Lessons learnt from EEF early years evaluations: Recommendations for evaluators and delivery teams

Version authors: Lauren Spinner and Aoife Duff

Contents

Version	Date	Reason for revision
2.1 **[*latest*]**	October 2025	- Parameters for power calculations updated on pg.6 - Table displaying parameters for EEF EY efficacy and effectiveness trials added to appendices (Appendix A) - Link to document reflecting on lessons learned from childminder pilot now included.
2.0	June 2025	Overhaul of paper to include lessons learned from recent EYs commissioning.
1.0 **[*original*]**	2019	N/A

Summary

This paper summarises lessons learned from the Education Endowment Foundation's (EEF) early years (EY) trials and provides practical recommendations for conducting EY evaluations, particularly randomised control trials (RCTs) with pre-reception children. The report includes five sections, covering the following recommendations for the planning, implementation, and management of EY evaluations.

1. Evaluation set up

- Identify a minimum sample size for the trial.
- Plan to recruit a slightly larger sample than needed to mitigate the risk of settings withdrawing pre-randomisation and use conservative assumptions when calculating the sample size required.
- Carefully consider the outcome measures selected for EY trials.
- Be mindful of the research burden on EY settings.
- Prepare alternative data collection methods.
- Plan according to the setting types in your sample.
- Consider examining longitudinal outcomes.
- Collect multiple child variables from the settings if matching to the National Pupil Database (NPD) or longitudinal tracking are expected, and for EEF data archiving purposes.
- Collect data on socioeconomic disadvantage.
- Consider stratification variables for randomisation.
- Be mindful of spillover effects if including settings from nursery groups or chains in the sample.
- Clearly define your analytical sample.
- Set a minimum threshold for the number of eligible children within a setting.
- In trials where a targeted intervention is being used, consider when and how eligible children are selected.

2. Recruitment

- Ensure sufficient time and resources are allocated for EY setting recruitment.
- Develop a strong recruitment team with efficient ways of working.
- Use a variety of approaches to promote the research project to settings to help reduce the risk of under-recruitment.
- Consider involving local authorities and other key local stakeholders in recruiting EY settings.
- Collate contact details for settings in the geographic areas you are targeting.
- Consider how to include a diverse range of setting types in your sample.
- Plan for the fact that PVI settings and maintained settings may operate differently.
- Adapt your recruitment strategy if targeting childminders.
- Provide clear, accessible recruitment documents.

- If time and resources allow, invite feedback on your recruitment documents drafts.
- Consider using multiple formats for consent forms.
- Ensure settings have fully understood the requirements of taking part in the evaluation.

3. Programme delivery period

- Monitor child attendance and movement to new settings during the delivery phase of the trial.
- Consider how attendance patterns may impact on intervention exposure.
- Consider facilitating engagement by offering incentives and contributions towards staff costs.
- Maintain strong communication channels with settings.
- Ensure programme delivery plans consider the context EY educators are working in.
- Bear in mind that some educators may be more comfortable participating in programme evaluations than others.
- If interviewing staff or observing EY practice, ensure the process and aims are clear to the settings and parents.

4. Pre-testing and preparing for randomisation

- Allow time for enumerating the sample in the autumn term.
- Give children time to settle in before testing.
- If necessary, consider concealed or batch randomisation as contingency measures.
- During recruitment or child enumeration, collect data on children's enrolled days/hours to inform testing schedules.

5. Considerations for both pre- and post-testing

- Be sensitive to children's needs around testing.
- Consider outsourcing some of the data collection if in-house resources are limited.
- Provide assessors with a randomised list of children to be assessed.
- Ensure a familiar member of staff is present or available during testing.
- Avoid narrow testing windows and plan additional 'mop up' visits.
- Avoid scheduling tests immediately before holidays.
- Provide settings with a testing schedule in advance.
- Have clear contingency plans if the evaluation includes the transition from nursery to reception.
- Ensure rigorous quality assurance procedures are in place, regardless of who completes the data collection.
- If assessments are to be administered online via a tablet, check with settings before testing begins that they can provide a Wi-Fi connection.
- Inform the EEF of any pressure points or emerging challenges as soon as possible so risk mitigation strategies can be agreed in good time.

Background

Early childhood education and care (ECEC) refers to education and childcare delivered by regulated providers such as childminders, nursery and reception classes, nursery schools, day nurseries, and preschools. In England, this covers the age range from birth to the end of the reception school year with 'early years' (EY) being the term often used to describe this key phase in children's learning and development. In England, parents are not required to engage their children in early childhood education: compulsory schooling starts the term after a child turns five.

Planning and implementing research projects with EY settings has all the challenges of education research in other key stages plus the complexities of working with a diverse variety of setting types, a relatively transient workforce, and very young children who may be still adjusting to attending the setting and whose attendance is non-compulsory. For RCTs, baseline testing is often needed due to the lack of prior education data and most testing with young children needs to be completed one to one. EY projects therefore need detailed planning and require careful management and monitoring from evaluators and delivery teams. However, given the paucity of research conducted in the early years compared to the primary school phase of education and the transformational role high quality education and care can have for children's development, there are significant benefits for educators and policymakers if we can gather robust EY evidence.

The EEF's remit was first extended to include pre-reception children in 2015 when the organisation's age range was expanded from 5 to 16 years old to 3 to 18. In 2022, it was extended further to include two-year-olds and, importantly, to explicitly include all types of provider of ECEC. Before 2022, the EEF had commissioned 30 evaluations focused on the EY phase. At the time of writing in 2025, a further 21 independent evaluations focused on the early years had been commissioned, substantially expanding the early years evidence base and representing over a quarter of the EEF's project commissioning in this period. Its commissioning since 2022 has included 17 projects targeting pre-reception children, more than doubling the number of studies conducted with this age group.

This work represents a big step forward in furthering the quantity and quality of EY research in England. Recent EEF commissioning has also included several pioneering ventures, including the first EEF evaluation targeting childminders, the first focused on two-year-olds within early years settings, the first where the primary outcome is a measure of personal, social, and emotional development (PSED), and the first 'choices' trial in EY settings.[1] Alongside this, all evaluations commissioned by the EEF for pre-reception age groups in group-based providers

[1] The EEF's 'choices trials' aim to produce causal evidence about the impact of different school-level approaches and policies, and teachers' classroom-level choices, on outcomes of interest.

within the last two years have included both maintained and private, voluntary, and independent (PVI) settings. This is a key research gap in England as most research has been implemented within the maintained sector, but the majority of children (including children from lower-income households) are accessing education and care in the PVI sector.[2]

In addition to the evaluations commissioned, from 2021 to 2025 the EEF has also supported 35 Early Stage Programme Development (ESPD) projects focused on the early years. ESPD projects support organisations to design and develop programmes that tackle educational disadvantage. A selection of the most successful projects developed through this work will be taken forward for an independent evaluation, helping to build the pipeline of evidenced-based programmes and support key areas of need for settings and children. This paper does not cover learnings from the ESPD projects but further information about this area of the EEF's work can be found on the **ESPD page** of the EEF website.

One of the catalysts for the sharp increase in the number of EY projects commissioned by the EEF since 2022 has been the EEF's collaboration with the Department for Education's early years Stronger Practice Hubs to support education recovery following the COVID-19 pandemic. While the appetite from educators to engage with support has been strong, the recovery environment has provided a challenging context to implement research as many settings are finding that their resources are more stretched than ever, with workforce recruitment and retention challenges being especially acute. Given this context, launching and implementing an unprecedented number of large-scale EY research projects has involved significant work from the EEF, independent evaluation teams, programme delivery teams, Stronger Practice Hubs, and early years settings, and much has been learnt from this work. The learnings from this period of rapid expansion in EY research projects will inform future strategic priorities at the EEF as it works to make a significant contribution to the evidence base on what works to support disadvantaged children.

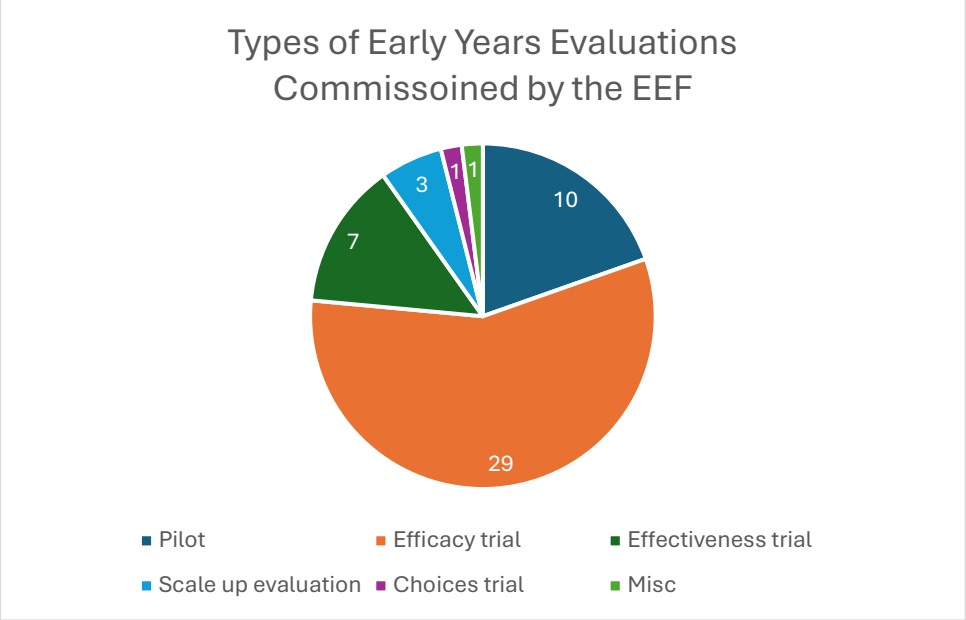
About this paper

In recognition of the substantial learning from the expansion of the EEF's EY commissioning, this paper provides an update to the 'Lessons Learnt from EEF's EY trials' paper published in 2019. The recommendations are based on the EEF's recent experience of funding and managing EY interventions and their subsequent evaluations as well as input collected from evaluation and delivery teams working on EEF-funded research projects. The paper aims to provide a set of focused, practical recommendations for delivery and evaluation teams planning to conduct EY independent evaluations, particularly with the EEF. It is intended as a summary of lessons learnt from past and live projects and does not aim to provide an exhaustive list of risks or mitigation

[2] DfE childcare and early years provider survey, 2023. Available at: https://explore-education-statistics.service.gov.uk/find-statistics/childcare-and-early-years-provider-survey

strategies. The recommendations in this paper focus on pre-reception evaluations, but some considerations will also apply to evaluations with reception-aged children.

The pie chart below summarises the types of EY evaluation commissioned by the EEF to date. Most EY evaluations that EEF has commissioned so far have been efficacy or effectiveness trials, usually RCTs that have been cluster-randomised at the setting level. The lessons learnt and recommendations in this document are therefore largely drawn from this methodology. However, we are also interested in adopting other methodologies where this is feasible and beneficial.



Recommendations

1. Evaluation set-up

For several reasons, designing evaluations in the early years is often more challenging than at later stages of education. For example, most EY trials will need to administer a baseline test when children have only just started at their setting due to the absence of earlier attainment data in administrative databases, and a lack of prior evidence means conservative assumptions will need to be used in sample size calculations, usually resulting in a higher number of settings needed. More settings are also likely to be needed because, often, not all children in a setting will participate in the evaluation. Therefore, cluster sizes tend to be smaller in EYs trials, resulting in higher minimum detectable effect sizes (MDES) that will be compensated for by increasing the number of settings. Outlined below are some additional considerations and recommendations to take on board when setting up EY evaluations.

- **Identify a minimum sample size for the trial**. Although to date most EEF trials have been powered to detect an MDES of 0.2 standard deviations, the EEF believes evaluations should be powered for an MDES that aligns with what is realistic and cost-effective,

considering prior evidence and the programme's Theory of Change (ToC). The appropriate MDES will depend on the nature of the programme being evaluated, with more resource-intensive programmes potentially requiring larger effect sizes to justify scaling them, while less intensive or light-touch programmes presenting smaller but still meaningful effect sizes. Moreover, and worth considering, there is some research to indicate that interventions which commence earlier in a child's development can lead to larger effects than those which commence later (e.g., Li et al., 2020), so you may reasonably expect to observe larger effect sizes in EYs than school trials. In addition to determining an 'ideal' sample size based on the 'ideal' MDES, evaluators should determine a lowest 'acceptable' sample size during the set-up stage, which may be lower than the 'ideal' sample size. This 'acceptable' sample size should be based on the largest acceptable MDES—that is, what both teams think is the maximum impact the programme could reasonably be expected to achieve with existing power calculation parameters. Alternatively, given that power calculations at study design stage tend to be based on conservative assumptions, teams could define the least conservative assumptions which they consider to be realistic, and thereby specify the minimum recruitment target they would be comfortable achieving. The 'acceptable' sample size should be used as a cut-off point to potentially stop the trial if there are difficulties with recruitment.

- **Plan to recruit a slightly larger sample than needed to mitigate the risk of settings withdrawing pre-randomisation and use conservative assumptions when calculating the sample size required. For the EY trials that have reported to date (excluding COVID-19 affected trials):**
    - The average (median) number of settings to either withdraw or be excluded from a trial pre-randomisation is 14, with a range from 0 - 32. Recruiting an additional 10% to 15% of settings to a waitlist is recommended so that settings that withdraw before randomisation can be replaced by settings on the waitlist.
    - Setting-level attrition post-randomisation has ranged from zero to 13%, with a mean of 3.5% and median of 3%, which is relatively low. Child-level attrition post-randomisation has ranged from 6% to 24%, with a mean of 16% and median of 14%. These figures suggest that pupil absence during endline testing (due to moving setting, sickness, or attendance patterns) is driving trial attrition rather than setting withdrawal. This should be considered when planning data collection, ensuring sufficient mop-up testing is included (further discussion on this can be found in the 'Considerations for Both Pre- and Post-Testing' section).
    - Setting-level ICCs have ranged from 0.02 to 0.35, with a mean and median of 0.17.
    - Child-level pre-post test correlations have ranged from 0.20 to 0.80, with a mean of 0.56 and a median of 0.58.
    - Cluster sizes have ranged from 5.6 to 29.6, with a mean of 13.3 and a median of 10.5 children per cluster.

- The combination of higher ICCs, lower pre/post-test correlations, and lower numbers of children per cluster observed here compared to school-based trials (on average), implies that more clusters are needed to achieve the required MDES compared to other key stages. Sample size decisions will depend on a number of constraints (delivery practicalities, burden on settings and educators, timeline, and cost) and will therefore need to be made on a pragmatic case-by-case basis. Note that the averages presented above include estimates from both pre-Reception and Reception trials. A breakdown of the estimates observed in EEF-published EY efficacy and effectiveness trials to-date can be viewed in the table in Appendix A.

- **Carefully consider the outcome measures selected for EY trials**:
  - Some measures may be suitable for use at endline but not at baseline due to the children being too young. Measures may not have been validated for children of baseline age and/or floor effects may be observed in baseline data. Piloting of proposed measures during the recruitment phase of the project can be beneficial to identify any issues before official data collection commences. It is also important to weigh up the pros and cons of including multiple secondary outcome measures, which can bring additional burden to settings and children and lead to fatigue or disengagement, affecting validity and the amount of missing data. Establish whether measures will require specialist administration (for example, from speech and language therapists) and what training assessors will need. Sometimes delivery teams are more familiar with the chosen measures than the evaluator, so factor in any time that the delivery team may need to contribute to training assessors or for the development of assessment guidance and resources.
  - One-on-one assessments with the child tend to be the most common data collection method for this age group. If administered by external assessors, they are expensive to administer. At baseline, however, some one-on-one assessments can be administered by educators if administration does not require extensive training or specialist expertise. Educator-report measures are completed by the educator based on their knowledge of the child. With these measures, keep in mind that educators may not know the child well enough at the start of the academic year to accurately report on them. At endline, if educators are not blind to condition, educator-report measures may not be suitable due to a potential high risk of bias.
  - Parent-report measures may be useful where knowledge of the child is required, but bear in mind that response rates may be lower than for other types of measures. When considering a parent-report measure it is important to refer to the programme's ToC and discuss with the delivery team whether impact of the programme is expected to be observed beyond the context of the EY setting during the trial period. EEF EY trials which have engaged with parents as part of the programme or the evaluation are **Peep Learning Together Programme**, **Early Years Conversation Project**, and **Emotion Coaching**.

- **Be mindful of the research burden on EY settings**. Educators will be very busy welcoming children to their settings in the autumn term and, for many children, starting nursery or reception will be a challenging time requiring significant support. If your evaluation involves educators administering tests to children or visiting them at home, ensure sufficient support is in place and timelines are realistic. This will have implications not only on participation and retention rates but also for the quality and reliability of the data collected. If the budget allows and working with a familiar member of staff is not necessary for children to complete the tests successfully, consider whether sending in your own researchers to complete the testing or outsourcing the service to a third party (for example, from EEF's data collection panel)[3] could reduce the research burden on early years educators. Administration of assessments by external assessors can reduce staff time needed for early years educators to assess children but may require additional time from them to schedule visits and facilitate testing, so pros and cons for EY educators should be weighed up alongside methodological considerations for the choice of assessment. Ensure you have a clear understanding of the time needed to complete all the evaluation activities and avoid placing more burden on children, settings, and parents than is necessary, including by selecting assessments that are quick and easy to administer. This is also in line with the minimisation principle (GDPR Article 5), which requires that data collection should be limited to only what is relevant and strictly necessary.
- **Prepare alternative data collection methods**. If planning to share materials and questionnaire links by email, or if programme implementation depends on accessing web resources, bear in mind not all settings have computer access or Wi-Fi and that not all EY educators will have advanced computer skills. Providing alternative options, such as paper versions of a survey, may help increase response rates. Planning for the fact that some EY educators will not have a workplace email address, or may share an institutional email address with colleagues, is particularly worthwhile and there may be data protection considerations here too regarding the use of personal email addresses.
- **Plan according to the setting types in your sample**. From an evaluation point of view, there are advantages in working with school-based nurseries, including lower staff turnover (and therefore children working with the same teaching staff over longer periods of time), children remaining in the same setting for longer periods, and availability of unique pupil numbers (UPNs), which are required for longitudinal impact tracking. Projects recruiting PVIs will need to be especially aware that (a) staff turnover is higher in these settings, (b)

[3] The EEF Data Collection Panel comprises organisations that have expressed interest in working with the EEF panel of evaluators to collect high quality data for EEF evaluations. They are independent of the EEF and have not been through an EEF vetting or recommendation process, but they do agree to Data Collection Panel terms and conditions. These T&Cs and the expressions of interest summarising the prior expertise and capacity of each Data Collection Panel member can be found **here**. EEF Evaluators should contact members of the Data Collection panel directly if they would like to work with them.

that they can sometimes close unexpectedly, and (c) that parents may be more likely to withdraw their children from these settings mid-year in order to move them to other nurseries. Children in PVIs will also move to a new setting when transitioning to reception, creating an additional challenge for trials that include the transition from nursery to reception. Despite the potential challenges of working with PVI settings, they bring much value to an evaluation, particularly an EEF evaluation, because PVI settings are more prevalent in areas of disadvantage and make up a large majority of EY providers overall, so it is important to ensure that they are represented. If engaging childminders in evaluations, teams should plan how and when to contact participants, considering their work patterns and the fact that they are more likely to be able to complete tasks associated with the evaluation outside of their working hours (see this note for more detailed reflections and recommendations on undertaking research with childminders). Overall, in recent EEF evaluations, PVI settings and childminders have been keen to participate in evaluation activities due to limited opportunities for this historically. It is important to consider whether or how implementation of the programme may differ in different setting types and how this will be captured in the evaluation (through the Implementation and Process Evaluation [IPE]).

- **Consider examining longitudinal outcomes**. Many EY programmes are centred around the professional development of educators and whole-setting change. Change in practice takes time to embed, so it may be that greater impact of the programme is observed in subsequent years following the initial delivery period. Longitudinal follow-up could be built into the evaluation either via collecting EYFSP data from the NPD once the cohort has completed reception year or by administering follow-up measures to staff, parents, or children directly. If collecting longitudinal data directly from participants, the follow-up sample may be much smaller than the sample from the main trial due to high rates of staff and child movement, especially if four-year-olds were included in the main trial and they have now moved on to school (which could be difficult to track). Longitudinal follow-up via the EYFSP will likely ensure greater retention of the sample but this measure would not capture the benefits of further embedding of staff practice if the children move to a reception setting having received the programme in their nursery at age three or four. EYFSP data is also likely to be linked to more distal outcomes than the primary outcome in the trial.
- **Collect multiple child variables from the settings if matching to the NPD or if longitudinal tracking is expected, and for EEF data archiving purposes**. Pupils in England are allocated UPNs at the point of first entry into the state-funded school sector (see DfE guidance **here**). This is typically when a pupil joins a maintained nursery or primary school (including joining nursery classes in a primary school) but can be later where entry to the state-funded school sector happens after this point. This means that children participating in early years trials, particularly those in PVIs, may not yet have been issued a UPN to facilitate matching to the NPD. In this case, evaluators should collect the

recommended set of identifiers below to facilitate subsequent matching to the NPD and archiving:

- first name;
- last name;
- home postcode;
- date of birth;
- unique pupil number (UPN) if available; and
- unique establishment number (URN) or LAESTAB if available, setting, local authority, or postcode.

Evaluators should submit the dataset containing these identifiers to FFT Education (FFT) at the end of the project. FFT will hold the data until participating children would be expected to appear in the NPD then match with NPD and replace these identifiers with the Pupil Matching Reference (PMR). The dataset containing the PMR (but no direct identifiers) will then be archived in the usual way. Data sharing procedures related to these variables and steps should be adequately described in privacy notices. For further information, see the **EEF's archive privacy notice**.

- **Collect data on socioeconomic disadvantage**. As with all EEF trials, a subgroup analysis should be performed on children from disadvantaged backgrounds. Free school meals (FSM) eligibility does not commence until children start school, so, in EY trials, disadvantage can be operationalised as a 9-month - to four-year-old who is in receipt of the Early Years Pupil Premium (EYPP). It may be particularly challenging to power evaluations for this subgroup because of lower uptake of EYPP compared to FSM, and due to potentially low numbers of disadvantaged children in smaller settings. The estimated average number of three- and four-year-olds registered for EYPP in each setting across England is 2.4 (the Department for Education reports that 116,500 three- and four-year-olds were in receipt of EYPP in 2022 across 47,121 providers: **https://explore-education-statistics.service.gov.uk/find-statistics/education-provision-children-under-5**). Note that EYPP was extended down to children as young as nine months in September 2024, in line with expanded childcare entitlements.

  While EEF efficacy trials are typically not powered to detect impacts on children eligible for EYPP, we do want to aim for our evaluations to provide meaningful information about this subgroup. Effectiveness trials should aim to power for this subgroup.

  - During set-up meetings it would be useful to discuss how you might maximise the size of the EYPP subgroup in the trial. This could include prioritising children eligible for EYPP if testing a smaller subset of children within the setting rather than selecting children randomly, and whether to include eligibility criteria around disadvantage at the setting

or pupil level (teams should be cautious of recruitment challenges when considering this).

- Teams should also consider during the set-up phase how their recruitment strategy could help to maximise the size of the EYPP subgroup. For instance, some teams have prioritised proactive recruitment efforts in local areas in deciles one to four of the Income Deprivation Affecting Children Index (IDACI) which are categorised as 'deprived' and 'most deprived' areas. Local area deprivation levels do not always mirror deprivation levels within specific settings in a given area, but in the absence of publicly available datasets on numbers eligible for EYPP by setting, IDACI scores can act as a useful proxy. See the 'Recruitment' section below for further information on available datasets and accessing IDACI scores.
- **When to collect**: EY providers are responsible for identifying children who might potentially be eligible for EYPP, however, parents and carers are the ones who must make a request to the local authority for their child's eligibility to be checked. EY providers facilitate this process by supporting families to complete necessary forms. Often this is a tick box giving consent for providers to conduct an EYPP eligibility check. From September 2024, any eligible child over nine months of age would receive EYPP from the beginning of the term after they apply. Eligibility must be checked annually. The number of children identified as eligible for and receiving EYPP will grow throughout the year as it takes time for all the administration to be processed after children join a setting. Therefore, the most reliable time to collect EYPP data from settings is likely to be once programme delivery has ceased, at or around the time of endline testing, however this data would need to be collected earlier (enumeration/baseline) if evaluators wish to examine EYPP patterns within settings and/or across the sample. In this case, we would recommend re-collecting EYPP data from settings at endline to ensure this information is up to date for analysis purposes.

- **Consider stratification variables for randomisation**. Evaluators may like to consider 'setting type' as a stratification variable to ensure the number of maintained and PVI settings are equally split across conditions. 'Region' may also be relevant in some cases, especially if it is important to ensure that there is an equal number of control and intervention settings within each region for practical delivery purposes.
- **Be mindful of spillover effects if including settings from nursery groups or chains in the sample**. Some settings are part of multi-academy trusts, PVI chains, or are affiliated with other middle-tier actors that support groups of settings. It is important when recruiting that the delivery team finds out if and how often staff move between settings that are part of a group because there could be spillover effects in a control setting if an educator from the intervention setting works there and, in addition, dilution of the intervention in their original setting. The approach taken to manage this on previous projects has been that instead of having a blanket rule about including or excluding these multi-chain settings, teams have assessed them on a case-by-case basis. So, in previous trials, the delivery

team has gathered information on staff movement: if settings did share staff regularly then only one setting in the chain was eligible to take part in the trial to avoid spillover effects or dilution. If settings did not share staff, then more than one setting in the chain was eligible to take part in the trial and each setting was randomised following the standard randomisation protocol for the trial.

The exact approach is likely to depend on the type of intervention and how substantial spillover and dilution effects due to staff movement are likely to be. For instance, there might be some interventions where staff movement is less likely to lead to spillover, such as interventions focused on parents or the home environment. In some cases, it may be possible to agree with the leadership of the nursery chain that staff do not move between settings within the chain during the period of the evaluation, although this would ultimately be outside of the evaluator's control to enforce.

- **Clearly define your analytical sample**. Agree at set-up whether the baseline test is an eligibility criterion for the child to participate in the trial—and if a child does not have a pre-test, but does have a post-test, will they be included in the ITT analysis? It is important to document this decision in the evaluation protocol as it will have important implications on how attrition is calculated at the end of the trial (and may, therefore, influence the security rating of the trial).
- **Set a minimum threshold for the number of eligible children within a setting**. For example, if ten children need to be recruited to the trial then you may require settings to have at least 15 eligible children to provide a buffer.
- **In trials where a targeted intervention is being used, consider when and how eligible children are selected**. It is worth discussing this in relation to the project timeline to make sure there is enough time at the start of September before testing or delivery begins. EY settings may not have a final register until the new academic year begins in September, so selecting children during the previous summer term can be difficult, because:
    - there tends to be a lot of movement of children between settings, so a child identified as eligible for an intervention in the summer term may move to another setting before testing or delivery begins in the autumn term; and
    - if educators are required to identify eligible children in September, many of them could be new to the setting meaning that the educator may need to spend a few weeks with them before they are able to identify them as eligible.

2. Recruitment

Recruiting settings and children for EY evaluations is challenging for several reasons. Staff capacity is very stretched in many EY settings so some feel unable to take on the requirements of participating in a research project and many who would like to take part struggle to find time to complete the sign-up process. Additionally, getting the word out to settings about the opportunity to be part of a research project is challenging as finding contact lists for settings can

be difficult and early years settings are not always as well networked as schools. The EEF's work and the concept of taking part in a research project may also be unfamiliar to some EY staff and to parents, who may be reluctant to sign up for projects that involve testing young children or that involve the possibility of being randomly assigned to a control group when they would like to be part of the intervention group. Considering these challenges, we would recommend the following to evaluation and delivery teams.

- **Ensure sufficient time and resources are allocated for EY recruitment**. Planning, drafting recruitment documents, and communicating with the settings will take time. This is the case for any evaluation but in our experience recruiting early years settings is often more time-intensive than recruiting primary and secondary schools. It is important to make sure that the team conducting recruitment have enough capacity to undertake this task. It is also advisable that at least six months are allowed for promoting the project to settings and getting settings signed up. If evaluation activities are scheduled to start at the beginning of the academic year, the recruitment period will coincide with the Easter and summer holidays, in addition to half-term breaks, during which PVIs may be open but school-based settings will be closed. Recruitment should ideally be completed by late June for programme evaluations starting in early autumn so that sufficient time is left to prepare for any training and baseline testing taking place once the new academic year has started.
- **Develop a strong recruitment team with efficient ways of working**. Successful teams ensure that the staff members working on recruiting settings have a shared understanding of how to promote the programme along with a good understanding of how the evaluation will work—for example, they are able to explain concepts like RCTs to settings and have excellent knowledge of the programme, and understanding of the context in early years settings. This ensures that the recruitment team can provide clear and accurate information to settings and can help them to understand how the project would work for their setting. Building personal connections with staff in settings can also be helpful, for instance, by having a named contact who consistently communicates with each setting. Effective ways of working within the recruitment team, clearly defined responsibilities, and careful tracking of engagement with settings are also key. Many teams set up shared email accounts and shared spreadsheets to track engagement with each setting, conversations with settings, and any concerns about the suitability of the setting for participating in the project. This helped to ensure continuity and promptness in follow-up with settings.
- **Use a variety of approaches to promote the research project to settings to help reduce the risk of under recruitment.** This could include:
  - directly contacting settings by email, phone, or mail;
  - using media and social media channels;
  - promoting the project at events; and
- collaborating with local stakeholders to promote the project.

Teams that recruited most quickly in the EEF's latest wave of early years trials generally put an emphasis on directly contacting settings but accompanied this with a range of other approaches. It is also useful to consider the phasing of these activities, for instance, early in the recruitment process the priority is to get the word out about the project and to generate expressions of interest. To support this, the launch of recruitment can provide an opportunity to make a splash with a news piece, and engaging key stakeholders who can support with recruitment early on can lay the groundwork for successful recruitment. After initial promotion work, the priority may shift from generating interest to converting expressions of interest into full registration to take part. At this stage, information sessions or phone calls for settings considering taking part may be useful ways to move settings forward through this process. Settings often require multiple 'touchpoints' before they will submit an MOU (such as initial contact via a newsletter, follow up emails to explain next steps or answer questions, and a phone call).

- **Consider involving local authorities and other key local stakeholders in recruiting EY settings,** such as local authority early years leads, Stronger Practice Hubs and Teaching School Hubs. Given their roles in convening educators and signposting them to relevant activities, these stakeholders can provide valuable support in disseminating information or facilitating access to settings via meetings or events. They may also be able to advise on whether there are any other local initiatives underway that could impact on the project such as other interventions that are recruiting or delivering in the area or local roll-outs of particular assessment tools, which may be relevant to consider as part of the evaluation design.
- **Collate contact details for settings in the geographic areas you are targeting** by drawing on published lists of early years settings in England. This enables targeted communications with settings likely to be eligible for the programme (note that promotional communications must adhere to GDPR and the ICO's direct marketing guidance). Collating contact lists for early years settings can be challenging, because there isn't a comprehensive central database that provides email addresses, phone numbers and postal addresses for EY settings in England. However, the following published datasets are a useful starting point.
    - The DfE's **Get Information About Schools** service provides a comprehensive list of schools in England that can be filtered by whether nursery provision is offered. It includes school addresses but not other contact details.
- The DfE's **Childcare Providers Register** provides similar information to the above service but covers other types of childcare provider. Addresses for non-domestic childcare providers are available in this dataset. Addresses are redacted for domestic-childcare providers. This register is also a useful resource for teams wishing to prioritise settings in disadvantaged areas as part of their recruitment strategy as it includes information about the setting's local level of deprivation based on the Income Deprivation Affecting Children

Index (IDACI). In contrast, the 'Get Information About Schools' dataset does not currently include information on settings' IDACI scores, however, the dataset does list each setting's Lower Layer Super Output Area (LSOA), so matching with other datasets that list IDACI scores by LSOA can be undertaken to check local area deprivation levels for school-based settings from this list.

  - The DfE's **Consented Address for Childminders and Domestic Childcare** provides information about domestic childcare providers who have consented to have their home addresses published.
  - The **Day Nurseries website** provides contact details for many nursery chains and groups.
  - Some local authorities also publish contact lists for early years settings in their area.

- **Consider how to include a diverse range of setting types in your sample**. If you are aiming to include a range of setting types within the evaluation, rather than focusing on evaluating a programme in a specific setting type, it may be useful to consider proportional targets for the types of settings recruited. For example, in recent EEF-funded EY trials teams have aimed for at least a third of recruited settings to be PVIs and at least a third to be maintained settings to ensure inclusion of both types. Another factor when considering the diversity of your sample is deciding the approach for nursery groups or chains. Targeting such organisations can be an efficient way to reach large numbers of settings, however, recruiting multiple settings that are part of the same group may reduce the diversity of the sample in the evaluation. Additionally, there can be risks of spillover effects, as outlined in the Set-Up section.
- **Plan for the fact that PVI settings and maintained settings often operate differently**. PVI settings may be open during school holiday periods and may have longer opening hours than maintained settings, meaning PVIs may be contactable at times when maintained settings would be closed. However, while school-based maintained settings often have dedicated staff members working on reception or providing administrative support, this is often not the case in some PVI settings. A positive implication of this is that communications with PVIs will sometimes go directly to key decision-makers, which can be conducive to an efficient recruitment process. On the other hand, PVI settings without dedicated administrative support may have limited capacity to respond to emails. Some recruitment teams from EEF-funded projects report that while emails have been effective for communicating with school-based maintained settings, phone calls have been more effective for communicating with PVI settings, although it is often advisable to avoid calls in the period around lunchtime (roughly 11am to 2pm) when adult-child ratios may be more stretched due to staff lunch breaks.
- **Adapt your recruitment strategy if targeting childminders**. Childminders may have limited flexibility to respond to communications or attend recruitment events during their working hours when they are likely to be busy caring for children. Offering phone calls or

recruitment information events in the evenings or at the weekends may help in this respect. Childminders often work independently or in small groups so the decision to proceed with registering for a project can often be made quickly without the need to consult wider staff teams. The limited availability of professional development opportunities for childminders may also contribute to prompt progress with recruitment as the opportunity to take part in a funded programme is likely to be an attractive prospect. However, getting the message out to childminders can be challenging. Childminder agencies are valuable stakeholders to engage and may be able to help with promoting opportunities to networks of childminders, although recruiting multiple childminders who are part of the same agency could reduce the diversity of the evaluation sample or pose a contamination risk. Many childminders are not part of an agency or formal network: approaches to reaching these could include advertising through local authorities, social media groups, or children's centres.

- **Provide clear, accessible recruitment documents**, explaining what participation entails for settings, parents, and children as well as the roles of the project teams and data processing details. Given that EY settings may be less familiar with educational research, it is very important that the enrolment form or Memorandum of Understanding (MOU), information sheets, and privacy notice are written and produced in an accessible style. Several evaluators have translated parent information sheets and consent or withdrawal forms into other languages (by recruiting university students to translate documents, for example, or using tools such as Weglot) to ensure parents/carers who have English as an additional language can still be informed about the trial and provide consent for their child to take part. It is also important to make it easy for settings to assess whether the programme being evaluated is right for their setting, staff, and children so clear information about who the programme is for, what needs it addresses, and what delivery involves is key. It will be important for providers to carefully consider how best to present the benefits of their programme for settings, EY educators, and the young children's outcomes. If planning to test or observe young children, explain how you will ensure this is done in a sensitive, age-appropriate manner. If specialists are involved (such as speech and language therapists), describe their role in an informative, reassuring manner. Helpful approaches to support with clear communication to settings could include producing video explainers that summarise key information and including tick lists on MOUs to remind staff of the key requirements they are signing up to and to check if they have understood these.

  If time and resources allow, invite feedback on your recruitment documents drafts from a few parents and EY educators willing to help in exchange for a small incentive.

- **Consider using multiple formats for consent forms**. Some evaluators have found it beneficial to use both paper-based and e-consent options when obtaining (opt-in) consent from parents/carers to provide optimal accessibility and maximise returns.

- **Ensure settings have fully understood the requirements of taking part in the evaluation** by inviting settings to a webinar or speaking to them over the phone before they sign an MOU. Lengthy recruitment documents can be challenging for busy setting leads to engage with so ensuring that there are opportunities to discuss before settings register is valuable for supporting retention. To ensure that key information has been understood, it may be useful to run through a checklist for settings or to ask questions that check for understanding, either as part of a phone conversation or on the MOU form. It is particularly useful to double check that settings meet the eligibility criteria. For instance, we have had several cases of early years settings signing MOUs for two different EEF evaluations despite both MOUs stating that settings are not eligible if they have already signed up to another evaluation that will be taking place at the same time. In some cases, two senior leaders have each signed the setting up for an evaluation without liaising with each other. To minimise administrative work associated with arranging phone calls, some teams have used online meeting booking tools that allow settings to book in a phone call for a time convenient to them.

3. The programme delivery period

Once settings have been recruited, the focus for delivery and evaluator teams shifts to preparing settings for the programme and evaluation activities, supporting them through delivery, monitoring how delivery is going, and retaining engagement through to the endline testing period. Effective communication, support, and monitoring are key to the success of this phase.

- **Monitor child attendance and movement to new settings during the delivery phase of the trial** so that a clear picture of likely attrition is obtained and can be used for planning purposes for endline testing. As we recommend intention to treat (ITT) analyses, endline testing all children who were baselined may be more feasible if it is known early on when children have moved settings. Some evaluators have monitored this at two specific points during the trial. As a strategy to minimise attrition, some evaluators have planned to follow up children at endline even if they leave their EY setting. This is primarily an option for trials using opt-in consent where, as part of the consent form, evaluators have asked parents/carers for permission to contact them or their child's new setting in order to still collect endline data.
- **Consider how attendance patterns may impact on intervention exposure**. Ensuring that children receive the intended level of exposure to an intervention can be challenging in EY evaluations due to children's varying attendance patterns. For example, if sessions of an intervention are due to be administered three times per week on a Monday, Wednesday, and Friday, some children may be in attendance on all of these days while others will not be, so their exposure to the intervention partly depends on their attendance pattern. Delivery teams should provide guidance to early years educators on how to approach intervention delivery in the context of varying attendance patterns. Some evaluators have attempted to use children's attendance patterns as a proxy for exposure, although

children's attendance data can be difficult to collect because this information is not recorded in a standardised way across settings and can be burdensome for educators to provide to evaluators: many settings, for example, still use paper records. Therefore, evaluators often obtain data on children's enrolment patterns and use this as a proxy for attendance and for exposure, although this will not be entirely accurate. However, in the pre-delivery phase of the **NELI Preschool trial**, the evaluator, the National Foundation for Educational Research (NFER), developed a 'Session Delivery Log' for educators to complete to record the number of whole-group, small-group, and one to one intervention sessions attended by children in their setting. The NFER has observed good completion rates of these logs with very few reported issues. The log template can be found in Appendix C (with permission from, and many thanks to, the NFER).

- **Consider facilitating engagement by offering incentives and contributions towards staff costs**. If budget allows, it is worth considering how incentives or other financial supplements can support participation and completion of research activities. Teams have sometimes found that chasing settings for payment information can be challenging so the process for settings to provide payment information needs to be simple and streamlined. For the EEF's recent early years evaluations, approaches to incentivising participation have included:
    - **Providing the programme to settings at a subsidised rate or free of charge**.
    - **Providing a contribution towards backfill costs for staff participating in training**. Participating in professional development activities can be difficult for early years staff because tight adult-child ratios, limited budgets, and lack of protected or paid time for professional development can make it difficult for settings to cover the costs of backfill when staff attend training related to a research project. As a result, recent EEF-funded evaluations targeting pre-reception age groups have provided a contribution towards these training costs. Where this is done, it is important for evaluators to include information on this in the cost evaluation. Backfill payments may contribute to higher attendance at training sessions or engagement with other elements of the programme so it is important for evaluations to be clear about the context in which the programme was provided.
    - **Offering incentives or thank-you payments for the completion of specific research activities**. Incentive amounts should be proportionate to the amount of work for settings in supporting research activities. In recent EEF EY trials, a typical incentive for the completion of baseline evaluation activities—this may include providing child data, facilitating child assessments, collecting parental consent, and completion of staff surveys—has been in the region of £150 per setting, while a typical incentive for completion of endline activities has been in the region of £250 per setting. Some teams have also included additional incentives for IPE activities where these are expected to be burdensome or where there are concerns about low response rates.

- **Considering incentives for parents participating in training or data collection**. If your project involves parents as research participants, bear in mind that engaging parents and ensuring a high response rate can require significant effort due to their multiple commitments and varying familiarity with research. You should carefully consider and budget for effective incentives for this group and plan how you will encourage participation and follow-up with non-responders. In line with the EEF's organisational mission, ensuring that parents from socioeconomically disadvantaged families are included and represented in research activities is especially important and financial incentives may facilitate this.

- **Maintain strong communication channels with settings**. While this is essential on all projects, some evaluators feel it is particularly important for EY evaluations. EY settings may be busier and less used to evaluation activities than other settings so regular, supportive, and straightforward communication is key. Build in multiple contacts in each EY setting to help mitigate high staff turnover and maintain engagement. This should include contact details for the staff members who will be directly involved with the project as well as contact details for a setting manager or other senior leader—buy in from senior leaders is often key for maintaining engagement. To keep in touch with these contacts, it can be useful to have a named project contact on the delivery and/or evaluation team who staff get to know leading on communication with the setting. Distributing regular project newsletters has been reported to be helpful (for example, once per month), especially where teams have tried to make these fun and engaging or used them to remind settings about upcoming project activities so that they can plan ahead. Some teams have sent greetings cards or short videos to settings to help keep communication varied. This is valuable both for the intervention settings and the control settings (that may not otherwise hear from the project teams for long periods and may become disengaged before post-testing). Creating a communications plan which is agreed between delivery teams and the evaluator is helpful so that is it clear who will be contacting settings when and for what purpose (see Appendix B for an example). This can prevent multiple messages being sent to settings at the same time, which can be confusing or overwhelming.

- **Ensure programme delivery plans take into account the context EY educators are working in**. Capacity is tight in many early years settings so it is helpful to provide a variety of dates and times for settings to engage in training sessions taking place as part of programme evaluations as well as ways for settings to catch up on sessions missed. Training multiple staff members in the setting may be necessary given staff turnover and the demands on individual staff members' time. For childminders, training sessions may need to be scheduled outside of usual working hours. If the evaluation or programme activities require settings to have access to technology, such as a strong Wi-Fi connection, a tablet, or a display screen, consider contingency plans for settings that do not have suitable technology in advance.

- **Bear in mind that some educators may be more comfortable participating in programme evaluations than others**. Evaluators working on EEF-funded early years projects have observed that while some early years professionals are confident taking part in a new training programme, others express feeling daunted by the prospect of large packs of training resources to digest, assignments to complete between training sessions, or being observed delivering activities with children. Access to training opportunities in nursery settings is often more limited than in schools, which may contribute to this. For example, in 2020, the Social Mobility Commission reported that only 17% of staff in early years settings reported having taken in part in training related to their work in the previous year.[4] Considering how professional development can be manageably sequenced and how support can be provided in an approachable way may be valuable for facilitating engagement with professional development programmes. For example, some teams have conducted a welcome visit or call with the staff who will be taking part in training (not just the setting leaders) at the start of the programme to facilitate this.
- **If interviewing staff or observing EY practice—for example, using Environment Rating Scales—ensure the process and aims are clear to the settings and parents**. EY staff may be reluctant to be observed if they are unsure about confidentiality or the aims of the observation (for instance, they may feel that *they*, rather than the programme, are being evaluated if this is not made explicit). Parents may also be concerned about safeguarding and data protection, especially if audio or video recordings will be used. Evaluation and delivery teams have also noted the importance of making sure that staff know how valued their expertise, knowledge of their setting, and experiences engaging with the programme are as part of the project. Some teams have also recommended conducting focus groups rather than one to one interviews in cases where they think staff may feel more comfortable participating alongside peers.

4. Pre-testing and preparing for randomisation

Given the lack of prior EY education data in administrative databases, the EEF recommends including a baseline test to reduce the MDES for a given sample size and to compare and control for differences in baseline attainment. However, a baseline test should not be added to the detriment of successful implementation of the training or intervention. If the delivery team believes that baseline testing would reduce the necessary delivery period to such an extent that the likelihood of finding an effect would be compromised, it would be preferable not to have a baseline. In this case, other data (socioeconomic indicators or other variables relevant to the intervention) could be used to reduce the sample size required for a given MDES and explore baseline imbalance and differential attrition. Evaluators may decide that the gains in statistical

[4] Social Mobility Commision (2020) *The stability of the early years workforce*. Available at: https://www.gov.uk/government/publications/the-stability-of-the-early-years-workforce-in-england/the-stability-of-the-early-years-workforce-in-england

power from administering primary or secondary baseline measures are not sufficient to justify the expense and burden on settings, educators, or parents/carers. If not using a baseline test, the evaluator should clarify the measures taken to mitigate the risks (such as using other covariates, more emphasis on multiple imputation, or restricted randomisation methods.)

Where a baseline is used, the key challenge (in addition to cost and research burden) will be time pressure. Programme delivery will often need to start as soon as possible in the academic year so that there is sufficient time for the intervention and any 'bedding in' period required before post-testing, but equally it is essential to allow plenty of time for enumerating children, conducting baseline tests, and conducting randomisation.

- **Allow time for enumerating the sample in the autumn term**. In EY trials, it is difficult to collect child data such as name or date of birth from settings and enumerate children before the autumn term because many of the children eligible for the trial will not join the setting until then. Enumeration at the start of the autumn term therefore needs to be factored into the timeline, which may condense baseline testing periods (if this is taking place). Some evaluators have found that two-year-olds often have a staggered introduction to their EY setting, which may lead to further delays in settings providing child data for this age group. Differences have also been found between setting types, with PVIs sometimes needing more time to provide child data than maintained settings due to differences in their data and technology systems: this may need to be factored into the approach taken to obtain data, with more time or a different strategy (for example, phone calls rather than emails) being employed for PVIs. Many EY settings will not be familiar with SharePoint or other cloud storage platforms where they may be asked to upload data: clear instructions on this should be provided by the evaluator.
- **Give children time to settle in before testing**. Testing children as soon as they have started in a new setting or moved to a different room, classroom, or key worker is not advisable as time is needed for them to settle in. Children's willingness to take part in assessments and the validity of the data obtained may be affected if children are unsettled. If setting staff are involved in completing baseline assessments, or in the recruitment or identification of children or families for the evaluation, they will need time to first get to know them to determine who is eligible and suitable to participate, and to be able to answer questions about them in educator-report measures. Children will also need time to get used to the new environment and staff before being comfortable interacting with any external assessors for one-to-one testing. It is recommended that at least three weeks pass between the children starting in their setting or transitioning rooms or classrooms and the start of baseline testing. Depending on the intervention timeline, October may be the best month to begin baseline testing. December should be avoided, if possible, due to settings being busy before the winter break and some children going on holiday early.
- **If necessary, consider concealed or batch randomisation as contingency measures**. While not advised in the first instance due to the design and administrative complexities

that these options bring, if more time is needed to conduct baseline testing, batch or concealed randomisation techniques could be employed as contingency options. The benefit of batch randomisation is that delivery teams can begin to contact intervention settings to arrange training sessions at an earlier point as there would be no need to wait for all settings to complete baseline testing before randomisation takes place. Instead of randomising the whole cohort at once, randomisation would occur in batches as settings complete testing.

With concealed randomisation, all settings would be randomised in one go but baseline testing would continue in a few settings after this point. Care must be taken not to reveal condition allocation to these settings until they have completed baseline testing. The benefit of this approach is that delivery teams can contact the majority of intervention settings in one go to arrange training dates and not have to wait for all settings to complete baseline testing; the few remaining settings can be contacted as and when they complete testing. This approach is most likely to be useful at the end of a term in that waiting for baseline assessments to come in from all settings would not leave enough time to contact all settings. With this option they could contact the majority of settings a week or two before the end of term and then the final few in the last week before the school holidays. The risk here is that settings that are randomised before they complete baseline testing (but not told of their allocation until afterwards) could withdraw from the trial before baseline testing is complete. These settings would be included in the attrition figures because they had been randomised.

Regardless of the randomisation approach, delivery team should aim to share potential training and delivery dates with *all* settings before randomisation so that settings can hold this time in their calendars and be available to take on the programme if they are allocated to the intervention condition.

- **During recruitment or child enumeration, collect data on children's enrolled days/hours to inform testing schedules**. This is to ensure testing is efficient and time in the settings is used wisely—you would want assessors to attend settings on days that most of the participating children will be present. This data should be collected before baseline testing begins and again before endline testing to inform planning in case enrolment patterns have changed between the two timepoints.

5. Considerations for both pre- and post-testing

Post-testing is perhaps the biggest challenge on all trials, but particularly on EY ones, due to the high risk of child-level attrition. In addition to the difficulties inherent in recruiting and retaining EY settings and children already mentioned, young children are more likely to move between settings during the year and take holidays during term time than older children as it is not compulsory for them to attend ECEC. However, some evaluators have found EY post-testing easier than pre-testing as children are slightly older and the settings have more planning time

and understand better what is expected of them. To ensure success across both testing periods, the EEF would recommend the following.

- **Be sensitive to children's needs around testing**. Decide in advance how you will respond to children with additional needs or those who are happy to be tested and then change their minds. It is critical that children do not feel coerced and understand clearly what they are asked to do. Some evaluators have found that using small incentives, such as stickers, can help, though not everyone will favour extrinsic motivators. It can be tempting to administer multiple tests during the same visit (for example, for primary and secondary outcomes), however, testing young children for longer than 20 minutes in one sitting should be avoided. Some evaluators recommend that, if testing were to take 30 minutes or more, it should be split into two sessions, which should be scheduled on different days if possible (rather than, say, in the morning and afternoon of the same day). This would, of course, increase the logistical challenge and the likelihood of setting and child drop-out. Multiple sittings on the same day and multiple testing dates for the same child should be considered very carefully due to the additional burden they place on the children and the setting. Evaluators may wish to pilot measures in a pre-trial phase with children outside of the identified sample to understand better the time taken to complete assessments and any barriers which may be revealed.
- **Consider outsourcing some of the data collection if in-house resources are limited**. As noted in the Evaluation Set-Up section, collaborating with organisations from the EEF's data collection panel or other test administrators can be considered if the additional benefit outweighs the likely higher cost. It is critical that test administrators are selected and trained very carefully. For instance, while supply teachers or PGCE students may be easy to recruit and provide a cost-effective option, it is very important they have the experience, training, and support necessary to understand the challenges of working with young children.
- **Provide assessors with a randomised list of children to be assessed**. To remove bias introduced by assessors selecting children not-at-random, assessors should be provided with a randomised class list by the evaluator and instructed to assess children in the order they appear on the list. This is particularly important if the intervention is universal or at a whole-class level and more children than are required for the evaluation will receive the intervention—for example, when 15 children are required to be part of the sample and complete measures but there are 30 children in the intervention group. If a randomised list is not provided, assessors may be tempted to select children who, for example, are more outgoing, reducing heterogeneity of the sample. Some children on the list may, of course, decline to take part and these wishes should be respected. In the whole-class example above, all 30 children (if eligible) could be included on the randomised list and the assessor could be instructed to work through the list in order, stopping when they have completed testing with 15 children (the required sample size). This allows for instances of child absence or refusal while still minimising the risk of bias. If the randomised list is deviated

from, assessors should record details of this so that evaluators are aware of the risk of bias. If a child is not present or available for their allotted assessment slot and it is a targeted intervention where all eligible children are included in the evaluation, the assessor should select the next child on the list and then return either during the same session or at a different time to assess the child who was skipped. As discussed in the earlier Set-Up section, for some evaluations the decision may be taken to prioritise EYPP children to maximise how many are in the sample. This approach means that the selection of children is not random, but selection does still follow a specific protocol and any bias introduced is therefore well understood (that is, the sample is deliberately biased to increase presence of EYPP pupils).

- **Ensure a familiar member of staff is present or available during testing**. The EEF's safeguarding guidelines require that testing is never unsupervised. An unsupervised environment means the researcher does not have an educator who would have safeguarding training and responsibility for children from the setting in the room. We do not expect any unsupervised activities to take place in EEF projects. If tests are administered by external staff, having a familiar educator around will also increase the children's comfort and willingness to participate in assessments. However, it is important that these staff members remain impartial during testing and do not introduce bias to responses, for example, by providing additional help on the assessment or nudging children towards particular responses.
- **Avoid narrow testing windows and plan additional 'mop up' visits**. In EY settings, some children are inevitably absent on testing days due to variations in patterns of attendance, holidays, or illness. Evaluators often find that EY testing takes longer than anticipated meaning that additional 'mop-up' testing visits are essential to reduce attrition. It is advisable to budget sufficient time for this and also to collect data from children who were absent when the assessors initially attend the setting. The type of intervention will need to be considered when calculating an appropriate amount of mop-up time. For targeted interventions, the cluster per setting is likely to be smaller than for universal interventions: for example, if eight children are eligible for a targeted intervention, for sufficient study power it would be important to ensure that data is collected from all eight. Factoring in mop-up may therefore be particularly important in these instances whereas in trials with universal interventions, assessors would likely have a much larger list of eligible children to select from on testing days and the same amount of mop-up may not be required at baseline. Sufficient mop-up at endline is arguably even more important due to the high stakes around the primary outcome measure, attrition, and security ratings of the trial. Recent EEF EY trials have, on average, had assessors visit each EY setting for two days at both baseline and endline, with additional mop-up of half a day to a day for 30% of settings. We would recommend budgeting for this amount of time as a minimum, and it is highly likely that mop-up time at endline will need to be higher than this to make sure as many children as possible are re-tested.

- **Avoid scheduling tests immediately before holidays**. July and December are usually high-risk months as settings are very busy before the summer and winter breaks with some children going on holiday early. All testing should ideally be completed by the end of June for a summer testing period and the end of November for a winter testing period.
- **Provide settings with a testing schedule in advance**, listing which children will be tested when, where, how, and by whom. This will need to take account of the setting's circumstances and children's usual attendance patterns: the schedule will, therefore, need to be agreed with the setting. It is important that setting leads are asked to share this information with all staff, including administrators, so that everyone is aware and to avoid confusion when assessors arrive. Teams have also found it useful to keep a log that is securely shared between the delivery team, evaluator, and testing organisation (if applicable) to record and monitor contact with settings and testing progress.
- **Have clear contingency plans if the evaluation includes the transition from nursery to reception**. If children move from their early years setting to a different school for the reception year, additional MoUs may need to be signed by the new schools if relying on them to provide child data for the evaluation. In addition, it may not always be possible to find out which school the child has moved to as the early years setting may be reluctant to share this information or may simply not know. Obtaining this information from the parents can also be difficult. Ways in which evaluators can reduce attrition between nursery and reception include gathering data from settings on the number of children likely to move school for reception, monitoring the likely moves as soon as this information is available, and planning for follow-up testing in new schools (particularly where children who have moved are clustered). In some trials, follow-up data has been collected from children in their homes or neutral settings such as a local library if visiting their new school has not been possible, though additional safeguarding risks may need to be considered in these instances.
- **Ensure rigorous quality assurance procedures are in place, regardless of who completes the data collection**. Some evaluators have found it useful to observe and audio-record the testing sessions and then rate assessment quality from a sample of recordings. This has enabled them to identify areas where assessors may not have followed the brief consistently and where more support is required. Ideally, this quality assurance would be included in the preparation phase to identify problems early, however, it is essential to have ongoing monitoring or recording during the actual outcome testing phase too. Acting quickly to ensure procedures improve for any further testing will be crucial. Evaluators and, in some cases, delivery teams—if familiar with administration of the outcome measure(s)—should be prepared to provide top-up training to assessors if a need is identified during quality assurance checks of the data or when issues are raised by settings. Checking assessment data as soon as it comes in is also recommended as a way to identify any issues as soon as possible and, ideally, take mitigating action before the testing has been completed (for example, through further training or briefings or additional

mop-up visits). Where variation in assessor quality is identified, evaluators can include a sensitivity analysis in the final report exploring whether assessor quality influenced the result in any way. If working with subcontracted testing organisations, evaluators should build clear expectations and milestones into their contracts with them. During data collection periods, the EEF requests weekly updates on testing progress and quality assurance findings from evaluators. Note that if evaluators intend to audio- or video-record testing sessions with children then this should be explicit in data privacy notices.

- **If assessments are to be administered online via a tablet, check with settings before testing begins that they are able to provide a Wi-Fi connection**, otherwise this could significantly impact the success of testing. If feasible, it may be advisable for assessors to take paper versions of the assessments with them too in case issues arise with the tablets or for evaluators to explore the option of using dongles to use cellular data on the tablets. Sometimes assessments are conducted offline via tablets but need to be connected to the internet afterwards for the data to be uploaded to the assessment platform or storage cloud. It is essential that assessors complete this step as soon as possible after completing an assessment to ensure data is not lost, and evaluators should work closely with subcontractors to ensure that specific checks are made before tablets are wiped.
- **Inform the EEF of any pressure points or emerging challenges as soon as possible so risk mitigation strategies can be agreed in good time**.

Conclusion

As evident from the numerous considerations and recommendations detailed in this paper, programme evaluations in EY settings are complex and require intentional and thorough planning. Despite the challenges, they are still critical for developing the evidence base around what EY practices and interventions are most effective at improving children's outcomes and attainment, particularly for disadvantaged groups.

We welcome feedback on this paper and advice on increasing the success of EY interventions and evaluations. Please send your feedback to **info@eefoundation.org.uk**, marking your email for the attention of the early years team.

Education Endowment Foundation

Appendix A: Parameters for published EEF EY efficacy and effectiveness trials

Programme (year)	Theme	Evaluation type	Age	Focus	*N* settings in analysis	Level of randomisation	ICC in report (setting-level)	MDES at analysis	Mean cluster size at analysis	Pre-post test correlation (pupil-level)	Pre-randomisation drop-out (inc. withdrawn and excluded)	Setting level attrition	Child level attrition
Concept Cat (2025)	Early Language	Efficacy	N	Whole class	89	Setting-level	0.14	0.18	10.1	0.6	1 setting	0.00%	13.30%
Counting Collections (2025)	Early Numeracy	Efficacy	R	Whole class	169	Setting-level	0.16	0.16	16	0.73	8 settings	5.56%	24%
EasyPeasy (2019)	Home Learning Environment	Efficacy	N	Whole class	101	Setting-level	0.18	0.20	10	0.52	0 settings	1%	6.4%
Family Skills (2018)	Early Language & Literacy skills of EAL pupils	Efficacy	R	Targeted	102	Setting-level	0.02	0.23	19.5	0.54	17 settings	11.3%	19.6%
Flexible Phonics (2022)	Early Language & Literacy	Efficacy	R	Whole class	118	Setting-level	0.13	0.17	22	0.54	18 settings	4.1%	20%
Glasses in Classes (2020)	Supporting glasses-wearing	Efficacy	R	Whole class	98	Setting-level	0.10	0.23	7	0.54	1 setting	1%	13.6%

Programme (year)	Theme	Evaluation type	Age	Focus	*N* settings in analysis	Level of randomisation	ICC in report (setting-level)	MDES at analysis	Mean cluster size at analysis	Pre-post test correlation (pupil-level)	Pre-randomisation drop-out (inc. withdrawn and excluded)	Setting level attrition	Child level attrition
Learning Language and Loving It™ (2023)	Early Language	Efficacy	N*	Whole class	138	Setting-level	0.17	0.21	13	0.76	7 settings	1.43%	12.8%
Maths Champions (first trial - 2018)	Early Maths	Effectiveness	N	Whole class	94	Setting-level	0.17	0.26	6.7	0.59	17 settings	13%	36%
Maths Champions (second trial - 2021)	Early Maths	Effectiveness	N	Whole class	134	Setting-level	0.16	0.19	9	0.61	19 settings	0%	7.3%
NELI (first trial - 2014)	Early Language & Literacy	Efficacy	R*	Targeted	34	Individual-level	N/A	0.25	N/A	0.55	0 settings	0%	11.2%
NELI (second trial - 2020)	Early Language & Literacy	Effectiveness	R	Targeted	192	Setting-level	0.35	0.26	5.6	0.75	14 settings	0.5%	7.3%
PACT (first trial - 2021)	Home Learning Environment	Efficacy	N	Targeted	45	Individual-level	0.23	0.17	10	0.67	5 settings	4.3%	22%

Programme (year)	Theme	Evaluation type	Age	Focus	*N* settings in analysis	Level of randomisation	ICC in report (setting-level)	MDES at analysis	Mean cluster size at analysis	Pre-post test correlation (pupil-level)	Pre-randomisation drop-out (inc. withdrawn and excluded)	Setting level attrition	Child level attrition
PACT (2024)	Home Learning Environment	Efficacy	N	Targeted	43	Individual-level	0.17	0.16	N/A	0.80	16 settings	0%	9%
ParentChild+ (2022)	Home Learning Environment	Efficacy	N	Targeted	223	Individual (family) - level	N/A	0.31	N/A	0.58	20 families	N/A	21.2%
Peep Learning Together Programme (2020)	Home Learning Environment	Efficacy	N	Whole class	134	Setting-level	0.25	0.27	8.6	0.37	32 settings	3.6%	20.2%
Reception Jigsaw (2021)	Early Maths	Efficacy	R	Whole class	138	Setting-level	0.24	0.21	17	0.59	22 settings	2.2%	13.8%
Success for All (2017)	Reading	Effectiveness	R & Yr 1	Whole class	52	Setting-level	0.13	0.31	29.6	0.20	0 settings	3.7%	24%
URLEY (2018)	Early Language & Literacy	Efficacy	N & R	Whole class	115	Setting-level	0.13	0.21	17	0.28	8 settings	4.2%	22%

Appendix B: Example communication plan for settings

Evaluation of the XX programme

Communication plan for settings—September 2024 to July 2025

Dates	Team sending email	Settings	Content
w/c 2 September		All	Staff enumeration email
2 September	e.g. Evaluator or delivery team	All	Parent letter + pupil enumeration email
10 September		All	Reminder email: parent letter + pupil enumeration
17 September		All	Baseline survey information and survey link
18 September		All	Reminder email 2: parent letter + pupil enumeration
20 September		All	Reminder email: Baseline survey
26 September		All	Reminder email 3: parent letter + pupil enumeration
27 September		All	Reminder email 2: Baseline survey
4 October		All	Reminder email 3: Baseline survey
10 October		All	Thank you email
14 October		All	Randomisation outcome email
w/c 14 & 21 October		Intervention settings	Send training invites
21 October		Sampled settings	Pre-implementation lead interviews invite
4 Nov		Sampled settings	Reminder invites for interviews
January		Intervention settings	Email to keep settings engaged, remind them to inform us in case they have any staff changes or drop-outs from training etc.
January		Comparison group settings	Email to keep settings engaged and remind them that we will contact them closer to data collection in the Summer
April		Intervention settings	Email listing key dates and activities for endline data collection
April		Comparison group settings	Email listing key dates and activities for endline data collection
May		All	Similar comms to April ahead of summer data collection period starting in June
June–July		All	Various comms to support data collection for the evaluation – details TBC

Appendix C: Example session delivery log

NELI Preschool Session Delivery Log

Setting Name:		NFER No:	

Please use this log to record the NELI Preschool sessions that have been attended by ALL children in your setting. The names and dates of birth of these children have been prepopulated below. These details cannot be edited on this sheet so please contact us directly if they need updating.

You will need to:
1. Record the week commencing date for each week the intervention is delivered in the pink box underneath the week number.
2. Record each child's attendance at the different session types for every week of delivery using the drop down buttons. Type 'Y' if a child has attended that session type on that day, otherwise leave the box blank. You should only record a child as present if they have attended the majority of a session.
3. If any child leaves the setting mid-way through the intervention, please record this in column F and add any additional details in column G.

						Week 1															Week 2														
						Please enter the week commencing date (dd/mm/yyyy) here:															Please enter the week commencing date (dd/mm/yyyy) here:														
						Whole Group Session					**Small Group Session**					**1 to 1 Session**					**Whole Group Session**					**Small Group Session**					**1 to 1 Session**				
Child ID	**Child's First Name**	**Child's Surname**	**DoB**	**If this child leaves your setting, please add their leaving date here.**	**Comment**	**M**	**T**	**W**	**T**	**F**	**M**	**T**	**W**	**T**	**F**	**M**	**T**	**W**	**T**	**F**	**M**	**T**	**W**	**T**	**F**	**M**	**T**	**W**	**T**	**F**	**M**	**T**	**W**	**T**	**F**