Lessons learnt from EEF Early Years trials:
Recommendations for evaluators

14 May 2019

Education Endowment Foundation

I.   Background

The Early Years (EY) key stage in England comprises the nursery year (ages 3-4) and the reception year (ages 4-5), with the Early Years Foundation Stage (EYFS) covering children from birth to 5 years old. Planning and completing evaluations in the EY has all the challenges of evaluations in other key stages, plus the constraints of working with very busy, often under-resourced settings, and very young children who are only just getting used to being at school. Baseline testing is often needed due to the lack of prior attainment data, and most testing with young children needs to be completed one-to-one. EY trials are therefore riskier, requiring particular support and attention from evaluators. However, given the paucity of relevant research and the importance of a positive start in life, there is a significant policy benefit to gathering robust EY evidence.

Since 2014, when the EEF expanded its remit to include EY settings, we have commissioned 17 EY projects, several more being in the process of appointing evaluators or setting up. The 4 EY trials that have reported to date have experienced high levels of pupil attrition: Nuffield ELI: 11%, Maths Champions: 36%, Success for All: 24%, Family Skills: 20%. We are working closely with our evaluators and developers to ensure lessons are learnt and appropriate contingency plans are in place for EY trials to be conducted as rigorously as practically possible, with minimum burden on the settings, pupils and parents.

In 2016, we commissioned a review of EY outcome measures and the UCL Institute of Education completed the work in 2017 – now available as an interactive measures database and downloadable report[1]. More recently, in 2018, we commissioned one of our former EY evaluation managers, Anneka Dawson (now at the Institute of Employment Studies, IES), to conduct a small study of evaluators and developers who had been involved in EEF EY evaluations. The aim of this study was to elicit experiences and perceptions of working on these evaluations, as well as challenges and solutions that the project teams had found helpful. The study was inevitably small-scale, given the relatively small number of EY trials we have

---

[1] https://educationendowmentfoundation.org.uk/projects-and-evaluation/evaluating-projects/early-years-measure-database/

commissioned to date, and the fact that some respondents were involved in more than one EEF EY evaluation. An online survey with a mix of closed and open questions elicited 10 responses from EEF evaluators, in addition to which the IES conducted telephone interviews with four evaluators and two developers. The developers interviewed were suggested by the EEF and the evaluators were those who responded to the online survey and were willing to take part in a follow-up interview (the other respondents did not provide contact details and therefore remained anonymous).

II. About this paper

This paper draws on the report produced by the IES (for internal EEF use), UCL's 2017 review, the EEF's experience of funding and managing EY interventions and their evaluations, as well as feedback received from evaluation teams, developers and EY settings. The paper aims to provide a set of focused, practical recommendations for teams planning to conduct EY evaluations for the EEF. It is intended as a concise summary of lessons learnt from past evaluations and does not aim to provide an exhaustive list of risks or mitigation strategies. We would be very grateful for any feedback on this paper and any 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 Evaluation team.

III. Key challenges and recommendations

Key challenges in EY evaluations tend to concentrate around recruitment and testing, given the difficulties of working with busy settings that may not necessarily be used to evaluation activities, the risks around testing very young children and the challenge of persuading teachers and parents that the benefits of the evaluation outweigh the research burden. We recommend the following aspects are considered carefully when planning and completing such evaluations.

1. Recruitment and set-up

Recruiting settings and parents for EY evaluations is generally more challenging than at later stages, for several reasons. Given the absence of earlier attainment data in administrative databases, most EY trials will need to administer a baseline test when pupils have only just started school. Lack of prior evidence mean conservative assumptions will need to be used in sample size calculations (e.g., higher intra-cluster correlations (ICCs), likely lower pre-test/ post-test correlations, smaller cluster sizes than in other key stages), resulting in a higher number of settings needed. More settings are likely to be needed also because, often, not all

children in a class will participate in the evaluation (therefore the cluster size will be even smaller, resulting in higher minimum detectable effect sizes (MDES) that will be compensated for by increasing the number of schools).

EY settings and parents can be reluctant to participate in projects that involve testing young children, as many think children are already tested too much and too early. Most EY tests will need to be administered one-to-one, putting additional pressure on already busy settings, especially if the tests are administered by the teachers themselves. Many EY settings will therefore be reluctant to sign up, or will be more likely to drop out later on. In settings with high staff turnover (particularly private, independent and voluntary settings), the person who signed the MoU and committed to the evaluation may be different from the person who is in charge by the time the evaluation starts, or ends. Parents may also be more likely to withdraw young children from data processing if concerns around safeguarding and data protection are not addressed sufficiently. A clear recruitment and communication strategy, with contingency plans, is essential.

- **Ensure sufficient resources are allocated for EY recruitment.** Planning, drafting recruitment documents and communicating with the settings will take time. It is advisable that at least six months are allowed for active recruitment to EY trials, wherever possible. If the evaluation is scheduled to start at the beginning of the school year, the recruitment period will coincide with the Easter and summer holidays, in addition to half-term breaks. While the summer months are difficult for recruitment regardless of the educational stage, most EY trials will need to start testing children as soon as possible after the start of the school year. Recruitment should ideally be completed by late June for interventions starting in early autumn, so that sufficient time is left to prepare for any training and baseline testing taking place as soon as the new school year has started. If your project involves parents as research participants, bear in mind evaluators consider them notoriously difficult to recruit, due to multiple commitments and varying familiarity with research.

- **Recruit a slightly larger sample than needed**, to mitigate the risk of attrition, if at all possible, and use conservative assumptions when calculating the sample size needed. An additional 10% is usually recommended for settings. Average pupil attrition on our EY trials that have reported so far is 23%. The EY trials that have reported to date have shown that the ICCs have been higher than expected, usually around 0.15. The combination of higher ICCs, lower pre/ post-test correlations and lower number of pupils per cluster imply that more clusters are required to achieve the required MDES, comparted to other key stages. Sample size decisions will depend on a number of constraints (delivery

practicalities, burden on settings and teachers, timeline, cost) and will therefore need to be made on a pragmatic case-by-case basis.

▪ **Be mindful of the research burden** on EY settings. Teachers will be very busy welcoming children to their schools in the autumn, and, for many children, starting nursery or reception will be a challenging time, requiring significant support. If your evaluation involves teachers administering tests to pupils 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 on the quality and reliability of the data collected. If the budget allows and working with a familiar member of staff is not necessary for pupils to complete the tests successfully, consider sending in your own researchers in to complete the testing, or outsource the service to a third party (e.g., from EEF's data collection panel), in order to minimise research burden on teachers. We have had schools drop out of EY trials due to the sheer amount of work they needed to complete during testing, which they had not realised when signing up. Ensure you have a clear understanding of the length of time needed to complete all the evaluation activities and avoid placing more burden on children, settings and parents than absolutely necessary. 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.

▪ **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 MoUs, information sheets and privacy notices are written in an accessible style. 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 (e.g., speech and language therapists), describe their role in an informative, reassuring manner. Audience awareness will be key. If time and resources allow, consider inviting feedback on your recruitment documents drafts from a few parents and EY teachers willing to help in exchange for a small incentive. (Critical details can be edited out before sharing the documents for this purpose.)

▪ **If observing EY practice (e.g., 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. Parents may also be concerned about safeguarding and data protection, especially if audio/ video recordings will be used.

- **Consider involving local authorities** in recruiting EY settings, if you have (or can obtain) the necessary contacts. Given their 'gatekeeper' role, they can easily disseminate information, support recruitment, and facilitate access to settings.

- **Maintain a good relationship with the delivery team and the settings**. Whilst this is essential on all projects, some evaluators feel it is particularly important on EY evaluations, given the nature and specific challenges of these projects, particularly around testing. EY settings may be busier and less used to evaluation activities than other settings, therefore clear regular communication is critical. Consider building in multiple contacts in each EY setting to help mitigate high staff turnover and maintain engagement.

- **Prepare alternative data collection methods**. If planning to share materials and questionnaire links by email, or if implementation depends on accessing web resources, bear in mind not all EY practitioners have email addresses (that they are happy to share), web/ computer access or advanced computer skills. Providing alternative options may help increase response rates and implementation fidelity. If planning to test children using computers or portable devices, ensure the test licence covers such uses, otherwise paper-based delivery may be needed. This needs to be clarified and arranged well in advance of the start of testing.

- **Consider offering incentives for parents** to participate in training or data collection. Depending on circumstances, shopping vouchers, book tokens or childcare support can help increase parental participation in projects that rely on this busy respondent group.

- **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, therefore children working with the same teaching staff over longer periods of time, children remaining in the same setting for longer periods, availability of unique pupil numbers (UPNs) required for longitudinal impact tracking. Projects recruiting private, voluntary and independent settings (PVIS) will need to be especially aware that staff turnover is higher in these settings, these schools can sometimes close unexpectedly, parents are more likely to withdraw their children from these settings mid-year in order to move them to school-based nurseries, UPNs are not typically available, therefore linking to the National Pupil Database (NPD) and longitudinal impact evaluation will be much more difficult to carry out. Pupils 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.

- **Collect multiple pupil variables from the settings** if matching to the NPD and/ or longitudinal tracking are expected. The DfE have confirmed they need a minimum of five

data points for high-quality pupil matching. These may include, depending on project and context, first name, last name, date of birth, school name or school ID, local authority, UPN.

- **Have named project contacts who communicate with the settings regularly**, throughout the project. This is very important both for the intervention schools and the control schools (who may not otherwise hear from the project teams for long periods and may therefore become disengaged before post-testing). Distributing regular project newsletters in collaboration with the delivery team has been found to help. It is important to remember EY settings are very busy and taking part in evaluations will not be their first priority.

2.   Baseline testing

Given the lack of prior EY attainment data in administrative databases, the EEF recommends including a baseline test to reduce the MDES for a given sample size, to compare and control for differences in baseline, and to estimate and control for differential attrition. However, a baseline test should not be added to the detriment of successful implementation. If the developer or delivery team consider that baseline testing would reduce the necessary implementation period to such an extent that the likelihood of finding an effect would be affected, it would be preferable not to have a baseline. In this case, other data (socio-economic indicators or other variables relevant to the intervention) could be used to reduce the MDES and explore baseline imbalance and differential attrition. If not using a baseline test, the evaluator should clarify the measures taken to mitigate the risks (e.g., using other covariates, more emphasis on multiple imputation, restricted randomisation methods.)

Where a baseline is used, the key challenge (in addition to cost and research burden) will be the time pressure. The project team will often be under pressure to start delivery as soon as possible after the children start nursery/ reception, so there is sufficient time for the intervention and any 'bedding in' period required before post-testing. However, testing children as soon as they have started in a new setting is not advisable. Especially if teachers are involved in baseline testing, or in recruitment of children or families for the evaluation, they will need time to first get to know the children and the families. Children will also need time to get used to the new environment and staff before being comfortable interacting with yet more new staff, often for one-to-one testing.

- **Give children time to settle in before testing**. It is recommended that at least three weeks pass between the children starting nursery/ reception and the start of baseline testing. Depending on the intervention timeline, October may be the best month for

baseline testing. December should be avoided, if possible, due to schools being busy before the winter break and some children going on holiday early.

- **Be sensitive to children's needs** around testing (both at pre-testing and post-testing). Decide in advance how you will respond to children with selective mutism, or children who are happy to be tested and then change their minds. It is critical 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 (e.g., 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 school and pupil 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 settings.

- **Ensure a familiar member of staff is present or available during testing**, whenever possible. If tests are administered by external staff, having a familiar teacher, teaching assistant or EY lead around will increase the children's comfort and willingness to participate in assessments.

- **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, and additional 'mop up' testing visits are essential to reduce attrition.

- **Clearly define your analytical sample**. Agree at set-up whether the baseline test is an eligibility criterion for the pupil to participate in the trial. (If a pupil doesn't 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 set-up minutes and 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).

3.   Post-testing

Post-testing is perhaps the biggest challenge on all trials, but particularly on EY ones, due to the high risk of attrition. In addition to the difficulties inherent in recruiting and retaining EY settings and pupils already mentioned, young children are more likely to move between

settings during the year than older pupils and it is not compulsory for them to be in school. Pupils in private and voluntary settings are more likely to move between settings and take holidays during term time than those in school-based nurseries, therefore additional care must be taken when planning to collect data in such settings. Trials that include the transition from nursery to reception will be likely to lose additional pupils that will be difficult to follow in their new settings. 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.

- **Allow sufficient time for planning post-testing**. Some evaluators have found they needed twice as much time to plan the post-testing than originally anticipated. If contracting staff for testing, it is essential this is arranged well before the testing is scheduled to start, to allow sufficient time for training and familiarisation with the project.

- **Avoid scheduling tests immediately before holidays**. July is usually a high-risk month, as schools are very busy before the summer break and many pupils will be going on holiday early. All testing should ideally be completed by the end of June.

- **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 school circumstances and children's usual attendance pattern, therefore the schedule will need to be agreed with the school.

- **Plan additional 'mop up' visits**. As with pre-testing, additional visits will need to be scheduled for pupils who are not at school the first time. This is arguably even more important at post-testing, due to the high stakes around the primary outcome measure. Evaluators have found that more 'mopping up' is required at post-testing than at baseline, between one third and half of the testing visits needing to be rearranged at post-test. To minimise attrition, this needs to be anticipated and built into the evaluation timeline, and communicated to settings in advance.

- **Have clear contingency plans if the intervention includes the transition from nursery to reception**. If pupils move from their nursery school 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 pupil data for the evaluation. In addition, it may not always be possible to find out which school the pupil has moved to, as the nursery school 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 schools on the number of children likely to move school for reception, monitoring the likely moves as soon as this information is available,

planning for follow-up testing in new schools (particularly where children who have moved are clustered).

- **Consider outsourcing some of the data collection** if in-house resources are limited. Collaborating with organisations from EEF's data collection panel may be a solution, 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.

- **Ensure rigorous quality assurance procedures are in place**, regardless of who completes the data collection. Some evaluators have found it useful to observe and/ or audio-record the testing sessions and then rate assessment quality from the sample of recordings. This has enabled them to identify areas where assessors may not have followed the brief consistently and required more support. Ideally, this quality assurance would be included in the preparation phase to identify problems early. However, it is also advisable to have ongoing monitoring/ recording during the actual outcome testing. Acting quickly to ensure procedures improve for any further testing will be essential. 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. Checking assessment data as soon as it comes in is also recommended by some evaluators as a way to identify any issues as soon as possible and, ideally, take mitigating action before the testing has been completed (e.g., further training or briefings, additional mop-up visits).

- **Inform the EEF of any pressure points or emerging challenges** as soon as possible, so risk mitigation strategies can be agreed in good time.