where the intervention takes place?
- **Implementation support system** – e.g. what strategies and practices are used to support high quality implementation?
- **Implementation environment** – e.g. what are the influential contextual and compositional characteristics in the setting where the intervention takes place?
- **Implementer factors** – e.g. what is the profile of professional characteristics, intervention perceptions and attitudes, and psychological characteristics among implementers?
- **Intervention characteristics** – e.g. what form does the intervention take?

The dimensions of and factors affecting implementation noted above are central to IPE and are referred to throughout this handbook.

Further reading

Click the article title to access it online

Lendrum, A. & Humphrey, N. (2012). The importance of studying the implementation of interventions in school settings. *Oxford Review of Education*, *38*, 635-652.

The above paper sets out the various reasons for studying the implementation of interventions in school settings.

Nilsen, P. (2015). Making sense of implementation theories, models and frameworks. *Implementation Science*, *10*, 53-68.

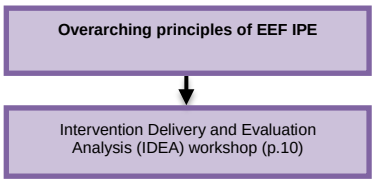
The above paper provides an up to date and comprehensive analysis of the range of theories, models and frameworks that are relevant to IPE, including process models, determinant frameworks, classic theories, implementation theories, and evaluation frameworks.

Durlak, J. & DuPre, E. (2008). Implementation matters: a review of research on the influence of implementation on program outcomes and the factors affecting implementation. *American Journal of Community Psychology, 41*, 327-350.

The above paper offers a comprehensive review of the evidence relating to dimensions of and factors affecting implementation.

[2] For example, if the only aspect of implementation that researchers assess is fidelity, it may be tempting to conclude that low fidelity is causal if there are null intervention effects. In fact, these poor outcomes may be more accurately attributed to an aspect that has not been measured, such as quality or participant responsiveness. Similarly, if not all aspects of implementation are considered, and fidelity is measured as high, poor outcomes may be incorrectly seen as programme or theory failure.

3 Overarching principles of implementation and process evaluation in EEF projects



The following key principles of IPE should be used at all stages of evaluation (e.g. pilot, efficacy, effectiveness).

1: Evaluators should use the checklist of implementation dimensions to identify how they might assess the full range of those dimensions and prioritise them where necessary.

Failure to assess implementation, or the assessment of a single dimension only (e.g. dosage), limits the utility of IPE and can lead to a Type III error (the inaccurate attribution of cause).

> *"Few studies… [have] collected data about all aspects [of implementation]. The exemplary use of a broader range of measures should be encouraged. Further, multimodal assessments, where multiple sources of data are used, should also be encouraged. All fields of intervention will mature more quickly if these higher standards become widely adopted" (Hansen, 2014, p.355).*

2: Evaluators should use the checklist of factors affecting implementation to identify how they might assess the full range of those factors, and prioritise them where necessary.

Implementation drives outcomes, and so finding out what drives implementation is an important consideration for evaluators, particularly if there is an expectation that the intervention will be brought to scale.

3: The use of mixed methods in IPE should be standard unless there is a sound reason for using a solely quantitative or qualitative approach

A review of IPE in EEF protocols and reports suggests that a primarily qualitative approach has been favoured to date. Qualitative and quantitative approaches are complementary and offer credibility and flexibility.

4: IPE and impact evaluation should be undertaken using an integrated approach, as opposed to being viewed as completely separate processes

This is particularly important at the reporting stage, where for example IPE findings can yield significant insights that contextualise and improve our understanding of the impact (or lack thereof) of an intervention.

5: The EEF adapted Template for Intervention Description and Replication (TIDieR) framework (below) should be used to ensure that an adequate description of the intervention being evaluated is provided, both in set-up and reporting stages.

The Template for Intervention Description and Replication (TIDieR) was originally published by Hoffman et al (2014). A version adapted for EEF projects is shown here:

1. Brief name
2. Why: Rationale, theory and/or goal of essential elements of the intervention
3. Who: Recipients of the intervention
4. What: Physical or informational materials used in the intervention
5. What: Procedures, activities and/or processes used in the intervention
6. Who: Intervention providers/implementers
7. How: Mode of delivery
8. Where: Location of the intervention
9. When and how much: Duration and dosage of the intervention
10. Tailoring: Adaptation of the intervention
11. How well (planned): Strategies to maximise effective implementation
12. How well (actual): Evidence of implementation variability*

*Included once an intervention has actually been piloted or trialled and implementation has been documented.

Use of the adapted TIDieR framework will ensure shared understanding of what is being evaluated among key stakeholders (e.g. EEF, delivery partner, evaluation team) and improve the consistency of intervention description in EEF reports. Put simply, knowledge of 'what works' will be greatly improved if the 'what' is articulated in sufficient detail. Lack of adequate intervention description is identified as a threat to the validity of findings in EEF's padlock security rating (EEF security classification guide can be accessed here)

A completed exemplar of the EEF adapted TIDieR framework can be found in Appendix 1.

6: Evaluators should develop an intervention logic model or theory of change in partnership with the delivery team to inform the evaluation. This should be more thorough than the basic logic model developed in the EEF set up meetings in the majority of cases.

It is important to know not just if an intervention 'works' in terms of producing desired outcomes, but also if it works in the manner theorised.

> *"Seasoned travellers would not set out on a cross country motor trip without having a destination in mind, at least some idea of how to get there, and, preferably, a detailed map to provide direction and guide progress along the way" (Stinchcomb, 2001, p.48)*

Suggested steps in the development and utilisation of an intervention logic model (adapted from Coffman, 1999) are as follows:

1. Determine the appropriate scope for the logic model (for EEF-funded interventions the default should be the intervention in its entirety, unless otherwise agreed)
2. Identify the model's components (e.g. inputs, processes/mechanisms, outcomes, and moderating factors)
3. Draft the logic model (a template is included in this guidance document; see below)
4. Use the logic model as an evaluation framework (e.g. develop data indicators of the different components and track these in the evaluation)

5. Revisit the model and use it as a learning tool (e.g. use indicator data to assess the extent to which the intervention is working as intended, and whether the model needs to be revised)

An exemplar logic model template (and a link to a step-by-step completion guide) can be found in Appendix 2.

Evaluators will need to collaborate with the delivery team and build the logic model approach into project plans and protocols e.g. discuss intervention theory in set-up and Intervention Delivery Evaluation and Analysis (IDEA) meetings (see guidance point 7); the delivery team co-produce or sign off on intervention descriptions in evaluation reports).

What aspects of IPE should be prioritised in EEF evaluations?

Whatever the stage of evaluation, IPE should be as comprehensive as possible. However, budget and data burden will mean that evaluators will need to prioritise and make informed decisions about what is essential and feasible in the context of a given project. This handbook provides guidance to aid these decisions.

For example, it may not be feasible to assess the full range of implementation dimensions and/or factors affecting implementation in a given project. Evaluators will need to determine what data will be of most value to the evaluation. Issues to consider include:

- What are the key research questions?
- Which are most meaningful implementation dimensions given the nature of the intervention?
- Is there naturally occurring implementation data (e.g. monitoring data collected by the delivery partner) that could be used (or developed/adapted for the purpose of the IPE)?
- What are the possible sources of data and how can data burden be minimised?
- Can mixed methods be used to maximise efficiency (e.g. implementation dimensions assessed quantitatively, factors affecting implementation assessed qualitatively)?
- Is there existing literature on implementation (e.g. what do we know already about the relative importance of different implementation dimensions for the intervention in question)?
- What is the stage of evaluation and is there any risk of interaction between implementation monitoring and implementer behaviour (e.g. given that increased monitoring influences implementation, IPE in effectiveness trials may need to be less intensive and comprehensive)?

3.1 Intervention Delivery and Evaluation Analysis (IDEA) workshop

7: Following the project set-up meetings, the evaluation team and delivery partner should meet for an Intervention Delivery and Evaluation Analysis (IDEA) workshop where possible

Whatever the stage of evaluation, this workshop will provide a strong foundation for a high quality IPE. Building upon the standard EEF set-up meetings, the aim of an IDEA

workshop is to enable the evaluation team and delivery partner to explore the intervention in greater depth and develop the IPE evaluation, including:

- *Co-construct and agree the TIDieR framework content for the intervention*
- *Interrogate or develop the intervention logic model (or theory of change)*
- *Examine intervention delivery and training materials (e.g. manuals)*
- *Explore existing literature and evidence about the intervention*
- *Discuss the intervention delivery history (within or beyond the EEF context)*

The information generated through these activities will enable the teams to:

- *Identify the most salient dimensions of and factors affecting implementation and consider when and how they may be assessed (e.g. how would 'dosage' be interpreted for this intervention? What is the optimal dosage and why? How should this be assessed?)*
- *Identify the most salient contextual factors that are likely to influence implementation and outcomes (particularly in effectiveness trials) so that these can be assessed as part of the IPE section of the protocol (see section 8.2)*
- *Start scoping data collection tools to provide evidence to support/challenge (in pilots) or empirically validate (in efficacy trials) the intervention logic model (or theory of change) (e.g. change mechanism indicators)*
- *Clarify which (if any) subgroups are likely to experience differential intervention benefits (and why) and should therefore be included in the trial protocol*
- *Generate a definition of 'on treatment' status, including data needed to determine this (in efficacy trials)*
- *Build a detailed IPE data generation and analysis protocol (within the constraints of the budget agreed by the EEF grants committee) for publication on the EEF website*
- *Consider the availability and specify naturally occurring data to support the above*
- *Agree any procedures for data sharing between the evaluation and delivery teams*

The IDEA workshop should take place prior to the finalisation of the evaluation protocol. Where it is difficult for a full IDEA workshop to take place due to time constraints or other restrictions, the evaluation and delivery teams should as a minimum ensure they have had an in-depth discussion about the IPE, including identifying any data that delivery partners have already collected that can be shared.

Some follow-up contact is likely to be needed following the IDEA workshop in order to finalise and agree the approach to be taken in relation to the above issues, although this may be achieved by email or teleconference with the delivery partner lead. With pilot projects, additional interviews or focus groups might be needed with key stakeholders in order to gain enough information to develop the theory of change/logic model.

8: Detailed descriptive data on implementation should be considered as a bare minimum reporting requirement, without which the validity of findings is undermined (EEF security classification guide can be accessed here)

Further reading

Click the article title to access it online

Kallestad, J. H., & Olweus, D. (2003). Predicting teachers' and schools' implementation of the Olweus Bullying Prevention Program: A multilevel study. *Prevention and Treatment*, 6, Article 21.

The above paper provides an illustration of how implementation variability in an intervention can be predicted by teacher and school-level factors.

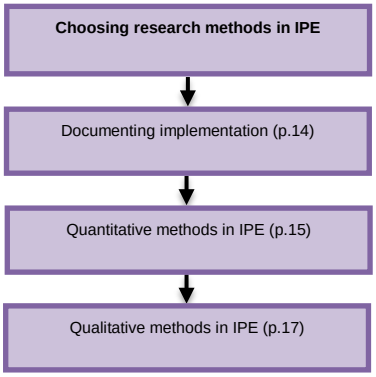
Humphrey, N. et al (2015). *Promoting Alternative Thinking Strategies (PATHS): Evaluation report*. London: EEF.

The above report provides an illustration of the application of many of the preceding overarching principles of IPE in EEF-funded research. IPE was an integral part of the PATHS efficacy trial. The evaluators made use of a determinant framework to guide their assessment of factors affecting implementation that drove their process-oriented questions. A range of implementation dimensions were assessed and used to examine their relative association with intervention outcomes. Finally, the IPE component of the trial utilised a mixed methods approach that incorporated surveys, structured observations and interviews, yielding both quantitative and qualitative data.

Fraser, M. W. & Galinsky, M. J. (2010). Steps in intervention research: Designing and developing social programs. *Research on Social Work Practice*, 20, 459-466.

The above paper considers the role of research and evaluation in informing the development of interventions. The steps outlined align broadly with EEF research stages (e.g. pilot, efficacy, effectiveness, scale-up).

4 Choosing research methods in IPE



This section discusses the methods that evaluators might use in IPE and some of the trade-offs between approaches that could be taken.

> *"At the heart of understanding how to develop [and evaluate] interventions is the realization that no one research method in isolation will suffice" (Borglin, 2015, p.29)*

The four mixed-methods designs in the social sciences that may be useful frameworks for EEF evaluators are listed in Table 1.

Design	Structure and priority	Integration of methods and data	Hypothetical IPE example
Convergent parallel	QUANT QUAL	Integration occurs during main interpretation of findings	Quantitative implementation data is captured through implementer self-report surveys while qualitative process data is generated in parallel
Explanatory sequential	QUANT□qual	First phase findings inform second phase	Quantitative data captured through structured observations indicates poor implementation (e.g. limited dosage). The reasons underpinning this subsequently form the central focus of interviews with implementers
Exploratory sequential	QUAL□quant	First phase findings inform second phase	Observations, interviews and document analysis in a small number of sites used to develop understanding of different implementation dimensions in a complex intervention. This data is subsequently used to develop a quantitative implementation survey with a view to conducting descriptive and/or relational analyses
Embedded and/or multi-phase	□ QUANT QUAL □	Methods and data blended from the outset Adoption of a cyclical process in which one or the other is prioritised at a given stage of evaluation, with each approach utilised providing a foundation for the next	Structured observations of implementation across intervention sites reveal a range of adaptations used by implementers. These are followed up in a small number of qualitative interviews to probe implementers' understanding of the adaptation process (e.g. what adaptations are made, when, how and why). This data is used to develop a quantitative survey focusing on adaptations which is implemented across intervention sites.

Table 1. Frameworks for mixed-methods research (adapted from Borglin, 2015).

In planning a mixed methods IPE, evaluators are advised to carefully consider the relationship between this aspect of their research and the evaluation as a whole. For example, in a trial context, how, when and where does the IPE strand fit in relation to the assessment of intervention outcomes? What purpose does the data generated through the

IPE serve, and how does this relate to the principal functions of IPE at different stages of intervention research and development? How will IPE and impact findings be integrated?

4.1 Documenting implementation

Detailed documentation of implementation serves to increase internal validity as it provides clear evidence that the intervention actually took place. This can be further enhanced through analyses that establish a relationship between implementation variability and outcomes (see sections 6.4, 7.1 and 8.1).

Qualitative and/or quantitative data could be used. For example, qualitative data can provide a detailed narrative summary, whereas quantitative data can generate descriptive statistics on the implementation dimensions assessed (e.g. fidelity, dosage) across schools. Quantitative data is essential if evaluators plan to assess the relationship between implementation and outcomes and/or conduct critical component analyses (see chapters 6-8).

> *"Accurate interpretation of outcomes depends on knowing what aspects of the intervention were delivered and how well they were conducted" (Durlak & DuPre, 2008, p.328)*

How might different implementation dimensions be assessed?

An important distinction here is how prescriptive the intervention is. For more prescribed, 'manualised' interventions, the following methods could be used:

- **Fidelity/adherence** may be assessed by rating, for example, the proportion of sessions covered, and/or extent to which the implementer followed the session protocol. A distinction may be made here between adherence in terms of content delivered and use of prescribed approaches to delivery. Quantitative ratings, such as those generated through surveys or structured observations, may be supplemented by interviews with implementers to explore their reasons for any lack of adherence, challenges to fidelity and the extent to which fidelity is feasible.
- **Dosage** may be assessed by rating, for example, the number or proportion of intervention sessions delivered and/or the amount of time spent delivering the intervention. Interview data may explain *why* dosage is variable.
- **Quality** may be assessed by rating, for example, implementer interest and enthusiasm, preparedness, clarity of expression, and responsiveness during the delivery of a session. Interviews with intervention recipients (e.g. students) or external support personnel (e.g. coaches) or semi-structured observations may also be used to explore factors that contribute to quality, and how this, in turn, may influence student engagement.
- **Participant responsiveness** may be assessed through focus groups or interviews with recipients and/or implementers, to explore whether participants engage particularly well with some aspects of the intervention and not others, and the reasons for this. It may also be examined by rating, for example, the extent to which recipients appear to be engaged and interested in the intervention materials and activities.
- **Reach** may be assessed by rating, for example, the proportion of the intended intervention recipients present in a given session. Interview data with implementers may explain why some intended recipients may not be present for the intervention

(e.g. students withdrawn from class for additional support, or for whom the intervention is considered unsuitable).

- **Adaptation** – see section 6.1
- **Programme differentiation** and monitoring of control groups – see section 6.2.

For flexible, less prescriptive interventions, and/or those that are complex and multi-component in nature, the process of measuring implementation is more complicated. For example, fidelity/adherence takes on a different meaning when there is no prescribed model against which to assess the delivery. In this case, fidelity may be assessed by rating, for example, the extent to which the key *principles* of the intervention are being adhered to as opposed to specific *practices*. Agreeing such principles with delivery partners from the outset is an important consideration, and as such is an ideal topic for exploration in an IDEA workshop (see guidance point 7).

4.2 Quantitative methods in IPE

There are essentially four main sources of quantitative data regarding implementation of school-based interventions. These are:

(i) researchers (e.g. EEF evaluators),
(ii) implementers (e.g. teachers),
(iii) external support staff (e.g. EEF delivery partners), and
(iv) intervention participants (e.g. students).

Use of the latter is confined almost exclusively to assessing participant responsiveness (and/or for triangulation purposes – e.g. to check whether the intervention has been delivered), so our discussion will focus on the first three sources.

Which source(s) of quantitative implementation data should evaluators use?

Table 2 summarises the advantages and limitations of gathering implementation data from the three primary sources listed above.

Source	Researchers	Implementers	External support staff (e.g. delivery partner)
Exemplar method(s)	Structured observation	Self-report survey	Informant-report survey and/or structured observation
Advantages	+ Rigour and objectivity + Stronger associations with intervention outcomes than other sources	+ Low cost and less intrusive than researcher observations + Can provide a summative account of a specified period of implementation (e.g. one year)	+ Minimal data burden on schools (if this is routinely collected as part of the intervention model) + Increased external validity if implementation monitoring is routine as part of intervention model
Limitations	- Time consuming and costly - Single observations provide only a snapshot of activity; multiple observations are burdensome and may increase attrition - Intensive observation may interact with implementation	- May be subject to bias in the form of impression management, demand effects, and/or differences in understanding of implementation requirements - Weaker associations with intervention outcomes than other sources	- Reduced evaluator control (e.g. focus, informed consent, inter-rater reliability) - Data may be biased by relationship between delivery partner and schools/teachers and their vested interest in the outcome of the evaluation.

Table 2. Advantages and limitations of different data sources/methods in quantitative IPE.

In ideal circumstances, a multi-method, multi-dimensional approach is preferable. However, budgetary and other constraints may mean that this is impractical. Hence, evaluators are advised to select an implementation data source that provides the optimal balance between rigour and feasibility.

In terms of the optimal frequency with which implementation data is collected, this will depend partly upon the source and method used, and the nature of the intervention itself. For example, implementer self-report surveys may only need to be administered once if the intention is for them to provide a summative account of implementation over a given period. Similarly, external support staff may provide implementation data spanning multiple points in time through the course of a trial by default (e.g. fidelity checklists completed as part of routine support visits).

What is the optimal approach to developing quantitative implementation measures?

> *"Even if the concept of implementation is not new, the idea of developing ways of measuring it certainly is" (Ogden & Fixsen, 2014, p.8)*

9: Evaluators should report on the development, piloting and psychometric properties (for example, inter-rater reliability statistics) of quantitative implementation measures in appendices in the EEF report

It is important that quantitative implementation measures are fit for purpose and can be shown to be reliable and valid. As a first step, EEF evaluators are advised to check with delivery partners and others connected to the intervention in question whether there are existing measures that can be used. For interventions that have been well validated in other countries and are being 'imported' and/or where implementation monitoring is routine, it is likely that there are tools that could be adapted to suit the EEF trial context.

Where there are no suitable instruments, or existing measures require significant adaptation, evaluators are advised to make use of the implementation science literature (e.g. see recommended chapter by Hansen (2014) below) and their knowledge and understanding of the intervention in question to develop an appropriate instrument(s). As with other aspects of IPE, this is likely to require collaboration with the intervention delivery partner (e.g. in the IDEA workshop).

Developing a quantitative implementation measure

By way of illustration, consider the development of an observation schedule to assess the implementation of the Good Behaviour Game (GBG), being trialled by EEF (2015-2017). The evaluators began by drawing together relevant literature on the measurement of implementation, alongside existing tools that had been used to assess the GBG in earlier studies. As implementation monitoring in the GBG is part of the intervention model, the evaluators also made use of an external coach fidelity checklist developed by the American Institutes for Research (AIR). The existing tools tended to focus primarily upon procedural fidelity and so the evaluators developed additional items focusing on participant responsiveness and quality. They used video footage of GBG implementation (recorded as part of the UK pilot that preceded the EEF trial) as a frame of reference throughout the schedule development process. Alongside the schedule, the evaluators developed a detailed rubric to aid observers. Once the schedule and rubric were complete, the evaluators used additional footage from the UK GBG pilot to assess inter-rater reliability (IRR) among members of the research team. Cohen's Kappa (for nominal items) and intra-class correlation co-efficients (for ordinal items) indicated excellent levels of agreement between raters (e.g. Cohen's Kappa of 0.95 for procedural fidelity). As direct observation of the GBG cannot provide meaningful data on dosage, the evaluators are collecting this via teacher self-report surveys (average number of times the GBG is played per week, and average length of each game) to be administered during fieldwork visits as part of the trial IPE.

A copy of the structured observation schedule developed for the EEF GBG trial can be found in Appendix 3. The accompanying rubric is available on request from the evaluation team.

4.3 Qualitative methods in IPE

Qualitative IPE can help evaluators to

(i) determine stakeholder views of the intervention and the way in which it is designed to be implemented
(ii) document the implementation of the intervention (e.g. the dimensions of implementation)
(iii) highlight contextual factors that influence implementation (see section 2.1), particularly those that may have implications for the transferability and/or sustainability of the intervention, and
(iv) validate or modify the intervention theory by illuminating the change process (e.g. perceptions of how and why the intervention improves outcomes).

In planning the qualitative strand(s) of an IPE, some fundamental issues for evaluators to consider include:

- What is the role of the qualitative researcher(s) relative to the rest of the trial team (e.g. nature and amount of contact such that any blinding is not compromised)?
- What specific research questions will be addressed in the qualitative strand of the study?
- Assuming a mixed-methods design, which research questions involve overlap with the quantitative strand?
- What overall approach (e.g. longitudinal case studies) is planned? What is the sampling strategy (e.g. to achieve maximum variation)?
- What data generation methods (e.g. observation, interview, document analysis) and sources (e.g. teachers, external support staff, pupils, parents, school leadership) will be used?
- How will the data be analysed?
- How will the results be integrated with the quantitative findings?

The primary sources of qualitative data are

(i) researchers (e.g. EEF evaluators),
(ii) implementers (e.g. teachers),
(iii) external support staff (e.g. EEF delivery partners),
(iv) intervention recipients (e.g. students).

Other stakeholders may also be important sources of data, depending on the questions to be answered (e.g. head-teachers, school-based intervention co-ordinators, non-implementing teachers and school support staff, and parents).

In terms of data generation, there are essentially four main methods available. These are

(i) observations (typically semi-structured),
(ii) interviews,

(iii) focus groups, and
(iv) document review.

Which source(s) of qualitative implementation data should evaluators use?

Table 3 summarises the principal advantages and limitations of generating implementation data using the methods and data sources listed above.

Exemplar method	**Semi-structured observations**	**Semi-structured interviews**	**Focus groups**	**Document review**
Exemplar sources	Researcher	Implementers School co-ordinator Head-teacher Parents External support staff	Intervention recipients (e.g. students)	General school documents (e.g. school improvement plan) Intervention-related documents (e.g. lesson plans) Students' work Intervention and training materials
Advantages	+ Flexibility allows re-focusing onto aspects of implementation which may not have been considered salient previously + Can reveal changes in implementation processes over time + Naturalistic data on dimensions of implementation and how these are influenced by contextual factors + Can increase understanding of what the intervention looks like in practice	+ Detailed explanations of decisions concerning processes and dimensions of implementation + Detailed explanations of how contextual factors affect implementation + Can generate insights regarding explicit and implicit beliefs about the intervention, perceptions of impact *et cetera* + Flexibility allows for progressive focusing on unanticipated issues and insights	+ Detailed explanations of pupil responsiveness to all components of an intervention + Use of visual supports/cues (e.g. programme materials) enables involvement of even youngest pupils + Can be important source of triangulation with data from other sources e.g. dosage +Can provide insights into factors affecting implementation	+ Increases understanding of status of intervention within school
Limitations	- Time consuming and costly - Single observations provide only a snapshot of activity; multiple observations are burdensome, may interact with implementation and increase attrition	- May be time consuming and costly to conduct and analyse - May be subject to bias in the form of impression management, demand effects, and/or differences in understanding of implementation requirements	- Reduced researcher control over direction of discussion - Selection of participants generally controlled by school -Pupils may not recognise all components of an intervention	- Time consuming - Limited data yield compared to other methods

Table 3. Advantages and limitations of different data sources/methods in qualitative IPE.

While a multi-method, multi-dimensional approach is preferable, evaluators should be aware of budgetary and data burden constraints (see section 4.2 above) and select data sources and methods that balance rigour and feasibility. The optimal frequency with which qualitative data is collected will partly depend upon the research questions to be answered. Implementation is dynamic, however, and data collected at different time-points allows progressive focusing and examination of the different 'phases' of implementation (Fixsen et al, 2005):

1. Exploration, e.g. exploring readiness, perceptions of needs and benefits
2. Installation, e.g. acquiring resources including training and setting the stage for initial implementation
3. Initial implementation, e.g. using the intervention for the first time, attempting to integrate it with existing processes and practices
4. Full implementation, e.g. broadening or scaling up of the intervention within the school, ensuring that it becomes 'embedded', sustaining implementation over time

Exploring the phases of implementation using qualitative approaches

For example, before and at the start of implementation (phases 1 and 2 above), interviews with implementers may provide insights into contextual factors that influence pre-planning and foundations for implementation (e.g. intervention acceptability – see section 5). Dimensions of implementation to be explored may include programme differentiation and planned adaptations (e.g. to improve goodness-of-fit). Document review will generate data on the perceived need and status/priority of the intervention.

As implementers begin to use the intervention (phase 3 above), interviews with them can usefully explore moderating and contextual factors, including the implementation environment, implementation support system and implementer factors (see section 2.1). These interviews may also explore changing priorities and feasibility (see section 5), while continuing to document the dimensions of implementation. Observations of the delivery of the intervention at this stage will support triangulation of data on implementation dimensions and processes if conducted after interviews, or may generate questions around processes if conducted beforehand. Document review (e.g. of timetables and lesson plans and/or students' completed work) will generate useful data through which to explore dosage.

Towards the end of the implementation period, at phase 4 above, interviews with implementers may explore perceptions of impact (and thus, utility – see section 5), and factors that may have influenced the full integration of the intervention. Focus groups with students at this phase can help evaluators gauge participant responsiveness and triangulate data on other dimensions of implementation (e.g. dosage).

What is the optimal approach to developing qualitative implementation assessment tools?

As with quantitative implementation measures, it is crucial that qualitative data generation instruments (e.g. interview schedules) are fit for purpose. As a first step, EEF evaluators are advised to check with delivery partners and others connected to the intervention in order to ascertain whether there are existing instruments (e.g. those used in previous evaluations) that can be adapted to suit the current evaluation context and research questions. Where there are no suitable instruments, or where those that are available require significant adaptation, evaluators are advised to make use of the aforementioned IPE review completed by the authors, available here and their knowledge and understanding of the intervention in question (including, but not limited to, information provided in the TIDieR framework and intervention logic model – see section 3) to develop an appropriate instrument(s). As with other aspects of IPE, this is likely to require collaboration with the intervention delivery partner (e.g. in the IDEA workshop and follow-up discussions – see section 3).

A copy of the implementation interview schedule developed for the EEF PATHS trial can be found in Appendix 4.

Developing qualitative implementation assessment tools

By way of illustration, consider the process undertaken to develop a semi-structured interview schedule for the assessment of the implementation in the Promoting Alternative Thinking Strategies (PATHS) trial (see recommended reading in section 3). The evaluators began by ensuring that they were familiar with the intervention materials and guidance on implementation. Relevant literature on the assessment of implementation was identified alongside existing interview schedules that had been used to assess PATHS implementation, or had been used by the evaluation team in previous evaluation projects and had been demonstrated to be valid and effective in generating appropriate data. The aims of the interviews (in relation to the study hypotheses and research questions) were determined and explicitly stated at the start of the schedule (e.g. dimensions of implementation, factors affecting implementation, sustainability, perceptions of impact). As the schedule was semi-structured, this served as a useful reminder for interviewers about the areas to be explored and returned to, so that they were free to probe interviewees' responses outside the constraints of a strict sequence of questions. The aims of the interview determined the areas to be explored and why, supporting the development of primary questions and further probes. A detailed schedule clarifying and explicitly stating the questions and reasons for them was developed. This was piloted and refined, and finally a complementary abbreviated schedule listing just the aims, questions and probes was developed for ease of use during interviews.

Further reading

Click the article title to access it online

Nastasi, B. K. et al (2007). Mixed methods in intervention research: theory to adaptation. *Journal of Mixed Methods Research*, 1, 164-182.

The above paper proposes a set of five mixed methods designs related to different stages of intervention research.

Gorard, S., Siddiqui, N., & Huat See, B. (2015). *Philosophy for Children: Evaluation report and executive summary*. London: EEF.

The above report provides a case example of the use of the delivery partners as a source of implementation data (in this case, dosage and quality).

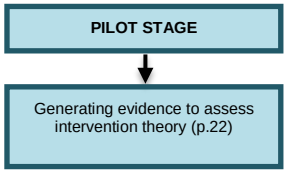
Schonert-Reichl, K. A., Smith, V., Zaidman-Zait, A., & Hertzman, C. (2012). Promoting children's prosocial behaviors in school: Impact of the Roots of Empathy program on the social and emotional competence of school-aged children. *School Mental Health*, 4, 1-21.

The above paper provides an illustration of detailed descriptive reporting of implementation (in this case, dosage, fidelity and participant responsiveness) in a trial context.

Hansen, W. (2014). Measuring fidelity. In Z. Sloboda & H. Petras (Eds.), *Defining prevention science* (pp.335-359). New York, NY: Springer.

This chapter explores different approaches to measuring different implementation dimensions (including exemplar items) in a range of studies. The chapter also includes discussion of the various sources of implementation data.

5 IPE at the pilot stage



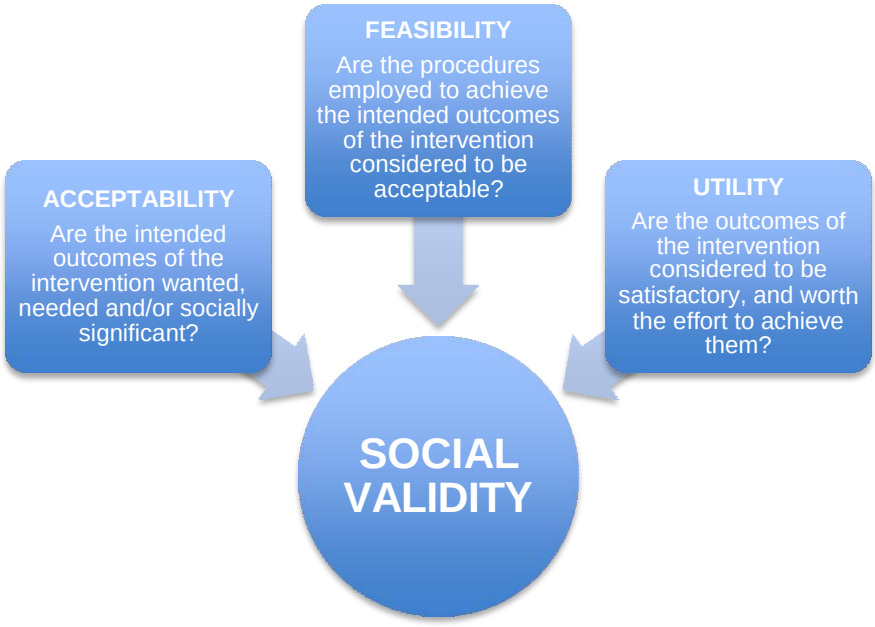
This section discusses the purpose of EEF pilots, research questions and approach. The purpose of EEF pilots is to evaluate whether the intervention fulfils the following three criteria:

(i) **Evidence of promise** (e.g. is there evidence of expected change happening?);
(ii) **Feasibility** (e.g. is the approach acceptable to participants?); and
(iii) **Readiness for trial** (e.g. is it replicable and affordable?).

The piloting stage involves the first 'real world' testing of an intervention and is typically conducted on a small scale. An intervention may be modified/refined following or during an EEF pilot study, including:

- Changes to intervention materials and procedures
- Changes to intervention theory (e.g. logic model or theory of change)
- Methodological changes to evaluation protocols (e.g. procedures for assessing implementation and/or outcomes)

Broadly speaking, the three EEF success criteria are equivalent to the establishment of the 'social validity' (Wolf, 1978) of the intervention. Social validity refers to the value and social importance attributed to an intervention by its direct or indirect consumers, and can be thought of in terms of the following dimensions:



5.1 Exemplar research questions for EEF evaluators at the pilot stage of intervention development

Pilot stage research questions under EEF's three criteria may include (but are not limited to):

Evidence of promise

- *What perceived needs does the intervention address? What potential benefits do stakeholders (e.g. implementers) identify? Are these proposed outcomes considered to be socially significant?*
- *Is there preliminary evidence of the impact of the intervention on intended outcomes? Of what magnitude? For whom? Do there appear to be any unintended consequences or negative effects?*
- *Is there evidence to support the intervention logic model or theory of change?*

Feasibility

- *Is the intervention feasible? For example, can implementers deliver the intervention as intended in the time allotted?*
- *Are suggested activities congruent with the context of delivery (e.g. target population, setting)? Does any sequencing of content and other aspects of intervention design make sense to implementers and recipients?*
- *Is the intervention reaching its intended target population? Are recipients engaged during the delivery of the intervention?*

Readiness for trial

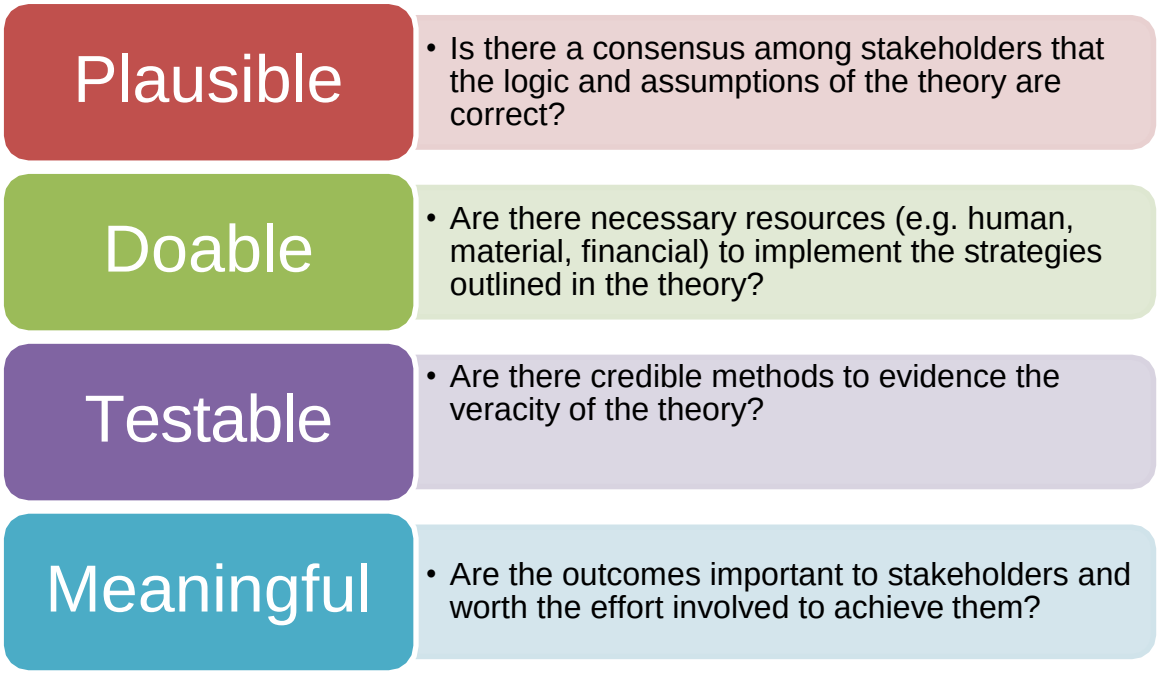
- *Are there identifiable 'components' within the intervention? How may these be assessed?*
- *What data indicators could be used to provide evidence of impact on intended outcomes?*
- *What can be learned from the above to take to the next stage of evaluation? For example, do changes need to be made to the intervention theory, materials and/or procedures? Are there any key contextual factors that appear to facilitate or impede successful implementation?*
- *Is the intervention considered to be affordable?*

While some early EEF pilots have included a control group (e.g. the Think Forward pilot), this will not be the case going forward. This is to avoid the over-interpretation of underpowered pilot trials, where the primary purpose should be about establishing evidence for the three criteria noted above. Instead EEF pilot evaluations might include some assessment of within-participant change and consider how best the outcomes referenced in the intervention logic model could be best assessed in a subsequent efficacy trial.

5.2 Generating evidence to assess intervention theory

The pilot stage provides a useful opportunity to interrogate the intervention theory (e.g. logic model or theory of change) and examine the extent to which emergent evaluation

evidence is congruent with it. A broad framework that may be useful for evaluators in this context is Connell and Klem's (2000) four criteria for the assessment of quality in theories of change. Using this, evaluators generate data to assess the extent to which the intervention theory is:



Clearly, there is overlap here with social validity principles outlined above. Alongside the above, the pilot stage provides opportunities to map out in detail the different sources of data that can be used to assess a given element of the intervention logic model or theory of change. Essentially, this involves examining each discrete stage and asking, 'how will we know if this is happening?'. This may be *relatively* straightforward when it comes to intervention inputs, evidence for which can be provided through a variety of methods as evaluators document implementation (see section 4.1 above). However, generating evidence to assess proposed mechanisms of change and/or interim outcomes is more complex, and will require evaluators to consider the use of various additional quantifiable indicators (e.g. attendance data), management information (e.g. service use data), bespoke quantitative instruments (e.g. depression survey) and/or qualitative data from intended beneficiaries (e.g. interviews with parents) in order to 'populate' the intervention theory with evidence. The same can also be said of the process of identification and mapping of contextual factors that influence different parts of the change processes triggered through the intervention. As with other aspects of IPE, the IDEA workshop (see section 3) will provide a helpful opportunity for this kind of mapping exercise.

10: Evaluators should map out the possible evidence indicators at each stage of the intervention theory and use the pilot evaluation to collect data so as to assess its veracity.

Further reading

Click the article title to access it online

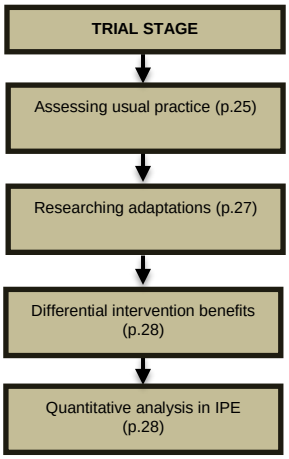
Kendal, S., Callery, P. & Keeley, P. (2011). The feasibility and acceptability of an approach to emotional wellbeing support for high school students. *Child and Adolescent Mental Health*, 16, 193-200.

This paper provides a useful case example of a qualitative, process-oriented pilot evaluation that focused primarily on acceptability and feasibility of a school-based emotional wellbeing intervention.

Connell, J. & Klem, A. (2000). You can get there from here: Using a theory of change approach to plan urban education reform. Journal of Educational and Psychological Consultation, 11, 93-120.

This paper outlines the theory of change approach to planning and evaluation of interventions. It provides a case example of an urban education reform initiative to demonstrate how, for example, evaluators can select appropriate indicators of change for different elements of the theory.

6 IPE in randomised trials



This section discusses best practice in IPE in randomised trials generally before focussing on elements of IPE that are more relevant in efficacy and effectiveness trials. IPE is critical for the interpretation of findings in trials, and, arguably, strengthens the ability to make causal inferences about the intervention's influence on outcomes.

> *"Changes in dependent variables can be attributed to the independent variable only when researchers can show accurate implementation" (Bruhn, Hirsch & Lloyd, 2015, p. 336)*

Exemplar research questions for evaluators at the trial stage of intervention development (whether efficacy or effectiveness)

Trial stage research questions may include (but are not limited to):

- How is the intervention implemented?
- How and why does the implementation (e.g. fidelity, dosage, reach) of the intervention vary?
- To what extent does this variability moderate intervention outcomes?
- Are there differential intervention benefits among pre-specified subgroups? If so, why?
- How is usual practice defined in relation to the intervention? To what extent and why does this change in the trial context, and does this influence outcomes?
- How and why is the intervention adapted, and to what extent do different adaptations moderate the achievement of expected outcomes?

6.1 Assessing usual practice

11: Evaluators should seek to assess usual practice in *all* trial schools prior to randomisation and at least once more at an appropriate future point in time (e.g. at post-test)

12: Quantitative usual practice data should be analysed in order to assess any changes that have occurred during the trial that may have implications for the

interpretation of trial outcomes (e.g. evidence of compensatory rivalry in control schools, or displacement in intervention schools)

Being able to define and document what constitutes 'usual practice' is a critical consideration for a number of reasons:

- Establishing the counterfactual is a necessary step in establishing causal effects
- Usual practice may change in response to randomisation to the control arm of a trial (e.g. compensatory rivalry, aka the 'John Henry' effect)
- It is important to establish the level of programme differentiation in the intervention group (e.g. how distinctive is the intervention? What has the intervention displaced?)

Unfortunately, documentation of usual practice – in terms of programme differentiation and/or monitoring of control conditions – is generally poor in intervention research. Typically, trials reports contain a short description with no indication of how this information was sourced, how much variability there was or whether any changes in practice occurred over time that may have influenced the trial outcomes.

How should usual practice be assessed?

As with other implementation dimensions, there are options here in terms of methods (e.g. observation, survey, document analysis). In a typical EEF trial, surveys are likely to offer the optimal balance between rigour and data burden. In designing quantitative usual practice surveys, evaluators are encouraged to consider the following issues:

- What is the appropriate level at which to survey usual practice (e.g. school or classroom/teacher?)
- What level of 'granularity' is required? For example:
    - usual practice in terms of named interventions in the same or related areas to the one being evaluated?
    - usual practice in terms of behaviours, strategies and approaches, curriculum content, and use of resources similar to aspects of the intervention being evaluated?
- What sources of information (e.g. existing research, evaluation team knowledge, delivery partner knowledge) are likely to be the most useful in developing a usual practice survey that is fit for purpose?

A copy of the usual practice survey developed for the EEF Good Behaviour Game trial can be found in Appendix 5

Qualitative usual practice data generates information on perceived similarity of a new intervention to previous practice. This is perhaps most useful for exploring the overlap between the new intervention and previous practices more generally, rather than a named intervention (see level of 'granularity' above). The level of similarity may be conceptualised as a moderating factor, in that a teacher who is already familiar with the concepts and processes of the new intervention may already possess the skills and self-efficacy required to deliver it effectively. Conversely, it may mean that this familiarity results in previous practices remaining in place, or that the new intervention is perceived as so similar that no change is required. Either of these may have implications for the acceptability of the

intervention and the way in which it is implemented. Questions within a semi-structured interview schedule relating specifically to usual practice could probe the extent to which implementers have changed their practice as a result of the new intervention.

6.2 Researching adaptations

Adaptations are seen as inevitable during the implementation of an intervention. While this is particularly true in effectiveness trial contexts due to the reduced developer control and increased range of implementation contexts, they can be expected at any stage of the intervention research process.

13: Evaluators should examine what adaptations are made, when, why, and the extent to which they are congruent with the intervention theory

How may adaptations be assessed?

The following is a useful framework for understanding and researching adaptations:

- Fit – are adaptations *logistical* (e.g. delivering an abbreviated version of a lesson in order to fit a pre-set timetable) or *philosophical* (e.g. changing intervention content where it is considered inappropriate for a given audience)?
- Timing – are adaptations *pro-active* (e.g. problems of fit to local context/need are anticipated and changes made in advance) or *reactive* (e.g. made 'on the fly' in response to issues that arise during the course of implementation)?
- Valence – are adaptations *positive* (e.g. in keeping with the goals and theory of the intervention), *neutral* (e.g. neither aligned with nor deviating from goals and theory) or *negative* (e.g. deviating from goals and theory) in nature? (Moore et al, 2013)

Quantitative assessment of adaptations may be integrated into the various methods and source noted in Table 2 above, in order to generate descriptive data and, potentially, assess the relationship between adaptation and outcomes. However, there are interpretive challenges posed here. For example, in the context of a structured observation, it may be possible to document adaptations that are made, but it will be difficult (if not impossible) to make accurate judgements as to whether these are made for logistical or philosophical reasons, or indeed whether they are pro-active or reactive.

Thus, in line with overarching principles of this guide, a mixed approach to the assessment of adaptations that incorporates both qualitative and quantitative data is optimal. Interviews are particularly useful for understanding *why* implementers have made adaptations. While they may be able to describe the changes that they have made, their reasons may only become clear with the further probing or questioning that is possible in semi-structured interviews. Specific questions around adaptations should aim to include additional probes to clarify whether they were pro-active or reactive, and made for philosophical or logistical reasons. It is important to note, however, that implementers may not always be *aware* of the adaptations that they are making, or their valence. Hence, skilled interpretation on the part of the evaluator will also be required.

6.3 Differential intervention benefits

From a technical standpoint, planned sub-group analyses form part of the outcome assessment protocol for a trial. However, conceptually they may be thought of as part of an IPE as the purpose they ultimately serve is to improve our understanding of how and why an intervention works, which includes knowing for whom it is more or less effective. As standard, EEF trials always include a subgroup analysis to determine if there are differential intervention benefits among disadvantaged pupils (defined using the EverFSM variable in the National Pupil Database). However, there may be other subgroups where theory and/or existing research suggests that additional analyses are warranted.

Are additional subgroup analyses (above and beyond FSM) warranted?

Clearly, a balance needs to be struck between advancing understanding of differential intervention benefits and the 'fishing problem' (mining trial data in the quest to find positive outcomes). Given this, any and all subgroup analyses should be planned (and thus included in the trial protocol), rather than being introduced post-hoc. In determining whether additional analyses (above and beyond EverFSM, such as gender, age and/or special educational needs) are warranted, evaluators should consider the extent to which a decision to focus on potential differential gains of a given subgroup is informed by theory (e.g. the intervention theory and/or the broader literature, such as risk and resilience theory), and/or informed by research (e.g. previous studies that have established differential intervention benefits).

14: All subgroups should be pre-specified and the choice of subgroups should be informed by the intervention theory and literature.

6.4 Quantitative analysis in IPE

The uses of quantitative IPE data can include documentation of implementation (see guidance point 8), on treatment analysis (see section 7.1), and critical component analysis (see 8.1). There is no single, agreed approach regarding the 'correct' way in which to model quantitative implementation data. The approach adopted will likely vary as a function of a variety of factors, including the design of the trial, the characteristics of the intervention itself, the nature and form of the implementation data collected, the skills and experience of the research team, and of course the intended purpose of the analysis (e.g. what are the research questions being addressed?). In the case of relational analyses designed to assess whether implementation variability predicts intervention outcomes, evaluators typically utilise regression-based analytical techniques (for example, multilevel modelling). Examples of methods used in the IPE literature can be found in the literature review that accompanies this guidance handbook (available here.)

Some general issues for evaluators to consider include:

- What is the appropriate level at which to model implementation data (e.g. school, classroom or pupil?)
- Is implementation data to be treated as continuous (e.g. percentage of lessons delivered) or categorical (e.g. low, moderate or high dosage)?
- Where implementation data is treated as categorical, what methods are used to determine category membership (e.g. statistical, such as place in overall distribution,

or externally imposed criterion, such as a threshold noted in the intervention materials or used in a previous study)?
- Will different implementation dimensions be treated as distinct, explanatory variables, or combined/reduced to provide a single proxy?
- Will control group data be used as a reference point in implementation analyses?

Whatever decisions are made in regard to the above issues, the analytical plan should be specified *a priori* in the same manner as is required for trial outcome analyses in order to avoid the 'data mining' problem.

15: Quantitative IPE analysis plans should be specified clearly in the trial protocol

16: Where quantitative implementation data is available, analyses should be undertaken to examine the relationship between implementation variability and intervention outcomes

Providing evidence that connects intervention implementation to outcomes can further increase the internal validity of a trial. This is particularly important given that we know that the implementation of a school-based intervention is likely to be highly variable, even in efficacy trial contexts.

Further reading

Click the article title to access it online

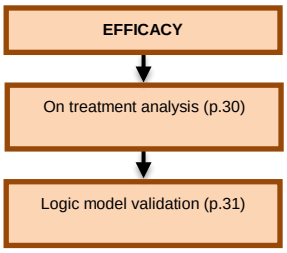
Askell-Williams, H., Dix, K. L., Lawson, M. J. & Slee, P.T. (2013). Quality of implementation of a school mental health initiative and changes over time on students' social and emotional competencies. School Effectiveness and School Improvement, 24, 357-381.

The above study illustrates the use of Latent Class Analysis to create distinct school profiles using data from a variety of sources and implementation dimensions, and the subsequent use of these profiles as explanatory variables used to predict student outcome variability in a hierarchical linear model.

Bonell, C., Fletcher, A., Morton, M., Lorenc, T. & Moore, L. (2012). Realist randomised controlled trials: A new approach to evaluating complex public health interventions. *Social Science & Medicine*, 75, 2299-2306.

This paper presents the rationale for an RCT model which – through incorporation of IPE approaches – is able to move evaluation beyond 'what works' to also ask 'for whom and under what circumstances', while also taking proper account of the complexity of social causation.

7 IPE in efficacy trials



At the efficacy trial stage, we examine whether an intervention achieves its intended outcomes under optimal conditions. In this context, trials tend to be at a smaller scale than effectiveness trials, enabling researchers to 'look inside the black box'. Above and beyond the general IPE considerations for trial-based evaluation, there are two particularly relevant aspects for evaluators to consider:

- On treatment analysis
- Logic model validation

Exemplar research questions for evaluators at the efficacy stage of intervention development and testing

In addition to those outlined in section 6, research questions at the efficacy stage may include (but are not limited to):

- *Do outcomes vary as a function of on-treatment status?*
- *Can the intervention logic model or theory of change be validated?*
- *What can be learned from the above to inform the effectiveness stage?*

7.1 On-treatment analysis

One approach to establishing the relationship between implementation variability and intervention outcomes is known as 'on-treatment' analysis (sometimes also called 'per protocol' or 'adherence to protocol'), in which only outcome data for participants known to have received the intervention as planned[3]. While often used in clinical trials, this kind of analysis can be difficult to apply in school-based evaluations because of the lack of established thresholds for describing the delivery of a given programme as 'on treatment' and the complex, multi-dimensional nature of implementation (for example, dosage may be per protocol but not fidelity/adherence).

Thus, if intending to conduct a formal on-treatment analysis, evaluators need to consider:

- What data is needed in order to determine whether implementation in a given class or school could be classified as 'on-treatment'?
- What is the appropriate threshold for on-treatment status?
    - Are there pre-existing thresholds that can be drawn from the programme materials, naturally occurring data (e.g. delivery partner accreditation or

[3] NB: 'As planned' may be interpreted in different ways according to the dimensions of implementation assessed for a given intervention.

similar), previous evaluations of the intervention, and/or the implementation science literature?
    - In the absence of (or in addition to) the above, can a threshold for the intervention be co-produced with the delivery partner?

An on-treatment status threshold should be articulated and agreed in advance of data collection (e.g. during the IDEA workshop – see section 3). At the most basic it could be a simple binary variable (completed intervention or not – with a clear and defensible description of the threshold for this classification). It can then applied to implementation data, such that evaluators can use it both *descriptively* (e.g. what proportion of trial schools/classrooms we classified as on-treatment?) and *comparatively* (e.g. are there differences in outcomes between on-treatment and off-treatment sites?).

17: Where appropriate and feasible, evaluators should conduct an on-treatment analysis alongside their intention-to-treat analysis. The on-treatment threshold should be articulated and agreed in advance of data collection in order to avoid data mining

7.2 Logic model validation

Another function of IPE in efficacy trials can be to inform empirical validation of the intervention logic model or theory of change. This may include modelling of mechanisms and processes that are thought to trigger change to demonstrate their association with outcomes, and/or exploring the temporal relationship between outcomes. The latter is particularly pertinent when the outcome of interest is distal to the intervention and change is assumed to be underpinned in some way by direct effects on proximal outcomes. This is particularly important at the efficacy trial stage as this will be the first opportunity to test assumptions about the factors associated with intervention outcomes.

How can evaluators evaluate intervention theory in a trial context?

One perspective on this question is that positive trial outcomes are an indication of the validity of the intervention theory in and of themselves. However, this makes numerous assumptions, including that the intervention was delivered as planned, and that change occurred in the manner originally theorised. Thus, a formal attempt to empirically validate the intervention theory is necessary. The IDEA workshop conducted in collaboration with the delivery partner during the planning stage can help to generate evidence indicators relating to both mechanisms and outcomes (proximal and distal) derived from the logic model or theory of change that can then be incorporated into the trial data collection and protocol.[4]

[4] This assumes, of course, that the IDEA workshop has taken place prior to the evaluation protocol being finalised. Where this is not possible for logistical reasons, evaluators should include a basic analysis plan that can then be modified or updated as necessary in agreement with the delivery team and EEF.

Further reading

Click the article title to access it online

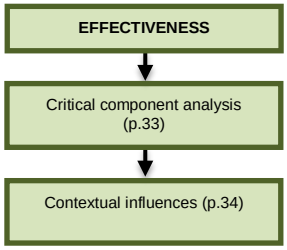
Berry, V. et al (2015). The effectiveness and micro-costing analysis of a universal, school-based, social-emotional learning programme in the UK: A cluster-randomised controlled trial. *School Mental Health*, Early View, 1-19.

This study provides a case example of the use of implementation data in a comparative 'on-treatment' outcome analysis. The authors set a threshold of 80% fidelity (sourced from coach/teacher ratings) as a means to compare the intervention outcomes of children in classrooms where implementation was deemed to fall below or exceed said threshold.

Banerjee, R., Weare, K. & Farr, W. (2014). Working with 'Social and Emotional Aspects of Learning' (SEAL): Associations with school ethos, pupil social experiences, attendance and attainment. *British Educational Research Journal*, 40, 718-742.

This paper analyses a proposed theory of change for the SEAL programme. Using multi-level path analysis of data drawn from a large number of schools and pupils, the authors demonstrated that implementation quality ratings were significantly associated with school ethos, which in turn mediated associations with pupils' social experiences, overall school attainment, and persistent absence.

8 IPE in effectiveness trials



If an efficacy trial asks whether an intervention *can* work, an effectiveness trial poses the question, “Will it work when delivered at scale?” Thus, there is a shift in emphasis from internal to external validity as the intervention is trialled in ‘real world’ conditions.

IPE remains a critical aspect of trial design at this stage. Some initial challenges for evaluators include:

- Rigorous documentation of the conditions of implementation in order to clearly articulate what is meant by ‘real world’
- Capturing the (inevitable) increased variability in implementation of the intervention
- Balancing the above with the need to maintain a ‘lighter touch’ so as not to confound the external validity of the trial (given that gathering implementation data is known to have a sensitizing effect on implementers)
- Examining a broader range of factors affecting implementation brought about by the effectiveness trial context (e.g. issues of funding and maintenance, and/or greater contextual variability)

Exemplar research questions for evaluators at the effectiveness stage of intervention development and testing

In addition to those outlined in section 6, research questions at the efficacy stage may include (but are not limited to):

- *What are the ‘real world’ conditions of implementation for the intervention being evaluated?*
- *To what extent does contextual variation influence implementation (and, subsequently, outcomes)?*
- *What can be learned from the above to inform the further scaling-up and dissemination process?*

8.1 Critical component analysis (CCA)

Critical intervention components (also known as core components or active ingredients) are those that actually make a difference to outcomes. Articulating and validating these components is a difficult and complex process. Interventions are often conceived of and evaluated in their entirety, making it difficult to isolate and manipulate individual components. However, knowing the active ingredients of an intervention is important as it enables us to focus our (sometimes scarce) resources on the right things, and increases the ability of implementers to make adaptations that are not likely to undermine outcomes. CCA can be thought of as an advanced version of on-treatment analysis. Where on-

treatment analysis conceptualises implementation as essentially binary (e.g. on-treatment vs. off-treatment), CCA offers a more nuanced view, concurrently examining the relationship between different intervention components (when present or absent to varying degrees) and outcomes.

> *"Research that focuses on operationalizing, measuring, and testing the efficacy of the independent variables (e.g., the core components) would improve our understanding of 'what works' and what is necessary for evidence-based programs and practices to produce outcomes" (Blasé & Fixsen, 2013, p.16)*

How might evaluators identify and validate critical intervention components?

Blasé and Fixsen (2013) suggest that we may think of critical components in terms of:

- contextual factors, such as the particular setting in which the intervention occurs
- structural elements, such as the obligatory number and/or sequence of sessions
- specific practices, such as techniques used to reinforce appropriate behaviour in a behaviour management intervention

The above framework could be applied during the IDEA workshop (see section 3) in an attempt to identify components that are hypothesised to be critical. For example, in relation to the structural elements noted above, part of the discussion could focus on the minimal dosage thought to be required in order to trigger change, and some exploration of the logic underpinning this (e.g. if it is proposed that 10 sessions is the minimum effective dosage, why is this the case?). Once the various proposed components (and the assumptions underlying them) have been mapped out, the discussion of how to determine their relative presence/absence in a given intervention site can take place. In terms of analysis, regression-based techniques that allow evaluators to assess the association between the extent to which various components are implemented and intervention outcomes are likely to be useful (see section 6.4).

In most trials, a CCA that follows the above process will likely rely on natural variation. However, for certain identified critical components, experimental manipulation may be possible. For example, consider an intervention delivered by teachers in which external coaching is utilised as part of the implementation support system. If coaching is considered to be critical (presumably because the coaching process facilitates higher quality implementation, and therefore better outcomes), the effects of its presence/absence on outcomes could be tested rigorously using an RCT design.

8.2 Contextual influences

One of the defining features of an effectiveness trial (when compared to an efficacy trial) is that the range of contexts of implementation is increased. IPE can help to establish, explore and explain contextual influences on intervention effects.

> *"Understanding the contexts in which complex interventions are delivered and received is critical to explaining why they do or do not work, or how we might expect impacts to differ if implemented elsewhere" (Moore et al, 2014, p.1)*

Some fundamental questions for evaluators include:

- What is the nature of the contextual variation being introduced at the effectiveness stage?
  - e.g. is the intervention being implemented in an increased range of settings and/or to a new/different population? What are the key indicators at the school, implementer and participant levels?
- How can this contextual variation be assessed?
  - e.g. through existing data, such as that contained in the NPD or EDUBASE, or through additional data?
- What level of association is of interest?
  - e.g. context-implementation-outcomes or simply context-implementation?
- Is the data to be used for summative or formative purposes?
  - e.g. establishing the magnitude of contextual moderation effects on implementation, or exploring what modifications are required in order to successfully replicate previously established intervention effects in a new context?

As with other aspects of IPE, an integrated, mixed-methods approach is likely to provide optimal insights in relation to contextual influences in effectiveness trials. For example, qualitative data drawn from key intervention stakeholders (e.g. implementers, recipients) may yield useful insights into the mechanisms underpinning any relationships established quantitatively (e.g. how and why do different contextual influences affect implementation and/or outcomes?), or help to map out the contextual influences at the school and classroom levels *ahead* of a quantitative investigation.

18: Evaluators should document clearly the contextual variation that is introduced in effectiveness trials, and where appropriate and feasible they should conduct analyses to explore context-implementation and/or context-implementation-outcomes associations.

8.3 Summary

Implementation and process evaluation refers to the generation and analysis of data to examine how an intervention is put into practice, how it operates to achieve its intended outcomes, and the factors that influence these processes. In the last couple of years there has been a greater emphasis on IPE at EEF and the production of this document demonstrates part of their investment in the importance of this aspect of intervention research. In it, the reader has been introduced to the eight dimensions of and five key factors thought to affect implementation that are evident in the IPE literature. Different sources of IPE data have been presented, and 18 key principles of IPE at the different stages of intervention research and development (e.g. pilot, efficacy, effectiveness) have been presented, alongside other important considerations for evaluators. Finally, in each section we have signposted sources of further reading.

It is our hope that this guidance handbook is useful to EEF evaluators (and, indeed, other researchers) as they seek to develop a more comprehensive understanding not only of 'what works', but also how, why, in what contexts or circumstances, and for whom things work (Slavin, 2012). IPE can help us to answer these important questions by providing researchers with theoretical, methodological and analytical tools that enable insights into

the processes and mechanisms underpinning the impact (or lack thereof) of educational interventions.

Further reading

Click the article title to access it online

Blase, K. & Fixsen, D. (2013). Core intervention components: identifying and operationalizing what makes programs work. *ASPE Research Brief*, February.

This paper explores the rationale for critical component analysis and discusses challenges and processes related to identifying and validating the 'active ingredients' of interventions.

Stallard, P. et al (2014). Classroom-based cognitive behaviour therapy (FRIENDS): a cluster randomised controlled trial to Prevent Anxiety in Children through Education in Schools (PACES). *The Lancet Psychiatry*, 1, 185-192.

The above study provides a useful illustration of how major components of interventions can be experimentally manipulated to determine if they are critical to the achievement of intended outcomes. In this case, the authors tested whether the teacher-led implementation of a mental health intervention was as effective as health professional-led implementation.

Pas, E. T., Waasdorp, T. E., & Bradshaw, C. P. (2015). Examining contextual influences on classroom-based implementation of Positive Behavior Support Strategies: Findings from a randomized controlled effectiveness trial. *Prevention Science*, *16*, 1096-1106.

The authors of this study identified a range of school-level contextual factors (e.g. school size, behavioural disruptions) and teacher-level factors (perceptions of school organizational health and grade level taught) that were significantly associated with variability in PBS implementation quality.

References

American Institutes for Research (2014). *Good Behaviour Game: Fidelity checklist*. Washington, DC: AIR.

Borglin, G. (2015). The value of mixed methods for researching complex interventions. In D. A. Richards & I. R. Hallberg (Eds.), *Complex interventions in health: An overview of research methods* (pp. 29–45). New York, NY: Routledge.

Blase, K., & Fixsen, D. (2013). Core intervention components: Identifying and operationalizing what makes programs work. *ASPE Research Brief*, *2*, 1–21.

Bruhn, A. L., Hirsch, S. E., & Lloyd, J. W. (2015). Treatment integrity in school-wide programs: A review of the literature (1993–2012). *The Journal of Primary Prevention*, *36*, 335–349.

Coffman, J. (1999). *Learning from logic models: an example of a family/school partnership model*. Harvard, MA: Harvard Graduate School of Education.

Connell, J., & Klem, A. (2000). You can get there from here: Using a theory of change approach to plan urban education reform. *Journal of Educational and Psychological Consultation*, *11*, 93–120.

Durlak, J. A., & DuPre, E. P. (2008). Implementation matters: a review of research on the influence of implementation on program outcomes and the factors affecting implementation. *American Journal of Community Psychology*, *41*, 327–50.

EPISCenter. (2014). *Promoting Alternative Thinking Strategies*. Retrieved from http://www.episcenter.psu.edu/ebp/altthinking

Evans, R., Scourfield, J., & Murphy, S. (2015). Pragmatic, formative process evaluations of complex interventions and why we need more of them. *Journal of Epidemiology and Community Health*, *69*, 925–6.

Fixsen, D., Naoom, S. F., Blase, K. a, Friedman, R. M., & Wallace, F. (2005). *Implementation research: a synthesis of the literature*. Tampa: University of South Florida.

Greenberg, M. T., & Kusche, C. A. (1993). *Promoting social and emotional development in deaf children: The PATHS Project.* Seattle: University of Washington Press.

Hansen, W. (2014). Measuring fidelity. In Z. Sloboda & H. Petras (Eds.), *Defining prevention science* (pp. 335–359). New York, NY: Springer.

Hoffmann, T. C., Glasziou, P. P., Boutron, I., Milne, R., Perera, R., Moher, D., Michie, S. (2014). Better reporting of interventions: template for intervention description and replication (TIDieR) checklist and guide. *BMJ (Clinical Research Ed.)*, *348*, 1687-1699.

Moore, J. E., Bumbarger, B. K. & Cooper, B. R. (2013) Examining adaptations of evidence-based programs in natural contexts. *Journal of Primary Prevention*, 34, 147-61.

Moore, G., Audrey, S., Barker, M., Bond, L., Bonell, C., Cooper, C., Baird, J. (2014). Process evaluation in complex public health intervention studies: the need for guidance. *Journal of Epidemiology and Community Health*, *68*, 101–2.

Ogden, T., & Fixsen, D. L. (2014). Implementation science: A brief overview and a look ahead. *Zeitschrift Für Psychologie*, *222*, 4–11.

Reupert, A. & Woodcock, S. (2010). Success and near misses: Pre-service teachers' use, confidence and success in various classroom management approaches. *Teaching and Teacher Education*, 26, 1261-1268.

Slavin, R. E. (2012). Foreword. In B. Kelly & D. F. Perkins (Eds.), *Handbook of implementation science for psychology in education* (p. xv). Cambridge: Cambridge University Press.

Stinchcomb, J. (2001). Using logic modeling to focus evaluation efforts. *Journal of Offender Rehabilitation*, *33*, 47–56.

Wolf, M. M. (1978). Social validity: the case for subjective measurement. *Journal of Applied Behavior Analysis*, *11*, 203–214.

Appendix 1: EEF adapted TIDieR framework

Completed example: Promoting Alternative Thinking Strategies (PATHS)

1. Brief name

Promoting Alternative Thinking Strategies (PATHS)

2. Why (rationale/theory)

Improved social-emotional competence is theorised to confer resilience to the onset, maintenance or progression of mental health difficulties (EPIS Centre, 2014). PATHS is based on the Affective-Behavioural-Cognitive-Developmental model of development, which emphasises the developmental integration of affect, emotion language, behaviour and cognitive understanding to promote social-emotional competence (Greenberg & Kusche, 1993). Core programme components are a taught curriculum, generalisation activities and techniques, and parent materials.

3. Who (recipients)

All children in a given class. The PATHS curriculum contains lessons for children throughout the primary phase of education.

4. What (materials)

Curriculum packs are provided for each class containing lessons and send-home activities that cover topics such as identifying and labelling feelings, controlling impulses, reducing stress and understanding other people's perspectives, in addition to associated physical resources and artifacts (e.g. posters, feelings dictionaries). In the current study, class teachers were also given an implementation guidance manual developed by the research team which emphasised the PATHS programme theory and the importance of effective implementation. PATHS curriculum packs are distributed in the UK by Barnardo's.

5. What (procedures)

PATHS lessons follow a common format that includes an introduction from the teacher (in which the lesson topic and objectives are introduced), a main activity (often built around a group activity or story), and a brief plenary/closure (in which learning is reviewed). Frequent prompts to elicit pupil responses and clarify learning are included throughout. The programme utilises a 'spiral' curriculum model, whereby (i) topics and concepts are revisited; (ii) units and lessons are developmentally sequenced; (iii) new learning is linked to previous learning; and (iv) the competence of learners increases with each successive visit to a topic or concept.

6. Who (implementers)

PATHS is designed to be delivered by class teachers.

7. How (mode of delivery)

Delivery of teacher-led PATHS lessons to all children is undertaken as part of the normal class timetable. Generalisation activities and strategies are implemented routinely throughout the school day.

8. Where (setting)

Regular classrooms in participating schools.

9. When and how much (dosage)

PATHS lessons last approximately 30-40 minutes and are designed to be delivered twice-weekly throughout the school year. Curriculum packs contain an average of 40 lessons.

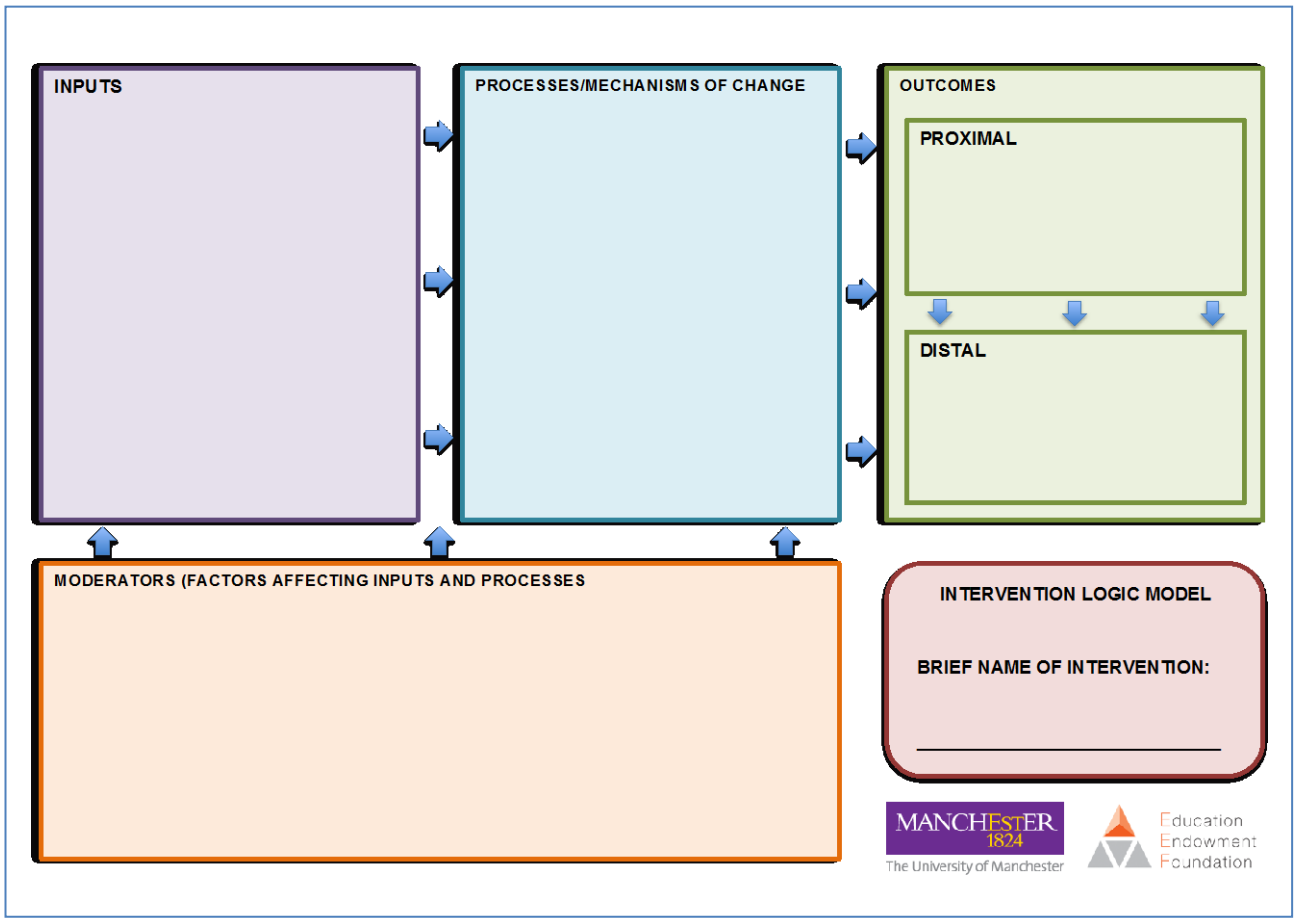
10. Tailoring

PATHS is a manualised intervention and optimal treatment fidelity is emphasised. Nonetheless, implementers are encouraged to make surface adaptations (e.g. changes of names in stories) in order to facilitate a sense of ownership and fit to local context.

11. How well (planned)

Strategies to maximise implementation effectiveness include one full day of initial training for teachers with a half-day follow-up approximately four months later. This is supplemented by technical support and assistance (e.g. lesson modelling, observation and feedback) from trained PATHS coaches.

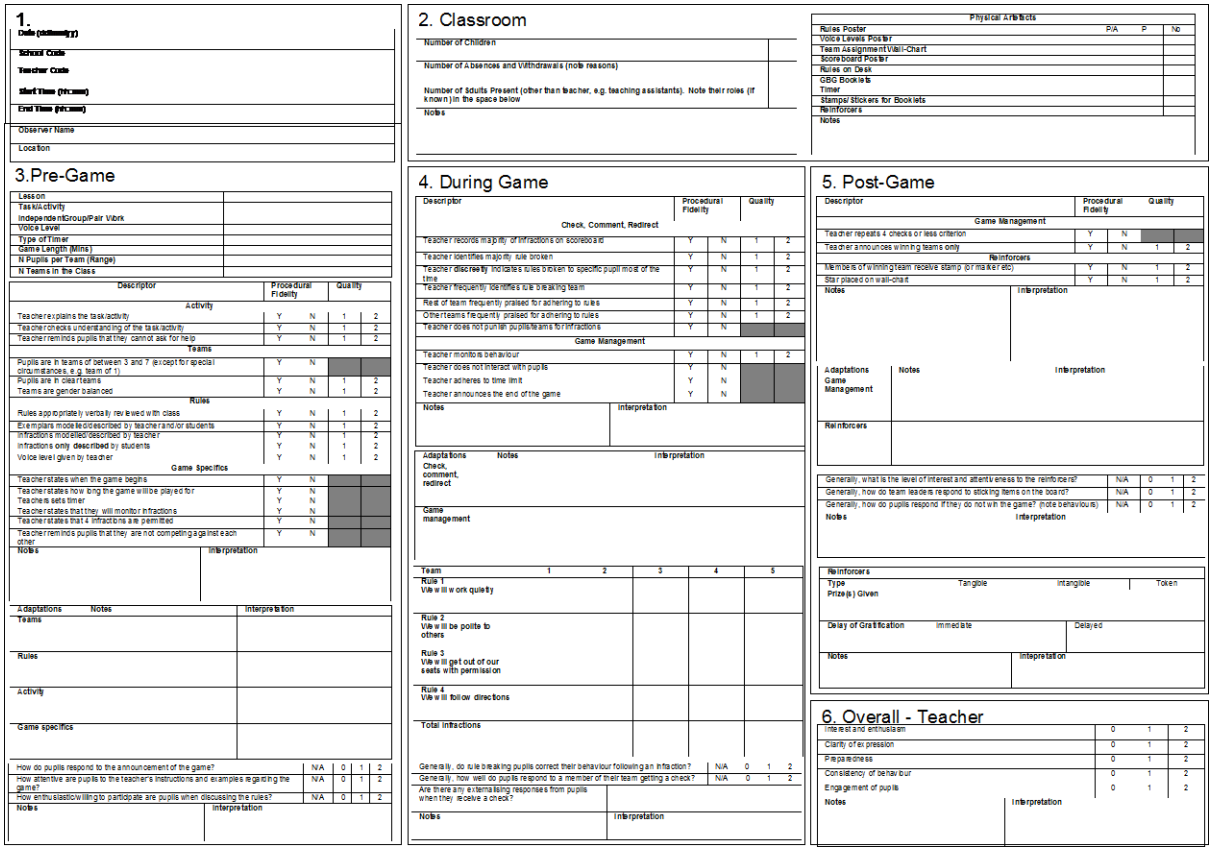
Appendix 2: Logic model template



A step-by-step guide to completing an intervention logic model can be found here.

Appendix 3: Exemplar structured observation schedule (GBG)

Developed by Neil Humphrey, Ann, Lendrum, Alexandra Barlow, Kirsty Frearson and Emma Ashworth. Adapted from GBG fidelity checklist published by the American Institutes for Research (2014).



1.

Date (dd/mm/yy)

School Code

Teacher Code

Start Time (hh:mm)

End Time (hh:mm)

Observer Name

Location

2. Classroom

Number of Children

Number of Absences and Withdrawals (note reasons)

Number of Adults Present (other than teacher, e.g. teaching assistants). Note their roles (if known) in the space below

Notes

Physical Artefacts

	P/A	P	No
Rules Poster			
Voice Levels Poster			
Team Assignment Wall-Chart			
Scoreboard Poster			
Rules on Desk			
GBG Booklets			
Timer			
Stamps/Stickers for Booklets			
Reinforcers			

Notes

3. Pre-Game

Lesson	
Task/Activity	
Independent/Group/Pair Work	
Voice Level	
Type of Timer	
Game Length (Mins)	
N Pupils per Team (Range)	
N Teams in the Class	

Descriptor	Procedural Fidelity		Quality	
Activity				
Teacher explains the task/activity	Y	N	1	2
Teacher checks understanding of the task/activity	Y	N	1	2
Teacher reminds pupils that they cannot ask for help	Y	N	1	2
Teams				
Pupils are in teams of between 3 and 7 (except for special circumstances, e.g. team of 1)	Y	N		
Pupils are in clear teams	Y	N	1	2
Teams are gender balanced	Y	N	1	2
Rules				
Rules appropriately verbally reviewed with class	Y	N	1	2
Exemplars modelled/described by teacher and/or students	Y	N	1	2
Infractions modelled/described by teacher	Y	N	1	2
Infractions only described by students	Y	N	1	2
Voice level given by teacher	Y	N	1	2
Game Specifics				
Teacher states when the game begins	Y	N		
Teacher states how long the game will be played for	Y	N		
Teacher sets timer	Y	N		
Teacher states that they will monitor infractions	Y	N		
Teacher states that 4 infractions are permitted	Y	N		
Teacher reminds pupils that they are not competing against each other	Y	N		

Notes

Interpretation

Adaptations	Notes	Interpretation
Teams		
Rules		
Activity		
Game specifics		

How do pupils respond to the announcement of the game?	N/A	0	1	2
How attentive are pupils to the teacher's instructions and examples regarding the game?	N/A	0	1	2
How enthusiastic/willing to participate are pupils when discussing the rules?	N/A	0	1	2

Notes

Interpretation

4. During Game

Descriptor	Procedural Fidelity		Quality	
Check, Comment, Redirect				
Teacher records majority of infractions on scoreboard	Y	N	1	2
Teacher identifies majority rule broken	Y	N	1	2
Teacher discreetly indicates rules broken to specific pupil most of the time	Y	N	1	2
Teacher frequently identifies rule breaking team	Y	N	1	2
Rest of team frequently praised for adhering to rules	Y	N	1	2
Other teams frequently praised for adhering to rules	Y	N	1	2
Teacher does not punish pupils/teams for infractions	Y	N		
Game Management				
Teacher monitors behaviour	Y	N	1	2
Teacher does not interact with pupils	Y	N		
Teacher adheres to time limit	Y	N		
Teacher announces the end of the game	Y	N		

Notes

Interpretation

Adaptations	Notes	Interpretation
Check, comment, redirect		
Game management		

Team	1	2	3	4	5
Rule 1 We will work quietly					
Rule 2 We will be polite to others					
Rule 3 We will get out of our seats with permission					
Rule 4 We will follow directions					
Total Infractions					

Generally, do rule breaking pupils correct their behaviour following an infraction?	N/A	0	1	2
Generally, how well do pupils respond to a member of their team getting a check?	N/A	0	1	2

Are there any externalising responses from pupils when they receive a check?

Notes

Interpretation

5. Post-Game

Descriptor	Procedural Fidelity		Quality	
Game Management				
Teacher repeats 4 checks or less criterion	Y	N		
Teacher announces winning teams only	Y	N	1	2
Reinforcers				
Members of winning team receive stamp (or marker etc)	Y	N	1	2
Star placed on wall-chart	Y	N	1	2

Notes

Interpretation

Adaptations	Notes	Interpretation
Game Management		
Reinforcers		

Generally, what is the level of interest and attentiveness to the reinforcers?	N/A	0	1	2
Generally, how do team leaders respond to sticking items on the board?	N/A	0	1	2
Generally, how do pupils respond if they do not win the game? (note behaviours)	N/A	0	1	2

Notes

Interpretation

Reinforcers

Type: Tangible | Intangible | Token

Prize(s) Given

Delay of Gratification: Immediate | Delayed

Notes

Interpretation

6. Overall - Teacher

Interest and enthusiasm	0	1	2
Clarity of expression	0	1	2
Preparedness	0	1	2
Consistency of behaviour	0	1	2
Engagement of pupils	0	1	2

Notes

Interpretation

Appendix 4: Exemplar implementation interview schedule (PATHS)

Developed by Ann Lendrum, Neil Humphrey, Craig Joyce, Kirsty Pert and Emma Stephens.

Interview data: aims

- To explore, understand and explain the processes of implementation of PATHS in English educational contexts
- To triangulate with and support interpretation of observation and impact data

Data needed for

- Examining processes of implementation
    - Fidelity – the extent to which the school is adhering to the intended treatment model
    - Dosage – how much of session delivered; number of sessions
    - Quality – how well different PATHS components are delivered
    - Participant responsiveness – the degree to which children and their parents engage with the intervention
    - Programme reach – rate and scope of participation
    - Monitoring of control conditions
    - Adaptation – the nature and extent of changes made to the intervention

- Identification of context specific factors affecting implementation
- Evaluating the feasibility of the future implementation of PATHS in English educational contexts

Preamble

- Check that the interviewee has received the information sheet and consent form and understands the project and his/her role in it.

Ask: Have you any questions about the project?

- Emphasise that:
    - The research team is speaking to a range of people involved in PATHS e.g. senior management, teachers, pupils at all of our (23) PATHS schools
    - We are interested in individual experiences and thoughts about PATHS, both positive and negative… "this is your opportunity to make your voice heard on PATHS … your comments may be helpful to others in your position at other schools at a later date"
    - However, we combine all the data we collect to provide an overall picture of PATHS and its implementation and any comments in the report are attributed very generally, for example, as "A (Year 3) teacher commented that…" Any comments/opinions will not be reported back to schools

Ask: Have you any questions about how we use your comments?

Ethics:

Remind interviewee:

- The interview will take about 30 minutes.
- You do not have to answer any questions that you are not comfortable with
- You can stop at any time, no explanation needed
- If any question doesn't make sense, ask for an explanation

Ask: Is it alright to record the interview? The transcript will only be seen by those working on the project. I will send you a copy too if you wish.

Ask: Are you able/willing to sign the consent form?

Explain procedure:

I will begin the interview with my name, the date, time and the identifying code we have assigned to your school - this is just to keep the recordings organised. All your details will be anonymised when the data is transcribed.

The first question will be about your role in school, followed by general questions about social and emotional learning in school, then moving on to PATHS more specifically

Ask: Have you any questions before we start?

Ask: Is it OK for me to start recording now?

Interview schedule

State researcher's name, date, time, school identifying code *(for data management)*

Can I just ask you to confirm your roles at school……..

……and in relation to PATHS *(e.g. Y3 teacher, co-ordinator, etc)*

A. Usual practice (Implementation - programme differentiation)

Aims: to clarify foundations for PATHS and school ethos round SEL; perceptions of benefits of PATHS/SEL; perceptions of need for PATHS/SEL; previous practice around social-emotional learning, whether starting PATHS has been integrated or resulted in changes to this.

1. Why has the school decided to implement PATHS?

Looking for information about:

- What sort of outcomes/change is the school aiming for? Is there a shared understanding?
- Are there specific needs within the school that PATHS is expected to address/meet?
- Whose decision was it to adopt PATHS?

2. What was done in school to develop social and emotional skills before you started doing PATHS?
   - Do you still do this?
   - Is/was this the whole school or just within your classroom?

Looking for information about:

a. Is PATHS part of a range of similar programmes/strategies? How does PATHS build on other local or national programmes/ interventions within school?
b. Has PATHS replaced previous programmes/ interventions/ approaches (eg SEAL)? Is it delivered alongside them? Are they integrated?
c. PATHS is just in KS2 – what does the rest of school do?

3. How would you describe the overall profile of PATHS in your school?
   - Is it just classroom teachers in Y3-5 (Y4-6) that are involved?
   - How involved is the headteacher? Senior management team?

Looking for information about:

a. Type of HT/SMT support
   - verbal only?
   - Active e.g. training time allowed, curriculum time allowed, included in planning *et cetera*

B. Implementation (dosage, fidelity/adaptation)

Aims: clarify implementation dosage and fidelity; modifications or adaptations and reasons for them; generalisation (link to quality)

4. How long have you been implementing PATHS?

5. How often do you teach PATHS? *Ask for example*

6. Is this a timetabled session? Same time every week?

Looking for:

- is timetabling is a problem
- (gently probe) status of PATHS? Competing priorities?

7. Are all pupils in the class present for PATHS?

Looking for:

- Participant reach
- Is the PATHS session used as withdrawal time? If so, do these pupils have PATHS at another time?
- Do some pupils have a more targeted approach eg SEAL small group work, nurture group? Is this in addition or instead of PATHS?

8. Are you able to cover all the lesson content in the time available? How long is a lesson on average?

    - Ask for examples

Looking for (probe gently):

    - Whether skipping content and why, e.g. competing priorities, lack of time, low status of PATHS

9. Have you repeated any lessons?

    - Ask for examples; gently probe reasons

10. Have you skipped any lessons?

    - Ask for examples; gently probe reasons

11. Have you been able to use PATHS outside of the specific lessons/ in other subjects (teachable moments)?

    - Ask for examples, e.g. which lessons, which concepts, in what ways?

12. Have you or the pupils been able to apply/generalise from PATHS in the classroom?

    - Ask for examples, e.g. do pupils use control signals, fingers linked, feelings faces, compliments, golden rule

13. Have you/the pupils been able to apply/generalise from PATHS outside the classroom, e.g. playtime?

    - Ask for examples, e.g. do pupils use control signals, fingers linked, feelings faces, compliments, golden rule

C. Attitudes to PATHS

Aims: clarify teacher and pupil attitudes to PATHS, including perceptions of impact; clarify fidelity and dosage, pupil responsiveness; describe and/or explain modifications or adaptations; inform interpretation of process data; inform future roll-out of PATHS in UK context

14. What do you think about the PATHS lessons and structure?
    - How useful do you find the lesson plans?
    - How much preparation is needed?

Ask for examples; probe for explanations, e.g. why like/don't like scripted lessons?
Looking for (gently probe):

- Do you follow the plans exactly? Make adaptations?
- Is it useful to have everything prepared? What would be more useful?

What do you think about the order of the lessons/structure of the programme? Have you changed the order around at all?

Looking for:

- Adaptations or changes
- Ask for examples – is this proactive, intended to enhance engagement and responsiveness? Is this reactive eg due to barriers (programme resources, lack of time?)

How familiar are the concepts, strategies?

Looking for:

- Changes to usual practice, foundations for PATHS

(*If Y4/5/6*) How useful/necessary were the Jump Start lessons? *Ask for examples*

15. What do you think about the PATHS resources *(if not included above)*

- How appropriate/suitable are the resources?

Ask for examples, e.g. availability of resources, age-level, particular class, SEN, emotional level, suitability for English context?

Looking for:

- How much do you adapt/make changes to the lessons?
- Ask for examples – want to identify whether this changes are proactive, intended to enhance engagement and responsiveness or reactive eg due to barriers (programme resources, time?)

How useful have you found the parent (send-home) activities?

How useful has PATHS been for meeting specific needs in your class?

- Ask for examples, e.g. improving relationships, empathy, inclusion, managing emotions, understanding feelings

Are there any aspects of PATHS that you have found particularly useful for your class?

- Ask for examples, e.g. pupil of the day, compliments, Golden Rules, talking about feelings, control signals, fingers linked

Are there any aspects of PATHS that you have found not useful/appropriate?

- Ask for examples, e.g. pupil of the day, compliments, Golden Rules, talking about feelings, control signals, fingers linked

16. What do the pupils in your class think about PATHS?

- Do they look forward to doing PATHS?
- Are they engaged by/do they enjoy the lessons?
- Are there any particular aspects they like?
- Are some groups more responsive than others *(eg SEN, EBD, quiet/withdrawn)*?
- PATHS has been designed for all the children in the class; have you found that it is useful for some groups more than others? *(eg EAL, SEN, EBD, withdrawn)*

Ask for specific examples (positive and negative)

17. Perceptions of impact: Has PATHS made a difference to your pupils? All pupils, or some groups of pupils particularly*?* The school more widely?

NB Acknowledge that may be too early to ask

- Ask for examples, e.g. improved relationships, social skills, understanding of emotions, behaviour, self-control, confidence and participation (eg quiet pupils more prepared to participate), classroom climate/ethos/atmosphere, learning, motivation for learning, attendance, SEN

D. Skills/knowledge/self-efficacy

Aims: teacher perceptions of self-efficacy, confidence, competence, skills and/or knowledge to implement PATHS; attitudes to training – quantity/quality, timing, content, utility etc; attitudes to support/coaching model - quantity/quality, timing, frequency, type of support available, utility etc. (NB to inform future roll-out)

18. Training: The PATHS programme provides a training package for teachers, with one day of training as you begin to deliver PATHS and a top-up half-day at the beginning of the second term. Were you able to attend the training?

Clarify whether first day, second (top-up) half-day or both

If yes:

- How useful did you find this? What particular aspects were useful?
- Was there anything missing?
- What additional/alternative training might have been useful?
- Were you required to 'cascade' the training/ brief colleagues?

Ask for examples (differentiate between first/second days)

If no:

- Did any colleagues attend the training? Who?
- Did they cascade the training/brief you on the training later at school?
- How useful was this?

Ask for examples (differentiate between first/second days)

19. *(If appropriate)* Are you planning to attend the second training day? *(why/why not?)*

20. Have you had any additional training relating to PATHS specifically?

21. Have you had any other opportunities for training/professional development around social and emotional learning?

    - Ask for examples

22. On-going support (coaching model): In addition to the initial training, the PATHS programme includes on-going support from a PATHS psychologist who has been assigned to your school. How useful have you found this on-going support?

23. Would you like to see more support?

24. Would you like to see less support?

25. Would you like to see different types of support?

    - Ask for examples

26. How important do you feel it is to have access to ongoing support?

Appendix 5: Exemplar usual practice survey (GBG)

Developed by Neil Humphrey, Ann Lendrum, Alexandra Barlow, Kirsty Frearson and Emma Ashworth. Sections 2-4 adapted from survey used by Reupert & Woodcock (2010).

This survey is designed to determine your 'usual practice' with regard to behaviour management in the classroom. The information you provide will be treated as anonymous and confidential. Please answer as honestly as possible. The survey should take no more than 10 minutes to complete.

There are five parts:

Part 1 asks general questions about you and your role in school.
Part 2 asks questions about general behaviour management strategies.
Part 3 looks at the reward systems, if any, that you have in place in your classroom.
Part 4 is about the ways in which you currently manage disruptive behaviour in the classroom.
Part 5 focuses on any proprietary interventions relating to behaviour management that you are currently implementing.

Part 1 – About You

1. Please specify your gender

Male ☐ Female ☐

2. How many years have you been practising as a qualified teacher?

_____

Part 2 – General Behaviour Management Approaches

Which of the following approaches do you currently use?

	No	Yes
I establish and maintain a set of classroom rules		
My pupils help to establish the rules of the classroom		
I use behaviour contracts		
I communicate clear expectations about rules and pupils' responsibilities e.g. through posters		
I give pupils positions of responsibility		
I alter the seating plan in my classroom as part of my behaviour management strategy		
I alter the curriculum to match pupils' interests and needs		
I promote good behaviour through PSHE lessons		
I use Circle Time to promote and help understanding of good behaviour		
I incorporate teaching of appropriate behaviours in lessons e.g. prosocial behaviours such as teamwork		
I attend behaviour management courses/CPD		
I use an anti-bullying policy		
I use buddying/peer mentoring techniques		
I use targeted behaviour management strategies for specific pupils		
I follow my school's behaviour policy		
I use the "silent and still" approach – stopping and waiting for pupils to respond		
I use signals e.g. clapping		
I use verbal redirection to engage pupils		
I reinforce our whole school behaviour policy/ethos/values		
I focus on good behaviour – "catch them doing the right thing"		
I observe and monitor pupils' behaviour in the classroom		
I respond to disruptive behaviour promptly		

Part 3 – Rewards Systems

How often do you use the following reward-based strategies?

	Never	Monthly	Weekly	Every day
I use a token reward system e.g. 'house points'				
I use an educational reward system e.g. free time, time on the computer				
I use prizes as rewards for good behaviour				
I use individual rewards				
I use group rewards				
I use whole class rewards				
I use special privileges				
I send notes/call/text parents about good behaviour				
I use praise and encouragement for good behaviour				
I/we hold assemblies in which good behaviour is recognised/rewarded, e.g. giving of certificates				

Part 4 – Managing Disruptive and Inappropriate Behaviour

How often do you use the following strategies for managing disruptive/inappropriate behaviour?

	Never	Monthly	Weekly	Every day
I ignore inappropriate/disruptive behaviours				
I use targeted small group work e.g. anger management				
I use vocal warnings e.g. raising/lowering voice, shouting				
I use body language e.g. frowning, physical proximity				
I remove privileges				
I use threats e.g. removal of rewards				
I use a warning/strike system				
I move pupils who are misbehaving to a different area of the classroom/make them stand up/send them out of the classroom				
I single out a child/group of children for misbehaviour				
I use a restorative justice system				
I use break-time supervision				
I use detention				
I use a behaviour report card				
I use a behaviour support base in the school				
I contact pupils' parents/carers				
I refer pupils to the head teacher/other professionals				
I send children home				

Part 5 – Proprietary Interventions

Which of the following proprietary interventions, programmes and approaches are you currently implementing in your classroom?

	Not implementing	**Just getting started**	**Well underway**	**Fully embedded**
Behaviour2Learn				
Cool In School				
The Motivated School				
Supportive Behaviour Management (SBM)				
Promoting Alternative Thinking Strategies (PATHS)				
Engage in Education (Catch22)				
Class DoJos				
Incredible Years Programme				
Behaviour For All				
IRIS				
5 Step				
Pivotal Education				
Humanutopia				
Inclusive Behavioural Programme (IBP)				
The Thrive Approach				
Behaviour Buddies				
Jogo Behaviour Support				
Achievement For All				
Nurture Groups				
Behaviour for Learning				
Other (Please Specify)				