

# Effectiveness Trial of White Rose Maths Reception Jigsaw (2025-26), a Two-Arm Cluster Randomised Trial Evaluation Protocol



**Evaluator (institution): National Foundation for Educational Research (NFER)**

**Principal investigator(s): Pippa Lord**

Template last updated: December 2022

## Evaluation summary

<b>Project title</b>	Effectiveness Trial of White Rose Maths Reception Jigsaw (2025 – 26), a two-armed cluster randomised trial
<b>Developer</b> <i>(Institution)</i>	White Rose Education (WRE)
<b>Evaluator</b> <i>(Institution)</i>	National Foundation for Educational Research (NFER)
<b>Principal investigator(s)</b>	Pippa Lord
<b>Protocol author(s)</b>	Aarti Sahasranaman, Chris Morton, Katherine Aston, Pippa Lord
<b>Trial design</b>	Two-arm cluster randomised controlled trial with random allocation at school level
<b>Trial type</b>	Effectiveness
<b>Pupil age range and Key stage</b>	Age 4 – 5, Reception (Early Years)
<b>Number of schools</b> <i>(at design stage)</i>	330
<b>Number of pupils</b> <i>(at design stage)</i>	6,600
<b>Primary outcome measure and source</b>	Early numeracy skills (Early Years Toolbox Early Numeracy App, Howard et al., 2022)
<b>Secondary outcome measure and source</b>	Early mathematical skills (Early Years Foundation Stage Profile maths ELGs from National Pupil Database (NPD))

## Protocol version history

Version	Date	Reason for revision
1.3 [latest]		
1.1		
1.0 [original]	29/08/2025	NFER share with EEF

- Any changes to the design need to be discussed with the EEF Evaluation Manager and the developer team prior to any change(s) being finalised. Describe in the table above any agreed changes made to the trial design. Also, please reflect any changes in the body of the document where appropriate.
- Please ensure that any changes to the design of the trial that affect the analysis to be undertaken are also reflected in the SAP.

## Contents

Evaluation summary .....	1
Protocol version history .....	2
Study rationale and background .....	5
Intervention: White Rose Maths Reception Jigsaw .....	7
Impact evaluation design.....	16
Implementation and Process Evaluation (IPE) design.....	28
Cost evaluation design .....	35
Ethics and registration.....	36
Data protection.....	37
Personnel.....	40
Timeline.....	43
References.....	45
Appendix A: Changes since the previous EEF evaluation.....	47
Appendix B: Baseline Early Numeracy (EN) Checklist.....	48

## Tables

Table 1: Trial design .....	16
Table 2: Sample size calculations before accounting for anticipated attrition .	<b>Error! Bookmark not defined.</b>
Table 3: Sample size calculations with school and pupil numbers reduced to account for anticipated attrition .....	<b>Error! Bookmark not defined.</b>
Table 4 IPE methods overview IPE methods overview .....	<b>Error! Bookmark not defined.</b>
Table 5 Timeline .....	<b>Error! Bookmark not defined.</b>

## Figures

Figure 1: Logic model.....	15
----------------------------	----

## Appendix tables

Appendix Table A: Changes since the previous EEF evaluation .....	<b>Error! Bookmark not defined.</b>
---	-------------------------------------

Appendix Table B: Baseline Early Numeracy (EN) Checklist ..... 48

## Study rationale and background

Mathematical attainment is an important determinant of educational, economic, social and health outcomes in later life (Werner, Acs and Blagg, 2024). Attainment gaps in mathematics between disadvantaged pupils and their more affluent peers appear early and these differences then remain stable but stubborn over time (Outhwaite, Anders and Herwegen, 2022). Children who are behind their peers in maths attainment at the start of primary school will typically remain behind their peers throughout school. Early intervention is, therefore, central to reducing these attainment gaps. Numeracy is an area of priority to achieve the UK Labour government's pledge to break down barriers to opportunity (The Labour Party, 2024). The government's proposals include a focus on early numeracy, where the practical, everyday applications of maths are taught to children from age 4 onwards. Inspired by the success of the evidence-informed phonics programme in improving reading, there is also a proposal for a "phonics for maths" programme in the early years and primary classes (Gilmore, 2023). While phonics is a very specific technique for teaching word reading, mathematical concepts, even in primary school, are quite broad and complex. Therefore, a single programme or technique might not be the solution for early maths teaching. However, phonics was informed by research and an understanding of how children learn, and this approach could be emulated in the teaching of early maths.

Early numeracy approaches that aim to develop number skills and improve young children's knowledge and understanding of early mathematical concepts typically increase children's maths learning by seven months (EEF, 2023). While these approaches benefit all pupils, their impact on children from disadvantaged backgrounds tends to be higher, particularly when they are targeted to these pupils (ibid). Early numeracy approaches, therefore, offer a potential solution to reduce the attainment gap in the very early stages of formal schooling.

Professional Development (PD) is likely to be important to support early numeracy approaches (EEF, 2023). Key areas of improvement for early years' practitioners include knowledge of mathematics, knowledge of children's development and development trajectories and understanding of the types of activities that support early mathematical learning (ibid). Research has shown that PD programmes that incorporate a set of 14 'mechanisms' (empirically-evidenced principles about how people learn and change their practice) are more likely to be effective (Sims *et al.*, 2021). PD programmes that incorporate more mechanisms tend to have a higher impact on pupil outcomes, whereas those that incorporate fewer or zero mechanisms have small to no effects.

White Rose Education (WRE) (previously known as White Rose Maths (WRM)) was established in 2017 within the Yorkshire-based Trinity Multi Academy Trust. WRE develop and offers a range of schemes of learning, resources and PD programmes. Since 2019, WRE schemes have been used in 140 countries worldwide and in over 90% of primary schools in the UK. The WRM Reception Jigsaw is a PD programme that aims to improve early maths teaching quality and outcomes in schools. The mechanisms incorporated within the Reception Jigsaw programme are detailed in the 'Intervention' section below. The programme is targeted to Reception practitioners (teachers and teaching assistants) and involves five two-hour training sessions and five coaching and mentoring visits spread throughout the academic year (see 'Intervention' section for details). All

training sessions and coaching and mentoring visits happen face-to-face with a designated WRE trainer visiting the school.

The efficacy of Reception Jigsaw has previously been evaluated using a randomised controlled trial (RCT) commissioned by the Education Endowment Foundation (EEF) (Tang *et al.*, 2024). This efficacy trial involved 138 schools, which were randomly assigned to either the control or intervention groups. Control schools continued with their usual practice, and intervention schools received the Reception Jigsaw training programme. The trial found that children in Reception Jigsaw schools made, on average, an additional one month's progress in maths compared to children in control schools. Although this finding had a high security rating, it was not statistically significant, i.e., the evaluator could not conclude that this represented a real effect. Children who were taught by teachers who had complied with the programme requirements (i.e., attended at least nine out of the 10 training sessions) made an additional one month's progress. This finding was statistically significant and highlighted that the programme had a positive impact when compliance to the programme was achieved. Children eligible for Free School Meals (FSM) in intervention schools made no additional progress in maths compared to FSM children in control schools. There was some evidence to suggest that the Reception Jigsaw programme led to Reception teachers being more confident in teaching early maths. A planned longitudinal analysis showed that the impact of the Reception Jigsaw programme on pupils' maths attainment may be sustained up to a year after delivery of the programme. Children who were in Reception Jigsaw schools in their Reception year made, on average, an additional one month's progress in maths at the end of Year 1 compared to children in control schools.

This effectiveness trial of the Reception Jigsaw programme is expected to involve 6,600 pupils from 330 primary schools in England. Schools will be randomly assigned to receive either the Reception Jigsaw programme ('intervention' school) or continue with their usual practice ('control' schools). Intervention schools will receive access to the Reception Jigsaw programme during the 2025 – 26 academic year. Reception practitioners in intervention schools will start receiving training from October 2025. The primary outcome is Reception pupils' early numeracy skills measured by the Early Years Toolbox Early Numeracy assessment. The secondary outcome is wider mathematical skills of Reception pupils measured by the Early Years Foundation Stage Profile (EYFSP), specifically the two Early Learning Goals (ELGs) for maths. To minimise the burden and cost of data collection, only a sample of up to 20 children across the Reception year group will be assessed for impact in each school. The integrated Implementation and Process Evaluation (IPE) will seek to explore how and in what circumstances Reception Jigsaw impacts teachers and pupils. We will explore these questions using a combination of qualitative and quantitative methods. The qualitative methods will consist of training observations, Reception walk-throughs, interviews with teachers and senior leaders and focus groups with trainers, while the quantitative methods will involve analysis of surveys of teachers. The cost evaluation will provide a robust estimate of the cost to schools per Reception pupil per year of delivering the Reception Jigsaw programme.

There are two important design differences between the efficacy and effectiveness trials. The efficacy trial was not powered to detect an impact on FSM pupils. However, one of the objectives behind increasing the size of this effectiveness trial was to ensure that the trial would be

adequately powered for FSM pupils. Our sample size calculations suggest that we are in a better position to detect an impact for FSM pupils in this effectiveness trial compared to the efficacy trial. Another key difference between the two trials is the primary outcome measure. NFER used the New PUMA (Hodder's Progress in Understanding Mathematics Assessment) as the primary outcome measure in the efficacy trial. Following changes to the Reception curriculum in 2021, it was recognised that alignment between New PUMA and the revised curriculum was not strong. Although extensive exploration of alternative primary outcome measures ensued, due to logistical and budget considerations, New PUMA remained the primary outcome measure of the efficacy trial. Feedback from WRE and some schools also suggested that some aspects of this measure were not in line with expectations of pupils of this age or with teaching approaches followed in Reception Jigsaw. Taking into account the issues with New PUMA and the feedback from WRE and schools, we have chosen the Early Years Toolbox Early Numeracy assessment as the primary outcome measure following a thorough examination of potential alternatives.

## **Intervention: White Rose Maths Reception Jigsaw**

White Rose Maths Reception Jigsaw is a professional development training package that aims to improve mathematics teaching quality and outcomes in Reception classes. The professional development programme is treated as a manualised intervention, while the expected changes in teaching are expected to depend on teacher professional judgement and contextual factors, so are not manualised.

### **Why (rationale/theory)**

Research shows the lasting positive effects of high-quality Reception teaching (Tymms, Merrell and Bailey, 2017), and early maths knowledge (Aubrey, Godfrey and Dahl, 2006), which may help to narrow the gap in achievement throughout life.

The programme is underpinned by strong principles of early years mathematics teaching, specifically: learning through play; opportunities to explore and investigate through classroom provision; starting from and building on children's interests; using real objects in meaningful contexts to introduce mathematics learning; and the role of the adult in supporting and enhancing learning through a balance of adult-led and child-initiated activities. It is designed to improve teaching practice within regular maths teaching time, as well as maths learning within continuous provision.

Reception Jigsaw comprises five two-hour PD sessions and five personalised half-day coaching and mentoring visits over an academic year, led by an early years maths specialist, with gap tasks to be completed between sessions. The programme draws on key mechanisms for effective professional development (Sims *et al.*, 2021):

- instilling insights by consistently revising prior learning across training sessions, gap tasks and coaching and mentoring visits
- motivating change through explicit goal setting and reinforcement in coaching and mentoring visits

- acquiring new skills or techniques through modelling, rehearsal and trying out maths pedagogies and activities
- embedding changes in practice through action planning, physical resources to cue the use of new strategies, and rehearsal and context-specific repetition in the coaching and mentoring visits.

## **Who (recipients)**

The target schools

- are primary state schools in England in one of the [selected Local Authority](#) areas participating in the trial
- have not received White Rose Maths Reception Jigsaw training in the last two years, or been in the intervention group for the efficacy trial
- are not participating in any other EEF early years trials.

The training is primarily aimed at Reception teachers. The Maths Lead at the school must also attend the core training, as they have responsibility for overseeing and developing early years maths in their setting. Year 1 (Y1) and Year 2 (Y2) teachers may also attend the core training, with the rationale being to support the transition from Reception to Key Stage 1 (KS1). Reception and Y1/Y2 teaching assistants are also encouraged to attend the training. Coaching and mentoring is primarily received by the Reception teachers (selection and distribution discussed below).

Pupils in Reception are the indirect recipients of the training, via changes in Reception maths provision and teaching practice.

## **What (materials)**

*Specific to the schools participating in the intervention group*

Journal including gap tasks – this is a tool for reflection by Reception Jigsaw participants. It is not collected in or reviewed by WRM, although practitioners might use it during the half-day visit by the trainer to talk about their progress. At the end of each training session, participants are set a ‘gap task’ related to the training given, to be completed before the next session (also see section on ‘what’ below).

Physical resources to support maths teaching, in the form of dot plates and number cards.

*Available to all schools*

Online resources – publicly available to all schools on the WRM website. This includes teacher guidance, schemes of learning and interactive whiteboards. Awareness of these resources may be higher in the intervention schools due to their ongoing participation in the training. However, as WRM schemes of learning are commonly used by primary schools, we expect that control schools may also be aware of, and use, some WRM resources. This is part of real-world circumstances in an effectiveness trial. Note, we will check the control schools’ awareness/use in the IPE.

## **What (procedures, activities or processes used)**

The training received by the schools comprises five modules, each focusing on a specific aspect of early years maths. Each module uses the same cycle of development, comprising a PD training session, followed by setting a gap task which supports teachers to use their learning to develop

early years maths in their school, followed by a coaching and mentoring visit for personalised and tailored support.

**Five PD training sessions (each 2 hours)**, delivered face-to-face at each school. Each school receives their own training (i.e. not grouped with other schools) – for attendees, see ‘who’ (section above). The training sessions will be delivered according to the school’s needs (e.g., during the school day if staff can be released or outside teaching time, in the form of twilight or during INSET days).

Sessions will be delivered using slides, and slide handouts will be provided to schools. Slides will be the same across schools, with discussions and focus tailored depending on the school. Trainers are able to add examples based on their own experience, but they are not able to add/take away any slides.

Sessions will be led from the front, with opportunities for discussions about issues currently faced by the school/approaches taken.

The training sessions include practical activities and games that schools can use in their own classrooms. These will be demonstrated and form part of an interactive training session.

**Five gap tasks** - to be completed between sessions by Reception teachers, to encourage implementation and reflection of the learning in their teaching. The gap tasks also encourage teachers to identify children who are at risk of falling behind and use suggestions from the training to provide further support to these children.

Gap tasks are based on the content of the five PD sessions. All of the schools are given the same gap task, but how they approach the tasks is fairly open.

The gap tasks are discussed with the early years maths specialist during the half-day coaching and mentoring visit (see below).

### **Five half day coaching and mentoring visits from the Reception Jigsaw trainer**

These sessions are attended by (one of) the Reception teachers. The Maths Lead also usually attends at least part of the coaching and mentoring session, given their role in supporting and developing early years maths. This is to reduce burden on schools and the need for classroom cover as coaching and mentoring visits are during the school day. In schools with more than one Reception class, the distribution of 1:1 coaching time between the Reception teachers will be agreed with the school, based on support needs.

These visits aim to support Reception teachers in developing effective practice in their own setting, and as such are highly tailored to each school. Schools can also provide feedback on the extent to which they have used the training in their classroom.

The half day coaching and mentoring visits are made after the gap tasks and are an opportunity to reflect on the gap task with the trainer. WRE recommend that the coaching and mentoring visits happen around four weeks after the gap task is assigned, but this will depend on the availability of school staff and trainers.

These visits are more tailored than the PD training sessions but are still based on the topic covered in those sessions. The school and early years maths specialist agree how to use the time – this could involve coaching, lesson observations, reviewing the organisation of the classroom in relation to maths. Different schools may focus the time differently, for example some schools may opt to skill up a Reception teacher to become more expert in maths teaching in the early years, while others may use the time to support newer or less confident Reception teachers of maths.

The Jigsaw will be delivered to each school from October 2025 to June 2026. This start timeline for the first professional development session is slightly later than usual for Reception Jigsaw outside of the trial, in order to allow for baseline data collection and randomisation. However, it was felt by WRE that this would not unduly affect pupil outcomes. Moreover, the timeline is similar to that in the efficacy trial (with randomisation in October and endpoint assessments in June). The same WRE trainer will deliver all the sessions and coaching and mentoring visits in a particular school, to enable continuity and relationship building.

The five modules are:

1. **Developing Early Number Sense:** Focuses on counting principles, subitising, composition of numbers, comparison and number relationships with gap task around identifying children who are not yet secure with an element of number sense and planning opportunities to work with these children to support their next steps in learning.

2. **Creating a Mathematical Classroom:** Themed around attitudes to learning and incorporating everyday maths through classroom routines and designing effective enabling environments. The module also covers the role of the adult in supporting learning, with the gap task considering classroom provision and which areas of the setting support maths and why. Then, choose one area to consider how it can be developed further to support mathematical learning.

3. **Exploring Pattern, Shape and Spatial Reasoning:** Emphasises the importance of spatial reasoning underpinning all areas of mathematics. The gap tasks focus on identifying two or three children who are not yet secure with an element of spatial reasoning and planning opportunities to provide additional support and practise.

4. **Mathematical Talk and Questioning:** Discusses why talk is important and focuses on developing sustained shared thinking. It also focuses on creating opportunities for talk using open-ended questions, examples and non-examples. The gap task is to identify children who have gaps in their understanding of number and plan a series of number talks to support their next steps.

5. **Reasoning and Problem Solving in the Early Years:** Focuses on classroom culture, developing reasoning, and problem-solving through games and stories. This session explores how problems can be adapted to enable children to access them or to build further challenge for more confident children. The gap task is about trialling some of the suggested starting points and strategies for developing reasoning and problem-solving in the classroom.

## **6. Who (implementers)**

The Reception Jigsaw is delivered by expert trainers from WRE. Teachers receiving professional development then lead maths teaching and provision for pupils throughout the Reception year.

All trainers are qualified teachers and specialist leaders in education. They are required to have a deep understanding of research into early maths and the developmental progressions that children follow in each area of maths.

Most trainers are part of the WRE training team. Trainers who have previously delivered Reception Jigsaw will have refresher PD support and access to team days during the programme, to reflect and share.

#### *Training of new trainers*

The core team at WRE, who have substantial expertise in early years maths teaching, runs the training of trainers. A team of 17 trainers (four trainers who have previously delivered Reception Jigsaw training and 13 newly recruited trainers) will be trained by the core team and deliver the programme in this trial.

There are four days of introductory training at the start, then for each module there is approximately 1.5 days training per module.

In train the trainer sessions, trainers will have the opportunity to explore each session in depth, discussing the key pedagogy and research which forms the basis of the training. They will have the opportunity to read and discuss the key background reading which supports each session. In addition to training on the session content, the trainers will also have training on presentation and delivery style.

As part of the training, the trainers then practice delivering each of the PD training sessions in front of the WRE core team, and WRE provide feedback based on the sessions to the individuals.

Trainers will receive some training prior to the start of delivery. The remaining training - about how to deliver each of the modules - will be staggered through the 2025/26 academic year to ensure trainers are trained to deliver a module shortly before they run that module with schools.

One difference in the training model since the efficacy trial has been the increase in the number of train-the-trainer team days from 5 to 10 days. The approach of taking one day per PD training session remains from the previous trial, but the extra days have been scheduled to upskill the new trainers around early years pedagogy, provision, for discussing what has been seen in schools and providing support for the coaching and mentoring visits. These days aim to ensure there is common understanding about the early years classroom across the team. The experienced trainers will facilitate these days through sharing their experiences and practice.

#### *Quality Assurance of trainers' delivery*

Training sessions, delivered by every trainer, are observed by the core team at WRE twice a year. Based on their observations, WRE collate their feedback on different aspects of the training. The feedback is then analysed by the WRE team to identify and suggest specific areas for improvement for every trainer.

As some new trainers are being recruited by WRE to deliver Reception Jigsaw as part of the trial, WRE plan for all newly recruited trainers to deliver a 'practice' session before they deliver to schools in the trial. WRE will observe this practice session and provide feedback and guidance to the trainer as needed.

WRE collects feedback from school staff after every session and analyses and uses it to inform both the course content/delivery across all trainers as well as the line management/development of the individual trainers.

The trainers have meetings with their line manager every three weeks for supervision and (two-way) feedback.

## **7. How (mode of delivery)**

The same trainer delivers all training face-to-face to the same school. Each trainer will deliver training to 6-13 schools. The school will be given the trainer's email address so they can ask queries between training sessions.

The participants use what they have learned in the training to change/inform their teaching and delivery of maths in the classroom. The trainers provide support to do this through the coaching and mentoring visits.

## **8. Where (setting of the intervention)**

Professional development training sessions, coaching and mentoring visits, and Reception maths teaching and continuous provision, are delivered in each participating primary school, in the [selected Local Authority](#) areas participating in the trial.

## **9. When and how much (dosage and duration)**

Practitioners: Five two-hour CPD sessions for all Reception practitioners and maths leads. Five gap tasks for Reception teachers (optional for TAs). Five half-day coaching and mentoring visits per school, distributed between Reception teachers based on support needs.

Pupils: Teachers are encouraged to adapt the training to their own classroom, so implementation will look different in different schools, and even different classrooms in the same school.

## **10. Tailoring (adaptation to the intervention)**

Training sessions are delivered using a standard set of materials that White Rose has developed, but discussion and examples can be tailored to individual schools if there are particular areas they need support with.

Coaching and mentoring visits are highly tailored to each school and teacher, aiming to support the teachers to implement the subject knowledge and pedagogy from the training. The coaching visits are tailored based on a review of relevant gap tasks, teacher confidence in implementing the new pedagogy, the school's ELG data, current opportunities for maths in the classroom

environment, and the teachers' view of what is working well and where they would like support. Trainers will also provide additional responsive support where schools/teachers raise additional questions.

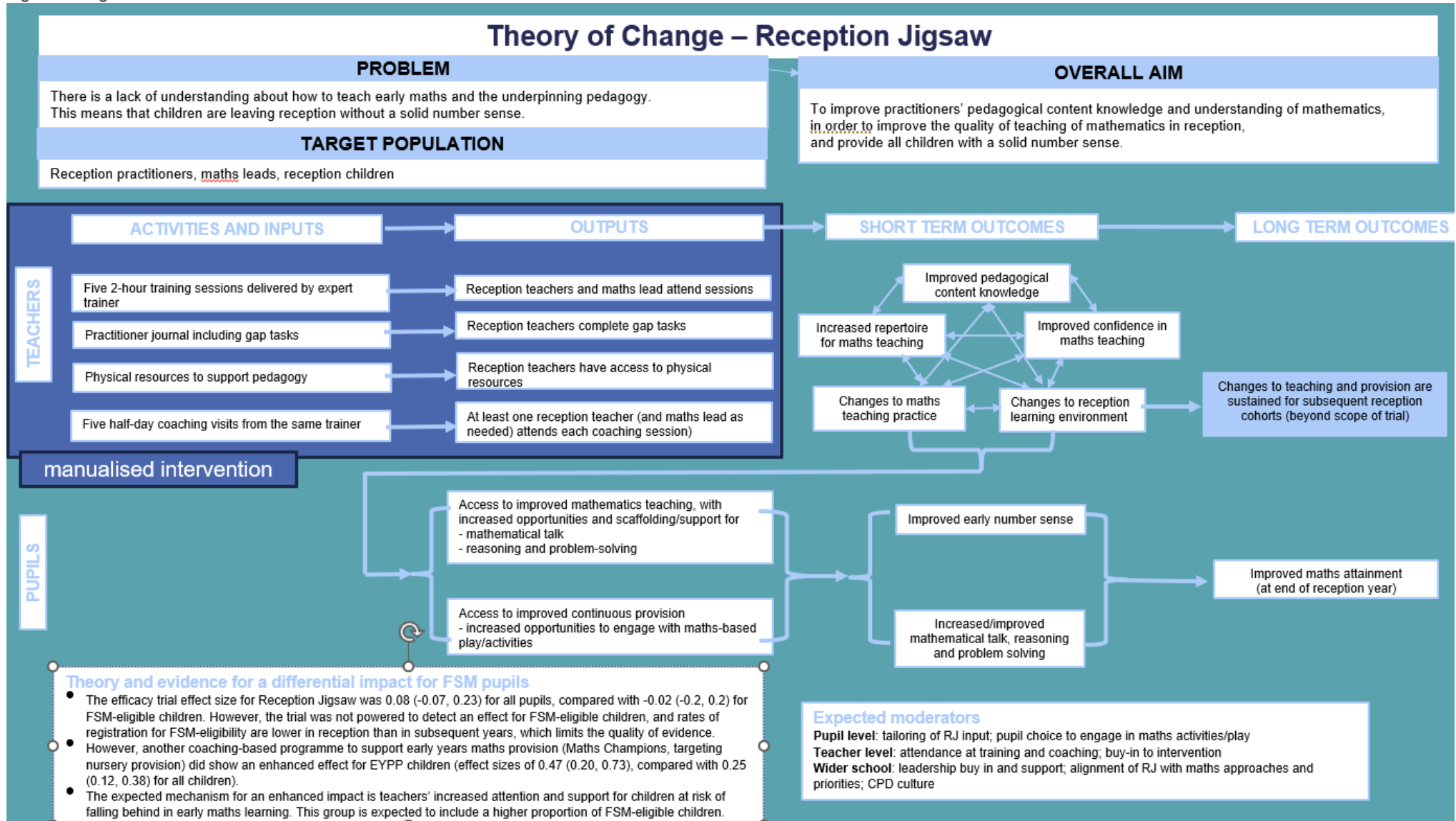
The training package provides adaptable ideas for new teaching strategies, activities and continuous provision that can be introduced, and teachers are encouraged to select from and tailor these to meet their pupils' needs. No strict lesson plans are provided.

### **11. How well (planned)**

WRE is able to (lightly) monitor schools' participation in the project during each visit, based on how teachers have used the training, and the extent to which they have completed the gap task and reflection. Trainers will feed this information back to the WRM project lead, who will have a follow-up phone call with the key contact at the school to discuss any issues that arise. The WRE project lead will keep a record of any issues and actions taken in response.

Below is a logic model, that outlines the hypothesised activities, outputs and outcomes. The potential mechanisms we are particularly interested in are the professional development mechanisms (see above) which are hypothesised to mediate changes in teacher practice, and the observed changes in teacher practice and provision, which are hypothesised to mediate changes in pupil actions and learning.

Figure 1: Logic model



# Impact evaluation design

## Research questions

### Primary research question

- RQ1. What is the impact of Reception Jigsaw on early numeracy (measured using the Early Years Toolbox) at the end of Reception?

### Secondary research questions

- RQ2. What is the impact of Reception Jigsaw on disadvantaged pupils' early numeracy (measured using Early Years Toolbox) at the end of Reception?
- RQ3. What is the impact of Reception Jigsaw on pupils' score on the Early Years Foundation Stage Profile (EYFSP) maths Early Learning Goals (ELGs)?
- RQ4: How does teachers' training/coaching attendance/compliance impact pupils' early numeracy (primary outcome)?

## Design

Table 1: Trial design

<b>Trial design, including number of arms</b>		Two-arm, cluster randomised
<b>Unit of randomisation</b>		School
<b>Stratification variables</b> (if applicable)		Geographic area to support delivery of the Reception Jigsaw programme
<b>Primary outcome</b>	<b>Variable</b>	Early numeracy (maths attainment)
	<b>Measure</b> (instrument, scale, source)	Early Years Toolbox – Numbers 2 App, 0-120, (Howard et al., 2022)
<b>Secondary outcome(s)</b>	<b>Variable(s)</b>	Early mathematical skills
	<b>Measure(s)</b> (instrument, scale, source)	EYFSP, taking the value one if a pupil meets or exceeds the two mathematics ELGs, and zero otherwise, National Pupil Database.
<b>Baseline for primary and secondary outcome</b>	<b>Variable</b>	Maths attainment
	<b>Measure</b> (instrument, scale, source)	Teacher assessment based on observation (Emerging Numeracy checklist), 10–60, bespoke instrument

This is an effectiveness trial to assess the impact of the WRM Reception Jigsaw programme on Reception pupils. The evaluation will use a two-arm cluster randomised controlled trial design. Participating schools will be randomly allocated in a 1:1 ratio into two arms. Randomisation will be stratified by geographic area to prevent some WRE trainers from having too many or too few schools delivering the programme relative to their capacity. Schools in the 'intervention' arm will

receive the WRM Reception Jigsaw CPD programme during the 2025-26 academic year. Schools in the 'control' arm will continue with their usual practice during the 2025-26 academic year.

Interested schools will complete an Expression of Interest (EoI), and schools identified to be eligible will sign a Memorandum of Understanding (MoU) to confirm their participation. Completion of the baseline assessment (i.e., the Emerging Numeracy observation checklist) by teachers for sampled pupils (sampling described later) will be a prerequisite for randomisation of the school. Although the Reception Jigsaw programme provides CPD training for Reception practitioners who will then teach their whole class, for the purpose of the evaluation, only a sample of Reception pupils from control and intervention classes will be assessed for impact (see 'Participant selection' section below for detail). This is primarily because sampling a subset of pupils allows us to reduce both the burden on teachers and the cost of assessing pupils.

The primary outcome for this trial is Reception pupils' early numeracy assessed by the Early Year Toolbox Early Numeracy app (the Numbers 2 app). At endpoint, the numeracy app will be administered by NFER test administrators who will be blinded to the school's random allocation. The secondary outcome for this trial is wider mathematical skills measured by the EYFSP maths ELGs.

All participating schools (intervention and control) that complete the baseline and endpoint testing requirements will receive a thank-you payment of £350. In addition, these schools will receive £250 of White Rose credit that can be spent on any White Rose subscription, resource or CPD and 12 months free access to any WRE webinar in any phase or topic after the trial.

## Participant selection

### School Eligibility Criteria

Schools can participate in the trial if:

1. They are located in one of the [selected Local Authority](#) areas participating in the trial.
2. They have not had the Reception Jigsaw training in the last two years or have not received the training as part of the previous Reception Jigsaw trial (2020-21).
3. They are not participating in any other EEF trials involving Reception teachers and pupils in 2025-26.

330 schools meeting the above criteria will be recruited to participate in the trial.

### Pupil Eligibility Criteria

Reception Jigsaw is a maths CPD programme that provides evidence-based knowledge to Reception practitioners on teaching early maths to *all* Reception pupils. Schools with mixed year classes will be allowed. While the programme can be delivered to all pupils within mixed classes only Reception pupils in those classes will be eligible to be assessed as part of the trial. Pupils at participating schools who are in Reception for the 2025/26 academic year will be eligible to participate in the trial. There are no specific eligibility restrictions amongst Reception pupils (e.g.

by maths attainment levels). However, as mentioned earlier, only a sample of Reception pupils from control and intervention classes will be assessed for impact.

Schools will complete an Excel template listing all pupils they expect to be in each Reception class for the 2025/26 academic year. Using this template, we will first sample up to 24 Reception pupils at random across all participating classes in the school for baseline Emerging Numeracy (EN) checklist completion (see ‘baseline measures’ below) in summer 2025. The EN checklist will be completed by the sampled pupil’s class teacher. Pupils without a completed EN checklist or who have been withdrawn from data processing will not be eligible to progress further in the pupil data collection for the trial (they would still continue to receive maths instruction from their teacher receiving the Reception Jigsaw training, if they are in a school that is randomised to the intervention group). At this point, it is expected that most classes will have completed the EN checklist for more than 20 pupils<sup>1</sup> (i.e. 21-24). However, only 20 of these Reception pupils will sit the Early Years Toolbox Early Numeracy test at the endpoint for cost-related reasons. These 20 pupils will be selected at random<sup>2</sup> from amongst the 24 pupils who were baselined and who are present on the day of endpoint testing. Any extra pupils above the 20 can effectively be used as substitutes for those with missing endpoint data (e.g. absent on the testing day), increasing the analysed sample size. The extra uncertainty as to which pupils will sit the assessment at the endpoint also reduces the possibility that analysed pupils are targeted with particular aspects of Reception Jigsaw more than other pupils in the class. In total, we expect 6,600 Reception pupils (20 pupils per school, 330 schools) to be assessed as part of this trial.

Unlike the efficacy trial of Reception Jigsaw, there will be no oversampling of FSM pupils as part of the pupil selection process. The efficacy trial oversampled FSM pupils with the aim of ensuring 25% of trial pupils were FSM-eligible. This was an adaptation made partway through the trial to achieve the Minimum Detectable Effect Size (MDES) in the FSM subgroup that was targeted in that trial’s initial design. We anticipate that the proportion of FSM pupils in the current trial will be lower than 25%, but by taking a random sample of reception pupils from across the whole year group in all the schools in the trial, the proportion of FSM pupils in the sample should reflect the proportion in the trial population. The MDES in the FSM subgroup will still be an improvement on the MDES of 0.265 seen in the efficacy trial, due to more schools being recruited (see ‘Sample size’ below).

## **Recruitment Strategy**

---

<sup>1</sup> There will be some schools with 20 or fewer pupils: at these schools all available pupils will sit the endpoint test as normal.

<sup>2</sup> The project statistician will randomly select 20 pupils for endpoint testing and the remaining 1-4 pupils will act as substitutes for any of the 20 that are absent on the day of testing. A priority order for the substitutes to be used will be determined, also at random. The random selection is therefore performed in advance, but exactly which pupils are tested is determined by who is present on the day.

WRE were responsible for the recruitment of schools to the trial. Schools were recruited from March to July 2025. We note that there was a six-week pause in active recruitment from the end of March to mid-May 2025 due to internal restructuring within WRE. This prompted a change in study design, which is explained in greater detail in the ‘Sample size’ section. Marketing emails were sent by WRE to targeted schools that met the eligibility criteria. Interested schools could either submit an EoI (and subsequently book a webinar) or directly book themselves to attend a webinar that provided more information about the trial. All schools that attended the webinar were sent an MoU. Schools that signed the MoU were then considered recruited to the trial. NFER supported WRE’s recruitment efforts by contacting a sample of English schools that met the trial’s eligibility criteria. These schools were directed to the trial’s EoI form and followed the same recruitment process as outlined above. WRE used the Monday.com<sup>3</sup> project management platform to monitor the progress of schools through the recruitment pipeline and contact them as required. We note that WRE met the recruitment target in July 2025; 338 schools agreed to participate by signing the MoU.

## Outcome measures

### Baseline measures

The baseline measure corresponding to the primary outcome and EYFSP secondary outcome is the Emerging Numeracy (EN) checklist (see Appendix B for the full checklist). The EN checklist was developed for use in the Reception Jigsaw efficacy trial (Tang *et al.*, 2024) as data from the Reception Baseline Assessment (RBA) is not available to researchers via the NPD (STA, 2025). This bespoke checklist was created to gauge numeracy skills in children in the early years without placing undue burden on teachers and pupils. It consists of 20 tasks based on the Early Years Outcomes and ELGs. The checklist is completed by teachers about their pupils - ideally after they have completed the RBA or after two weeks of the pupil attending school. It takes 5–10 minutes to complete per pupil. Teachers do not require any training to complete the baseline EN checklist as it is similar to the RBA assessment, however, NFER provides teachers with detailed instructions on completing the checklist. A Cronbach’s alpha of 0.95 suggests it is a reliable measure. It had a correlation of 0.59 with the outcome measure in the Reception Jigsaw efficacy trial, and so we consider it a more cost-effective way to improve statistical power, compared to using a commercial test at baseline (even though this might have slightly higher correlation). As mentioned previously, we only require the EN checklist to be completed for a sample of pupils across the Reception year group, to reduce both the teacher completion burden at baseline and the follow-up cost of test administrators.

### Primary outcome

NFER used the New PUMA as the primary outcome measure in the efficacy trial of Reception Jigsaw. However, New PUMA was not chosen as the primary outcome measure in this

---

<sup>3</sup> <https://www.monday.com/>

effectiveness trial for two important reasons. First, following changes to the Reception curriculum in 2021, it was recognised that alignment between New PUMA and the revised curriculum was not strong. Second, feedback from WRE and some schools suggested that some aspects of this measure were not in line with expectations of Reception-age pupils or with teaching approaches followed in Reception Jigsaw. Following an extensive exploration of other measures, we concluded that the Early Years Toolbox (EYT) is the most suitable measure for this effectiveness trial. The EYT is widely used in Australia; in England, the EYT was the subject of a pilot study co-funded by the Department for Education and was used as an outcome measure in the Study of Early Education and Development (SEED) study (Melhuish and Gardiner, 2020).

EYT measures key areas of early development, including numeracy, language, self-regulation, social-emotional development and executive function<sup>4</sup>. Each measure is a game-like assessment that is brief (5–10 minutes to complete) and will likely sustain the attention and interest of young pupils. Each measure is downloaded as an app for iPads. The Numbers task covers children’s emerging numeracy skills, including numerical language, spatial and measurement concepts, counting, matching digits and quantities, completing number lines, ordinality, subitising, patterning, numerical word problems and equations. The EYT Numbers task aligns well with the topics covered in the Reception Jigsaw twilight training sessions. The EYT Numeracy assessment has demonstrated construct validity, concurrent validity with established measures and high test-retest reliability (Howard *et al.*, 2022).

At endpoint, trained NFER Test Administrators will administer the Numbers 2 app individually to each sampled pupil using an iPad. The app is designed to be user-friendly and as such administrators do not require specific qualifications. The Numbers 2 app will be downloaded in advance of endpoint testing to all iPads that will be used. Test Administrators will be trained in the administration of the test, which will include practice runs with the Numbers 2 app on the iPad assigned to them to minimise the risk of app failure on the day of testing. The Number 2 app does not require wi-fi connection for data collection, thereby reducing the likelihood of disruptions to the test because of connectivity issues. However, stable internet connection is required to transfer the collected data to the database. NFER will monitor the status of data transfer to ensure that the data has been transferred to the correct location. If the automatic data transfer process fails for any reason, we will follow the manual data extraction process that is built into the app.

### **Secondary outcome**

We plan to use an aggregation of the two mathematics ELGs from the Early Years Foundation Stage Profile (EYFSP) as a secondary outcome measure. If pupils meet both the number and numerical patterns ELGs, this secondary outcome will be a 1 (ELGs met); otherwise, it will be 0 (ELGs not met). The NPD variables ‘FSP\_MAT\_E11’ and ‘FSP\_MAT\_E12’ will be used to determine whether each ELG has been met. This will provide another, though wider, i.e. less proximal, measure of mathematics attainment (expected/emerging level of development).

---

<sup>4</sup> <https://www.eytoolbox.com.au/download.html>

## Sample size

All sample size calculations were performed using the PowerUpR package (Bulus *et al.*, 2021) in the R statistical software (The R Foundation, 2025), using the function 'mdes.cra2'. This function computes the MDES for a cluster-randomised trial.

We required a trial design that could detect a relatively small intervention effect, given the effect size of 0.08 seen for the primary outcome in the RJ efficacy trial (Tang *et al.*, 2024). It seems reasonable to anticipate a slightly higher effect size than 0.08 for the current trial, given that the primary outcome measure may be more appropriate for Reception pupils (see 'Primary outcome'). Additionally, the impact of RJ in the efficacy trial may have been affected by Covid-related disruption, although the trial itself did not find evidence of this, based on data collected in the IPE. However, we also note that effect sizes are typically smaller in effectiveness trials than at the efficacy stage.

### Original trial design

Originally the evaluation design was for 14 Reception pupils per school to be selected for the trial, with 10 completing the EYT endpoint assessment. The choice of including no more than 10 pupils per school was made for cost reasons, as a maximum of 10 EYT assessments can be completed per test administrator per day. Including more schools in the trial was considered a more cost-efficient way to ensure the trial was adequately powered than including more than 10 pupils per school. At this stage, the recruitment target was 400 schools, which is higher than most EEF-funded trials, but was considered achievable from a recruitment perspective. This number of schools allowed for a MDES of 0.137, after accounting for attrition and specifying further parameters as given for 'Current trial design' below<sup>5</sup>.

During the recruitment phase, it became increasingly clear that recruiting 400 schools before the autumn 2025/26 term would not be possible. This was in part due to a pause in active recruitment for 6 weeks, due to internal restructuring within WRE and confirmation of the continuation of RJ in its current form. By July 2025, it was deemed possible to recruit approximately 330 schools, rather than the 400 intended. This prompted discussions around whether the trial could be redesigned to achieve the originally intended MDES of 0.137 with only 330 schools. The only way to achieve an MDES of 0.137 was by increasing the number of pupils assessed using the EYT, requiring two endpoint testing visits per school. This was agreed with EEF, and pursuing this strategy led to the current trial design, which is described below.

### Current trial design

---

<sup>5</sup> The exception is the anticipated school-level attrition rate, which was reduced from 10% to 5% between the original trial design and the current trial design. This was due to the evaluators reconsidering the probable attrition rate, rather than new information becoming available about the attrition rate.

For the current trial design, 24 pupils will have a baseline EN checklist completed, of which 20 will be assessed using the EYT at endpoint. The number of schools recruited has been reduced to 330, to reflect the number found to be achievable in practice. We estimated ICC,  $R_1^2$  and  $R_2^2$  values based on the endpoint data from the RJ efficacy trial (Tang *et al.*, 2024). Although we are proposing a different outcome measure for this effectiveness evaluation, the baseline measure, intervention and pupil year group will be the same as in the efficacy trial, so this was considered the best available source for these parameters. This results in an MDES of 0.132, as shown in Table 2.

In the FSM subgroup, the MDES is 0.170, assuming an average of 3.5 FSM pupils per Reception year group. This average is based on 17.5% of the 20 reception pupils sampled and tested at the endpoint being eligible for FSM (see [DfE statistics](#)). As an effect of 0.137 (see Table 3) is approximately what is expected in the general population of Reception pupils, and there is no reason a priori to suppose a greater impact in the FSM subgroup, it is likely that the power for the FSM subgroup analysis is less than 0.8.

Table 2: Sample size calculations before accounting for anticipated attrition

		Overall	FSM
<b>Minimum Detectable Effect Size (MDES)</b>		0.132	0.170
<b>Proportion of variance explained</b>	level 1 - pupil ( $R_1^2$ )	0.396	0.396
	level 2 – school ( $R_2^2$ )	0.147	0.147
<b>Intracluster correlations (ICCs)</b>	level 2 (school)	0.183	0.189
<b>Alpha</b>		0.05	0.05
<b>Power</b>		0.8	0.8
<b>One-sided or two-sided?</b>		Two-sided	Two-sided
<b>Average cluster size</b>		20	3.5
<b>Number of schools</b>	Intervention	165	165
	Control	165	165
	<b>Total</b>	330	330
<b>Number of pupils</b>	Intervention	3300	577.5
	Control	3300	577.5
	<b>Total</b>	6600	1155

The figures in Table 2 above do not account for attrition between the recruitment and analysis stages: the MDES is based on the number of schools and pupils we aim to recruit. To see the anticipated impact of attrition, school and pupil numbers in Table 3 below have been reduced to allow for 5% school-level attrition and 15% pupil level attrition. It is assumed that 95% of schools remain in the trial, and within these schools 85% of pupils will be included in the primary analysis. This shows that due to attrition, an MDES of approximately 0.137 is anticipated at the analysis

stage (0.181 in the FSM subgroup). The pupil-level attrition of 15% is approximately the level observed in other NFER trials where the primary outcome relies on an assessment delivered by NFER test administrators. This estimate may seem too high, given that the strategy of keeping up to 24 pupils in the trial until endpoint testing is designed to reduce pupil attrition (see ‘Pupil eligibility’ above). However, in practice there are likely to be schools with fewer than 24 (or even fewer than 20) Reception pupils, or schools that complete 20 or fewer EN checklists for their 24 selected pupils, so we have opted for a conservative attrition estimate to allow for this.

Table 3: Sample size calculations with school and pupil numbers reduced to account for anticipated attrition

		Overall	FSM
<b>Minimum Detectable Effect Size (MDES)</b>		0.137	0.181
<b>Anticipated attrition rate</b>	level 1 (pupil)	15%	15%
	level 2 (school)	5%	5%
<b>Average cluster size</b>		17	2.98
<b>Number of schools</b>	Intervention	157	157
	Control	157	157
	<b>Total</b>	314	314
<b>Number of pupils</b>	Intervention	2669	468
	Control	2669	468
	<b>Total</b>	5338	936

The remaining sample size parameters are the same as in the previous table.

## Randomisation

Schools that complete the EN checklists for sampled pupils will be eligible for randomisation. The randomisation will occur at the school level: schools will be randomised into the intervention and control arms on a 1:1 basis. Randomisation will be carried out by an NFER statistician using R Code, which will be QA'd by another statistician and stored for reproducibility and transparency. Neither statistician will be blinded to group allocation. The allocation data will then be passed over to NFER's Operations team, who will liaise with schools and WRE. Randomisation will be stratified by government office region to reduce the likelihood of some trainers needing to deliver the Reception Jigsaw training to more schools than their capacity allows.

## Statistical analysis

All analyses will be conducted in line with EEF's [Statistical Analysis Guidance](#). Apart from the missing data analysis, all analyses will be restricted to 'complete cases': pupils with any missing analysis variables will not be included. Analyses will be conducted on an 'intention-to-treat' basis (except the compliance analysis), with pupils analysed according to their school's randomised allocation, regardless of whether their teacher completed the Reception Jigsaw training. The analyst will not be blinded to treatment allocation for any of the analyses described.

## Primary analysis

The primary outcome that will be used to answer RQ1 is the total raw EYT score, measured via a test administered at the endpoint by NFER test administrators. This outcome will be the dependent variable in a two-level (pupil, school) linear regression model and will be regressed on a randomisation group indicator variable representing the effect of the Reception Jigsaw intervention. Baseline EN score will be included as a covariate, and there will be dummy covariates representing each geographic region that was used to stratify the randomisation.

### **Secondary analysis**

A secondary analysis will explore the impact of the Reception Jigsaw training on whether pupils meet both mathematical ELGs in the EYFSP (RQ3). The sample of pupils will be the same as the primary analysis, aside from differences due to missing outcome data. To minimise the burden of baseline EN checklist completion for schools, a maximum of 24 pupils will be analysed per school, despite ELG data being available on the NPD for all Reception pupils. Having baseline data for the sampled pupils improves the power for this outcome, compared to including all Reception pupils (but without baselines included in modelling). Our privacy notice and information for parents explain that outcome data is being processed for the sampled pupils only. A two-level (pupil, school) logistic regression will be run in which the dependent variable is an indicator for whether pupils met the expected level in both mathematics ELGs. Baseline EN score will be included as a covariate, as will dummy covariates representing geographical regions.

### **Estimation of effect sizes**

The effect estimates for each binary predictor - that is, the beta coefficient conditional on model covariates - will be converted into an effect size, dividing by the square root of the total variance (a similar effect size to Hedges'  $g$ ). The total variance will be calculated by summing the pupil- and school-level variance components from a two-level (pupil, school) model with no covariates. A 95% confidence interval for the effect size will be calculated by dividing the endpoints of a 95% confidence interval for the estimate by the square root of the total variance. Estimates for continuous predictors will not be converted into an effect size.

### **Sub-group analyses**

The effect of the programme amongst disadvantaged pupils (RQ2) will be explored by repeating the primary analysis model, restricted to the subset of pupils who are eligible for FSM. The NPD variable 'EVERFSM\_6\_P' from the school census will be used to identify whether pupils are eligible for FSM. Additionally, we will investigate the differential effect of the programme for FSM pupils relative to non-FSM pupils by repeating the primary analysis model with an indicator for FSM eligibility added, as well as an interaction term between the FSM and the randomisation indicators. This interaction term represents the differential effect of the Reception Jigsaw training for FSM pupils compared to non-FSM pupils.

### **Analysis in the presence of non-compliance**

Compliance will be defined by the teacher attendance of the training sessions. The five PD training sessions and five coaching and mentoring visits are all assumed to be of equal value, so

each teacher's training session attendance will be a number between 0 and 10. Attendance registers will be used to record each teacher's attendance at the training sessions. For modelling purposes, each pupil's compliance will be determined by the number of PD training sessions attended by their class teacher AND the number of coaching visits attended by their class teacher and/or a key person at their school. The key person is identified by the school as someone (e.g., Maths Lead, senior leader, class teacher) who will consistently attend all training sessions to provide continuity in the event that the Reception class teacher(s) is unable to attend all training sessions and coaching visits. For example, the first PD training session will count as 'attended' for a pupil only their class teacher attends. The first coaching visit will count as 'attended' if their class teacher attends, or the key person attends, or both. There are two reasons for using the pupil's teacher and/or key person's attendance of the coaching sessions, rather than solely considering the attendance of the pupil's teacher. Firstly, WRE do not consider it an essential part of the intervention that every class teacher attends every RJ coaching visit: the role of the key person is important in ensuring the learning from the training is incorporated in teaching practice when the class teacher is unable to attend. Secondly, if the attendance of the coaching sessions was based solely on class teacher attendance or solely on the key person's attendance it would be possible for a pupil to benefit from coaching sessions via the other staff member's attendance. This would be a violation of the exclusion restriction, which is discussed in more detail below.

There will therefore be one compliance measure for this evaluation: the impact of each additional Reception Jigsaw training session attended by a pupil's teacher and/or the school Maths Lead. This is a continuous compliance measure and will be estimated using instrumental variable modelling in which the units are pupils.

We will not take the approach of dichotomising the number of sessions at a certain threshold to create a binary compliance measure. This is due to what is known as the 'exclusion restriction': randomised group assignment can only influence EYT score via the compliance measure. One consequence of the exclusion restriction is that when dichotomising the number of training sessions delivered to create a binary measure, we must assume there is no benefit from Reception Jigsaw below the chosen threshold. For example, if we dichotomised at eight sessions, we must assume children whose teacher and/or Maths Lead attended seven or fewer sessions get no benefit from Reception Jigsaw. Our analysis instead assumes that the relationship between the number of training sessions delivered and EYT score is approximately linear: we believe this to be a more realistic assumption.

### **Missing data analysis**

Pupils without a completed EN checklist will not be included in the trial at the point of randomisation. This means no trial pupils will have a missing baseline, so the only variable with potential missing values in the primary analysis will be the EYT outcome. Exploration of the pattern of missing data and implications for the reliability of the primary analysis result will be conducted (e.g. via multiple imputation) in accordance with the EEF analysis guidance.

Where schools drop out and do not provide endpoint EYT scores, we will try to establish the reason for this, where possible, and provide a summary of reasons in the report. Test administrators will record an absence code for pupils missing at the endpoint EYT assessments,

providing a high-level reason for the absence (e.g. no longer at school), which will be described in the report.

## Implementation and Process Evaluation (IPE) design

Our IPE focuses on areas of the ToC and causal assumptions that are not yet fully understood from the efficacy trial, or where there is reason to expect a difference at a larger scale.

### Research questions

#### **RQ5a. To what extent does Reception Jigsaw maintain fidelity of delivery in relation to its PD training and coaching at scale?** (*fidelity, dosage*)

We will explore how fidelity is maintained at scale, including i) recruitment of trainers at scale and trainers' compliance with attending their training, and ii) the extent of training and coaching attendance by reception staff and maths leads. This will support us in explaining any 'voltage drop' in impact in moving to larger-scale, nationwide delivery.

- How many trainers were recruited?
- What is the extent of trainers' attendance at training?
- What is the extent of staff attendance at PD training and coaching?
- To what extent are the gap tasks completed?

#### **RQ5b. To what extent does Reception Jigsaw maintain quality of delivery in relation to its training of trainers, and of staff training and coaching, at scale?** (*quality, context*)

- What was the nature of the train-the-trainer model at scale?
- How do trainers perceive the quality of the training they received? How do White Rose perceive the quality of the trainers recruited?
- How do school staff and trainers perceive the quality of Reception Jigsaw PD training and coaching and mentoring, when delivered at scale?
- How do trainers tailor Reception Jigsaw training/coaching to school/staff needs?
- What enablers and barriers do staff experience when engaging with Reception Jigsaw PD training, coaching and mentoring and gap tasks?

#### **RQ6. What are the observed changes in teaching practice and provision in Reception Jigsaw schools? How and when do they occur? What seem to be the key causal mechanisms for those changes?** (*mediators, perceived change*)

As Reception Jigsaw is a professional development programme working with teachers, the causal chain to pupil impact is long: from CPD, through practice change, to outcomes for children. Building on the efficacy trial logic model, we will explore the 'black box' of changes in Reception maths provision, which mediate pupil impact, which have not yet been fully characterised. We will particularly explore how coaching attendance (explored as a moderator in the efficacy trial) mediates impact and whether it is part of other mechanisms such as commitment/engagement.

- How does Reception Jigsaw support staff to develop teaching practice and provision?
- What changes in teaching practice and provision occur in Reception Jigsaw schools? When do they occur? And how?

- To what extent does coaching attendance act as (a) a proxy for broader engagement with Reception Jigsaw (b) a causal mediator for Reception Jigsaw? How?

**RQ7. How is Reception Jigsaw moderated by school priorities, senior leadership support, and other contextual factors?** (*context, moderators*)

A key moderator for the impact of professional development is how the programme aligns within the school ecosystem, including the CPD culture, staff support, and the capacity to embed changes within teaching and learning. For example, the efficacy trial posited school/senior leadership support as a moderator but was not able to fully explore this.

- How does the Reception Jigsaw programme align with school priorities for developing Reception maths? Do the PD training, coaching and mentoring and gap tasks align with school needs? Is tailoring perceived to meet those needs?
- Does the level of senior leader engagement and support moderate changes in Reception maths and perceived impact?
- Is senior leadership engagement perceived to mediate changes in teachers' practice?
- What contextual enablers and barriers do staff experience in engaging with and delivering Reception Jigsaw?

**RQ8. What is the extent and quality of tailoring of the Reception Jigsaw approach/delivery to children's needs?** (*adaptation, tailoring, moderators*)

The efficacy trial found that Reception Jigsaw's flexibility, through adaptation and tailoring to pupils' needs, supported schools to embed the strategies in teaching and provision, and supported perceived impact for pupils. This focus on adaptation and tailoring has been enhanced in the professional development programme being used for the effectiveness trial and will be a line of enquiry for this evaluation. As effectiveness trials typically encompass a wider range of school contexts, we will also explore any differences in how Reception Jigsaw is used in different contexts (within and between schools), and any challenges encountered.

As for all EEF evaluations, we will particularly explore the experience of FSM-eligible pupils.

- How and why do teachers adapt Reception Jigsaw strategies with their classes?
- To what extent, and why, do teachers use Reception Jigsaw in targeted group teaching?
- What is the experience of FSM-eligible pupils?

**RQ9. What is the perceived impact of Reception Jigsaw?** (*perceived impact*)

- What are the perceived impacts of Reception Jigsaw for staff (e.g., improved pedagogical knowledge, increased confidence in maths teaching, changes to maths teaching practice)?
- What are the perceived impacts of Reception Jigsaw for pupils (e.g., improved early number sense, improved mathematical talk, reasoning and problem solving)?
- What are the perceived impacts of Reception Jigsaw for FSM-eligible pupils?

**RQ10. How does Reception Jigsaw relate to usual practice in Reception maths? (usual practice, differentiation)**

We are aware of the changing context of Reception maths provision, with an increasing number of professional development and curriculum programmes available in recent years. Therefore, it is particularly important to understand the current ‘baseline’ of usual practice in schools. We will explore:

- What is usual practice in Reception maths (including the use of early number programmes, and WRE teaching schemes and resources)? (intervention and control schools)
- How do staff integrate Reception Jigsaw with usual practice?
- How are Reception Jigsaw practices similar to and different from usual practice, and how do they add value to Reception maths?

**RQ11. What is the cost of implementing Reception Jigsaw in schools? (cost)**

- What are the financial and time costs for implementing Reception Jigsaw?

Please see the cost evaluation design for further information.

## Research methods

IPE instrument design will draw on those from the efficacy trial (in particular surveys and logs) adapted to reflect the RQs/themes of this effectiveness trial. Qualitative instruments will focus on the new lines of enquiry relating to mechanisms and context.

We used the EEF review of CPD mechanisms (Sims *et al.*, 2021) as a primary tool to co-map CPD mechanisms (instilling insight, motivating change, developing change and embedding in practice).

### Usual practice survey

A survey on usual practice will be sent to all schools at baseline, and to control schools at endpoint, for completion by the key project contact, typically the KS1 maths lead or Early Years Lead. It will explore use of mastery practices, early maths programmes, targeted group teaching, continuous provision and CPD. The endpoint surveys of headteachers, teachers and Maths Leads in intervention schools will also cover usual practice (see later section).

### Professional development (interviews, observations, attendance register, focus group)

We will focus primarily on trainer CPD sessions and coaching, with a light touch developer interview at baseline to explore WRE’s reflections on trainer recruitment, train-the-trainer sessions, and intentions for supporting training/coaching attendance at scale.

We will conduct seven observations of CPD delivery of the two hour training sessions, organised so that we observe all topics at least once.

An attendance register for training/coaching (designed by NFER in consultation with WRE and data provided by WRE) will focus on Reception staff and maths leads as the core participants. This will support the compliance measure (see Impact Evaluation) as well as the IPE.

A focus group with trainers will gather their perspectives on the CPD, including delivering CPD at scale, reflections on the minor changes to the Reception Jigsaw CPD since the efficacy trial, and changes in Reception maths teaching and continuous provision. We will hold this focus group up to three times to facilitate all trainers, each attending once (the number of trainers is expected to be more than can be accommodated in one session).

### **Case-study visits**

Case-study visits will be valuable for exploring changes in Reception maths practice, which is the key pathway to impact expected in the theory of change. We will visit eight Reception Jigsaw schools. We will draw a purposive sample of intervention classes, aiming to capture a maximum range across key moderators and contextual factors in the theory of change, including the proportion of FSM-eligible pupils, prior reported use of mastery approaches, attendance (to date) at Reception Jigsaw training/coaching, use of mixed Reception/Y1 classes, timetabling of maths teaching, and whether schools are using Mastering Number. Where a school declines to participate, we will contact a similar school. From prior experience, we expect the achieved sample of qualitative case visits to skew towards more engaged schools and teachers. We will use survey responses and attendance data to characterise the classes and teachers involved, in comparison with all participating intervention schools.

Visits will comprise:

- an interview with the trial lead (usually an early years or maths lead), focusing on the alignment of Reception Jigsaw with school development priorities and usual practice, and collecting cost data
- a ‘walkthrough’ interview with a Reception teacher and TA, covering both practice and continuous provision), covering all IPE dimensions
- photographs of classroom equipment/provision (not including children) to illustrate classroom environments
- an observation of Reception children working with Reception Jigsaw pedagogy or resources. This will focus on understanding their maths learning experience in comparison with key pupil outputs in the theory of change.

In addition to intervention case study visits, we will conduct up to four comparison visits in control schools to enable comparison of differences in practice, and explore the range of ‘usual’ practice including the use of other programmes. Should control schools not be forthcoming to host visits, we will conduct these as online interviews, omitting the observations of practice. The sampling process will be the same as for the intervention schools (described above) minus the variable on Reception Jigsaw training attendance, and where feasible, we will select control schools with a similar range/mix of characteristics (from baseline survey and demographic data) to intervention case study schools. Visits will comprise:

- an observation of Reception children learning maths
- a ‘walkthrough’ interview with a Reception teacher and TA
- photographs of classroom equipment/provision (not including children) to illustrate classroom environments

This data will be valuable in interrogating the disparity between the developer’s contextual assumption that the early number approaches promoted by Reception Jigsaw are highly differentiated from usual practice, and Reception practitioners’ perceptions during the efficacy trial that Reception Jigsaw approaches were only slightly different from their usual practice. As well as contributing to this trial, this usual practice data will provide context for developing Reception maths programmes and practice, which is particularly important given the lack of other promising programmes from recent trials.

### **Endpoint survey (intervention schools)**

The endpoint survey for teachers, TAs and maths leads in intervention schools will focus on views on coaching and mentoring, implementation within the school context (e.g. leadership support), adaptation/tailoring, and perceived impacts including practitioners’ perceived changes to pedagogy and practice from Reception Jigsaw. The survey will primarily comprise closed questions (e.g. Likert scales) which will be sent via a direct email link to reception class teachers and maths leads. The main contact will be asked to send a multi-user link on to TAs who work in reception.

The endpoint survey for the headteacher in each intervention school will focus on school support/buy-in, the alignment of Reception Jigsaw with school priorities and usual practice, integration into school planning, and views on cost-effectiveness/cost sustainability (see Cost evaluation). From previous experience, we expect a lower response rate for headteacher surveys as they are less directly involved in the trial. Our staff survey which will include the maths lead (described above) will also provide a leadership perspective on Reception Jigsaw within the school context.

### **IPE analysis**

A common deductive analysis framework will be developed from the research questions and used to interpret all data.

Qualitative data (interview transcripts and observations) will be analysed thematically using the data analysis software MAXQDA, using inductive summaries and coding within the overarching deductive framework to characterise experiences and perspectives. We will develop a descriptive case study for each intervention case study teacher and class (n=8). These will draw on interviews, observations and key attendance and survey data (where available), to help understand their CPD engagement, changes in teacher practice, and perceived impact, in comparison with the theory of change. These case study descriptions will be for analyses purposes, including to triangulate data sources; they will not form part of the published output. We will report any cross-case patterns from these case studies.

Quantitative datasets (survey and attendance registers) will be analysed using descriptives, cross-tabulations and simple inferential statistics. This will include a comparison of baseline and endpoint data on usual practice, and cross-tabulation of survey responses by teacher-reported engagement with the CPD provision.

We will triangulate data sources, providing a comprehensive assessment of programme implementation, and map the evidence against the theory of change.

Table 4: IPE methods overview

IPE Dimensions	Research questions	WRE interview	CPD observations	Trainer focus group	Training / coaching attendance register	Maths/EY lead interview	Reception teacher/TA walkthrough	Reception children observation	Cost proforma	BAU <i>BL - all, EP - control</i>	Staff survey <i>EP - treatment</i>	Head survey <i>EP - treatment</i>
Fidelity, quality, CPD-dosage	RQ5a. To what extent does RJ maintain fidelity of delivery in relation to its PD training and coaching, at scale? RQ5b. To what extent does RJ maintain quality of delivery in relation to its training and coaching, at scale?	✓	✓	✓	✓	✓	✓					
Mediators, perceived change	RQ6. What are the observed changes in teaching practice/provision in RJ schools? How and when do they occur? What are the key causal mechanisms of change for RJ (incl. coaching)?	✓	✓	✓	✓	✓	✓	✓			✓	✓
Context, moderators	RQ7. How is RJ moderated by school priorities and senior leadership support? What other factors moderate RJ?	✓		✓		✓	✓				✓	✓
Adaptation, tailoring, moderators	RQ8a. What is the extent and quality of tailoring of RJ to children's needs?			✓		✓	✓	✓			✓	
	RQ8b. How does RJ support disadvantaged pupils?			✓		✓	✓	✓			✓	
Perceived impact	RQ9. What is the perceived impact of RJ for teachers, pupils and disadvantaged pupils?			✓		✓	✓	✓			✓	✓
Usual practice, differentiation	RQ10. How does Reception Jigsaw relate to usual practice in Reception maths?			✓		✓	✓			✓	✓	✓
Cost	RQ11. What is the cost of implementing Reception Jigsaw in schools?					✓	✓		✓		✓	✓

## Cost evaluation design

A full cost evaluation was conducted as part of the recent efficacy trial (academic year 2021/22). Financial costs and time costs were reported separately, and we intend to take the same approach in this trial. Since the previous evaluation, anticipated time and financial costs have remained the same, including the programme model, resources and staff time, except for adjustments for inflation. Therefore, we would like to start from the assumption that financial and time costs remain similar.

The IPE will test this assumption, by briefly collecting information about

- Direct financial costs incurred for school staff to attend training/coaching, or purchase resources and equipment, to be collected from the sample of case study schools. (As the majority of direct financial costs reported in the efficacy trial related to the invariant programme costs, we do not believe we need to collect this information from all schools.)
- Staff time costs for engaging with CPD training/coaching, to be collected via endpoint surveys.

If reported costs are similar, we will primarily use the costs calculated by the previous trial and adjust the financial costs for inflation (as described in EEF Cost Evaluation Guidance, 2023).

If reported financial and/or time costs are different to the previous trial, we will conduct uncertainty analysis to consider different scenarios, to calculate cost per pupil per year in line with EEF guidance.

In addition to quantitative cost estimates, we will gather staff perceptions of costs through the endpoint surveys. This will focus on sustainability, including to what extent schools are able to resource and staff the programme and its implementation, their willingness to pay for it without subsidy, and views on its cost-effectiveness.

## Ethics and registration

This evaluation will be conducted in accordance with NFER's [Code of Practice](#). All of NFER's projects abide by its Code of Practice, which is in line with the Codes of Practice from BERA (the British Educational Research Association), MRA (the Market Research Association) and SRA (the Social Research Association), among others. NFER is committed to the highest ethical standards in all of its activities and ethical considerations are embedded in its detailed quality assurance processes.

### Ethical Agreement

Each participating school's headteacher will provide their agreement to participate in the trial by signing the Memorandum of Understanding (MoU) that outlines the responsibilities of all parties involved in the trial. MoUs will be signed by schools on a rolling basis as they are recruited by WRE from March – July 2025. NFER will share a parent letter and withdrawal form via WRE with schools to be sent to parents/carers of all Reception pupils in the school. Through the withdrawal form, parents/carers will have the opportunity to withdraw their child from the evaluation and associated data processing at any stage of the trial. Unless explicitly instructed, NFER will use all data collected up to the point of withdrawal in analysis. Withdrawal of pupils' data will not be possible after analysis has commenced (planned for August 2026).

Teachers completing surveys as part of this trial will be provided with full information about the research and how their responses and data will to be used. Separate privacy notices<sup>6</sup> have been drawn up for pupils and school staff giving details of any personal data being collected as part of the trial and how they were being transferred, stored, used, and reported.

### Registration

This trial will be registered at the [Open Science Foundation \(OSF\)](#) registry and the link to the project page on the OSF registry will be added to this protocol as soon as it becomes available. The trial registry will also be updated with outcomes at the end of the project.

---

<sup>6</sup> <https://www.nfer.ac.uk/for-schools/participate-in-research/participate-in-research-projects/the-white-rose-education-reception-jigsaw-trial-2025-26/>

# Data protection

## Data Protection Statement

All data gathered during the trial will be held in accordance with the data protection framework created by the Data Protection Act 2018 and the General Data Protection Regulation 2016/679 and will be treated in the strictest confidence by NFER, WRE and EEF. No individual or school will be identified in any report, however as we will use quotations or brief descriptions of practice, an individual may be able to identify themselves from the text.

WRE is the data controller for the WRM Reception Jigsaw programme. NFER and WRE have signed a Data Sharing Agreement that will govern the collection and sharing of personal data during this trial. This agreement includes a description of the nature of the data being collected and how it will be shared, stored, protected and reported by each party. In addition, WRE will provide an MoU to schools, explaining the nature of the data being requested of schools, teachers and pupils, how it will be collected, and how it will be passed to and shared with NFER. Three separate Privacy Notices, one for schools, another one for parents and a third for WRE trainers, are available [here](#).

## Legal Bases

The legal basis for processing **personal data** is covered by: GDPR Article 6 (1) (f) which states that ‘processing is necessary for the purposes of the **legitimate interests** pursued by the controller or by a third party except where such interests are overridden by the interests or fundamental rights and freedoms of the data subject which require protection of the personal data’. We have carried out a legitimate interest assessment, which demonstrates that the evaluation fulfils one of NFER’s core business purposes (undertaking research, evaluation and information activities). It also has broader societal benefits and will contribute to improving the lives of learners and teachers by providing evidence of the impact of early maths programmes on pupils’ maths learning/attainment outcomes. Therefore, it is in our legitimate interest to process and analyse personal data for the administration of this RCT and the analysis of its impact on maths outcomes for pupils. Details of all data processed by NFER for this project are also recorded in the project’s data log that is overseen by NFER’s Compliance Officer.

## Data Collection and Sharing

### School staff Data

At recruitment, personal data will be collected directly from participating schools by WRE. When a school signs up to join the trial, WRE will collect staff personal data about the Headteacher and key project contact directly from the schools. WRE will share this data with NFER via a secure digital data sharing portal. NFER will collect further personal data directly from schools including the Reception class teacher details for all participating Reception classes via a Teacher and Pupil data template and using the online survey platform, Tivian.

The following personal data will be collected about key project contacts and Reception teachers:

- Name
- Contact details
- Job role

The above personal data will also be collected for Headteachers, but without explicitly asking for job role (Headteacher will suffice).

WRE will also collect attendance data for Reception teachers, teaching assistants and maths lead at the PD training and coaching and mentoring sessions in the intervention schools and share this with NFER to support the compliance analysis.

NFER will also collect qualitative and quantitative online survey data from school staff (Headteachers, key contacts – usually the maths lead/early years lead – and Reception teachers and teaching assistants/support staff) to gather their views on the Reception Jigsaw programme. We will collect additional qualitative data on teachers’ attitudes towards the programme via interviews.

### **Pupil Data**

NFER will collect the following personal data for *all* children in Reception in 2025/26 directly from the participating school:

- Full name,
- Date of birth,
- Free School Meal (FSM) status.

In addition, we will collect the following personal data about each of the (up to) 24 sampled pupils from each school (across all participating classes):

- Unique Pupil Number (UPN),
- Responses to the baseline observational measure (Emerging Numeracy checklist completed by their class teacher),
- Early Years Foundation Stage Profile (EYFSP) outcomes from the National Pupil Database (NPD)

For the 20 sampled pupils who are selected for endpoint testing, we will collect the following additional data:

- Responses to the Early Years Toolbox Early Numeracy assessment (completed with an NFER Test Administrator)
- We will collect the following additional data from the DfE’s NPD at the end of the Reception year: sex; and eligibility for Free School Meals (FSM).

NFER will share the full name, date of birth and UPN of the *sampled* pupils with the DfE's NPD team to request the additional background data listed above. At the end of the evaluation, NFER will share pupils' outcome data (EYT Early Numeracy scores) with schools.

Three months after the publication of the evaluation report, pseudonymised<sup>7</sup> matched outcome data will be added to the EEF archive, which is managed by the Fischer Family Trust (FFT) on behalf of EEF and hosted by the Office of National Statistics (ONS). NFER will retain personal data for one year after report publication in case there are any queries about the report. One year after the report publication (expected to be May 2027), all personal data will be securely deleted.

## Data Processing Roles

NFER and WRE are independent data controllers. NFER is the independent data controller for the evaluation, which means that NFER determines the purposes of data processing for the evaluation. After trial data has been archived, EEF will become the data controller and is responsible for taking decisions about the means and purposes of processing. At that point, NFER will no longer be responsible for the data and will no longer be the data controller. FFT Education (<https://fft.org.uk/about-fft/>) and the designated EEF archive evaluator will act as data processors for the archive.

WRE is the independent data controller for programme delivery; for the purpose of the evaluation, WRE is the data processor.

---

<sup>7</sup> Pseudonymisation is a technique that replaces or removes information (like names or other meaningful identifiers) in a dataset that identifies an individual.

## Personnel

Name	Organisation	Role and Responsibilities
<b>Programme Management</b>		
Daniela Alvarado	EEF	Evaluation Manager – key point of contact for NFER
Ben Sillitoe	EEF	Programme Manager – key point of contact for WRE
<b>Evaluation Team</b>		
Pippa Lord	NFER	Project Director – strategic leadership and responsible for overall delivery and quality assurance of the trial
Aarti Sahasranaman	NFER	Trial Manager and IPE manager – day-to-day management of the trial and delivery of the trial design including integration of the IPE design, main point of contact for EEF
Katherine Aston	NFER	IPE Researcher – responsible for IDEA workshop, ToC development, qualitative IPE data collection
Eleanor Bradley	NFER	IPE Researcher – responsible for quantitative IPE, instrument design, qualitative IPE data collection
Chris Morton	NFER	Statistician
Kathryn Hurd	NFER	Research Operations Lead - overall data collection and school communications strategy
Katharine Stoodley	NFER	Operations Manager – day-to-day operations including preparation of recruitment documents, coordinating data collection and point of contact for schools participating in the trial
<b>Delivery Team</b>		
Kat Ellis	WRE	Head of EYFS – programme content and delivery
Jane Brown	WRE	Head of Primary Maths – programme content and delivery
Lloyd Ffrench	WRE	Managing Director – oversee and manage trial logistics
James Cheater	WRE	Director of External Programmes – oversee and manage trial logistics
Harry Metcalfe	WRE	External Programmes Manager – oversee and manage trial logistics

## Risks

Risk	Assessment	Controls, countermeasures and contingencies
<p><b>Insufficient schools recruited particularly due to the pause to the trial</b></p>	<p><b>Likelihood:</b> high <b>Impact:</b> high</p>	<ul style="list-style-type: none"> <li>• NFER provided input into the recruitment documentation.</li> <li>• Regular contact with WRE and EEF during recruitment period to monitor targets.</li> <li>• NFER provided recruitment support by sampling schools in the geographic area covered by the trial and contacting them with information about the trial and links to the EoI.</li> <li>• Sample size calculations revisited and design of the study altered to accommodate a lower recruitment target of 330 schools (original target was 400 schools). Revised design required 20 pupils at each school to be assessed (compared to 10 pupils in original design).</li> </ul>
<p><b>School attrition particularly due to the pause in the trial</b></p>	<p><b>Likelihood:</b> low <b>Impact:</b> moderate</p>	<ul style="list-style-type: none"> <li>• Clear initial and ongoing communication with schools explaining principles and expectations.</li> <li>• Clear communication and keep in touch emails sent to schools as soon the status of the trial was confirmed to reduce the likelihood of schools dropping out prior to randomisation.</li> <li>• Low testing burden (see below).</li> <li>• Named key contact(s) in school for direct communication.</li> <li>• Termly keep-in-touch to update any changes in contact and to keep schools informed of next steps.</li> <li>• Sample size calculations account for 5% school attrition.</li> </ul>
<p><b>Schools do not provide complete pupil data in the summer term thereby delaying baseline assessment timelines in autumn term</b></p>	<p><b>Likelihood:</b> medium <b>Impact:</b> high</p>	<ul style="list-style-type: none"> <li>• Remind schools of trial timelines particularly around pupil data submission and completion of baseline testing. Send reminders in summer term.</li> <li>• Pupil data email reminders and phone calls to schools to encourage submission of pupil data in summer.</li> </ul>

		<ul style="list-style-type: none"> <li>• In the summer, complete random sampling of pupils and preparation of baseline EN checklists in schools who have provided complete pupil data so that schools are ready to start baseline testing at the start of the autumn term.</li> <li>• Extend pupil data collection into the first (and if required) second week of September. Notify schools of September data submission deadline at the end of August.</li> <li>• Do two rounds of pupil sampling and baseline EN checklist preparation in September to ensure quick turn around and completion of baseline assessments in September 2025.</li> </ul>
<b>Pupil attrition from measurement</b>	<b>Likelihood:</b> low <b>Impact:</b> moderate	<ul style="list-style-type: none"> <li>• Random sample of 20 pupils per school accounts for 15% pupil-level attrition.</li> <li>• One or more of the up to 4 reserve pupils will be assessed at endpoint if some of the 20 randomly selected pupils are not available.</li> <li>• TA attendance is two days.</li> </ul>
<b>ICT issues with outcome measure (e.g. data transfer not successful) (NB the app does not require wi-fi in situ)</b>	<b>Likelihood:</b> medium <b>Impact:</b> high	<ul style="list-style-type: none"> <li>• TAs will test the EYT Numbers app on the iPads assigned to them ahead of the endpoint visit to minimise the risk of app failure during the visit.</li> <li>• Monitor the status of data transfer to ensure that data has been transferred to the correct location, e.g. study database.</li> <li>• Follow manual data extraction process if automatic data transfer fails.</li> </ul>
<b>Intervention is not implemented well</b>	<b>Likelihood:</b> low <b>Impact:</b> moderate	<ul style="list-style-type: none"> <li>• Clear information provided to schools explaining the principles of the trial and expectations.</li> <li>• Good communication with delivery team to provide strong implementation.</li> <li>• IPE will monitor implementation.</li> </ul>
<b>Changes to the project team due to sickness, absence or staff turnover</b>	<b>Likelihood:</b> moderate <b>Impact:</b> moderate	<ul style="list-style-type: none"> <li>• NFER has a large research department with numerous researchers and research associates experienced in evaluation who could be redeployed.</li> <li>• Clear and accurate project documentation would support continuity in the event of any team changes.</li> </ul>

## Timeline

Table 5: Timeline

Dates	Activity	Staff responsible/ leading
November – December 2024	Project Set-up <ul style="list-style-type: none"> <li>IDEA workshop and set-up meetings</li> </ul>	Pippa Lord, Aarti Sahasranaman, Katherine Aston
January – February 2025	Recruitment Documents	Pippa Lord, Aarti Sahasranaman, Kathryn Hurd, Katharine Stoodley
April – July 2025	Recruitment	WRE
	Grant Agreement	Aarti Sahasranaman
	Data Sharing Process <ul style="list-style-type: none"> <li>Finalise data flow diagram</li> <li>Sign Data Sharing Agreement</li> <li>Receive recruited school data from WRE</li> </ul>	Kathryn Hurd, Katharine Stoodley
	Pupil and Teacher Data Collection <ul style="list-style-type: none"> <li>NFER collect Reception pupil and teacher data from all recruited schools, with further data collection in September (see below)</li> </ul>	Kathryn Hurd, Katharine Stoodley
August – September 2025	Trial Protocol <ul style="list-style-type: none"> <li>Submit first draft of trial protocol in August 2025</li> </ul>	Aarti Sahasranaman
	Complete Pupil and Teacher Data Collection <ul style="list-style-type: none"> <li>NFER to collect pupil and teacher data from schools that did not provide data in the summer term</li> </ul>	Kathryn Hurd, Katharine Stoodley
	Random Sampling of Pupils <ul style="list-style-type: none"> <li>NFER to randomly sample up to 24 pupils across the Reception year group in all participating schools</li> </ul>	Chris Morton
	Baseline Data Collection <ul style="list-style-type: none"> <li>Reception teachers complete baseline EN checklist for sampled pupils</li> </ul>	Kathryn Hurd, Katharine Stoodley
	IPE <ul style="list-style-type: none"> <li>WRE staff interview</li> <li>Baseline BaU survey (all schools)</li> </ul>	Katherine Aston, Eleanor Bradley
October 2025	Randomisation <ul style="list-style-type: none"> <li>Schools randomised to control or intervention group</li> </ul>	Chris Morton

October 2025 – June 2026	WRM Reception Jigsaw Programme Delivery	WRE, Schools
	IPE <ul style="list-style-type: none"> <li>• CPD training observations</li> <li>• Trainer focus groups</li> <li>• Case study visits</li> </ul>	Katherine Aston, Eleanor Bradley
	Training Attendance Data <ul style="list-style-type: none"> <li>• WRE to collect attendance at PD training and coaching sessions and share with NFER</li> </ul>	WRE, Kathryn Hurd, Katharine Stoodley
	Statistical Analysis Plan (SAP) <ul style="list-style-type: none"> <li>• Submit first draft of SAP (January 2026)</li> </ul>	Chris Morton
	Submit NPD Application (February 2026)	Chris Morton
June – July 2026	Endpoint Data Collection <ul style="list-style-type: none"> <li>• NFER Test Administrators complete Early Years Toolbox Early Numeracy app-based assessment with 20 sampled pupils</li> </ul>	Kathryn Hurd, Katharine Stoodley
	IPE <ul style="list-style-type: none"> <li>• Endpoint BaU survey (control schools)</li> <li>• Endpoint Reception practitioner survey (intervention schools)</li> <li>• Endpoint headteacher survey (intervention schools)</li> </ul>	Katherine Aston, Eleanor Bradley
July – August 2026	Data Processing <ul style="list-style-type: none"> <li>• Data cleaning, DQA and preparation of final dataset</li> </ul>	Kathryn Hurd, Katharine Stoodley
September 2025 – January 2026	Impact and IPE Data Analysis	Chris Morton, Katherine Aston, Eleanor Bradley
February 2027	First Draft of Report to EEF	Aarti Sahasranaman
February – July 2027	Four review rounds and publication of final report <ul style="list-style-type: none"> <li>• EEF review, peer review, delivery partner review and proofreader/publication-ready review</li> </ul>	Aarti Sahasranaman, Pippa Lord
Autumn 2027	Data Archiving	Chris Morton, Aarti Sahasranaman

## References

- Aubrey, C., Godfrey, R. and Dahl, S. (2006) 'Early mathematics development and later achievement: further evidence', *Mathematics Education Research Journal*, 18, pp. 27–46. Available at: <https://doi.org/10.1007/BF03217428>.
- Bulus, M., Dong, N., Kelcey, B. and Spybrook, J. (2021) 'PowerUpR: power analysis tools for multilevel randomized experiments'. Available at: <https://cran.r-project.org/web/packages/PowerUpR/index.html> (Accessed: 12 August 2025).
- EEF (2023) *Early numeracy approaches*, EEF. Available at: <https://educationendowmentfoundation.org.uk/early-years/toolkit/early-numeracy-approaches> (Accessed: 12 August 2025).
- Gilmore, C. (2023) *Labour's plan to focus on early maths is solid – gaps in achievement start even before primary school*, *The Conversation*. Available at: <http://theconversation.com/labours-plan-to-focus-on-early-maths-is-solid-gaps-in-achievement-start-even-before-primary-school-215498> (Accessed: 22 August 2025).
- Howard, S., Neilsen-Hewett, C., de Rosnay, M., Melhuish, E. and Buckley-Walker, K. (2022) 'Validity, reliability and viability of pre-school educators' use of early years toolbox early numeracy', *Australasian Journal of Early Childhood*, 47(2), pp. 92–106. Available at: <https://doi.org/10.1177/18369391211061188>.
- Melhuish, E. and Gardiner, J. (2020) *Study of early education and development (SEED): impact study on early education use and child outcomes up to age five years*. Available at: [https://assets.publishing.service.gov.uk/media/5e4e5c10e90e074dcd5bd213/SEED\\_AGE\\_5\\_REPORT\\_FB.pdf](https://assets.publishing.service.gov.uk/media/5e4e5c10e90e074dcd5bd213/SEED_AGE_5_REPORT_FB.pdf) (Accessed: 12 August 2025).
- Outhwaite, L., A., Anders, J. and Herwegen, J.V. (2022) *Mathematics attainment falls behind reading in the early primary school years*. Working Paper No. 22-06. Available at: <https://repec-cepeo.ucl.ac.uk/cepeow/cepeowp22-06.pdf> (Accessed: 12 August 2025).
- Sims, S., Fletcher-Wood, H., O'Mara-Eves, A., Cottingham, S., Stansfield, C., Herwegen, J.V. and Anders, J. (2021) *What are the characteristics of effective teacher professional development? A systematic review and meta-analysis*. Available at: <https://d2tic4wvo1iusb.cloudfront.net/production/documents/pages/Teacher-professional-development.pdf?v=1727862311> (Accessed: 12 August 2025).
- STA (2025) *Reception baseline assessment: privacy notice for pupil data from September 2025*, GOV.UK. Available at: <https://www.gov.uk/government/publications/reception-baseline-assessment-privacy-notice/reception-baseline-assessment-privacy-notice-for-pupil-data-from-september-2025> (Accessed: 8 September 2025).
- Tang, S., Bradley, E., Martin, K., Schwendel, G. and Styles, B. (2024) *Randomised controlled trial evaluation of the White Rose Maths Reception Jigsaw*. Available at: <https://d2tic4wvo1iusb.cloudfront.net/production/documents/projects/Reception-Jigsaw-addendum-report-final.pdf?v=1737993539> (Accessed: 10 June 2025).
- The Labour Party (2024) *Change: Labour party manifesto 2024*, *The Labour Party*. Available at: <https://labour.org.uk/change/> (Accessed: 14 August 2025).
- The R Foundation (2025) 'The R project for statistical computing'. Available at: <https://www.r-project.org/> (Accessed: 12 August 2025).
- Tymms, P., Merrell, C. and Bailey, K. (2017) 'The long-term impact of effective teaching', *School Effectiveness and School Improvement*, 29(2), pp. 242–261. Available at: <https://doi.org/10.1080/09243453.2017.1404478>.

Werner, K., Acs, G. and Blagg, K. (2024) *Comparing the long-term impacts of different child well-being improvements*. Available at: [https://www.urban.org/sites/default/files/2024-03/Comparing\\_the\\_Long-Term\\_Impacts\\_of\\_Different\\_Child\\_Well-Being\\_Improvements.pdf](https://www.urban.org/sites/default/files/2024-03/Comparing_the_Long-Term_Impacts_of_Different_Child_Well-Being_Improvements.pdf) (Accessed: 12 August 2025).

# Appendix A: Changes since the previous EEF evaluation

Appendix Table A: Changes since the previous evaluation

Feature		Efficacy to effectiveness stage
Intervention	Intervention content	N/A
	Delivery model	Some minor changes have been made to the intervention model. For example, the spatial reasoning PD training session has been moved from module 5 to module 3 and some gap tasks have been changed.
	Intervention duration	N/A
Evaluation	Eligibility criteria	<p>There are two differences in the school-level eligibility criteria between the efficacy and effectiveness trials:</p> <ol style="list-style-type: none"> <li>(1) The efficacy trial was limited to state schools in Yorkshire or Essex whereas this effectiveness trial has a wider geographic spread. The list of Local Authority regions from which schools were recruited to the trial has been provided in the main body of the protocol.</li> <li>(2) The efficacy trial only included schools with standalone Reception classes. Schools with mixed year groups were not included. This is not an inclusion criterion in the effectiveness trial.</li> </ol> <p>There are two differences in the school-level pupil sampling approach between the efficacy and effectiveness trials:</p> <ol style="list-style-type: none"> <li>(1) An initial random sample of 28 pupils including a group of eight replacement pupils was selected from each trial school in the efficacy trial. The eight pupils in the replacement group were treated as standbys in case any of the 20 trial sample pupils in their school could no longer participate in the trial (e.g., they did not start school or were withdrawn from the trial by their parents). The replacement process could only happen prior to randomisation. In the effectiveness trial, an initial random sample of up to 24 pupils will be selected from each trial school. Schools will be asked to complete the checklist for all 24 pupils (if possible). At endpoint, 20 pupils will be sampled from the baselined pupils, for EYT testing, with up to 4 baselined replacements available (e.g. should a pupil be absent on the day of testing). The replacement process will happen at endpoint, aiding the size of the analysed sample.</li> <li>(2) Pupils identified by schools as FSM-eligible were oversampled in the efficacy trial, so that 25% of sampled pupils were eligible for FSM. This will not be the case in the effectiveness trial.</li> </ol>
	Level of randomisation	N/A
		Outcomes and baseline

		effectiveness trial. Teachers' confidence in teaching early maths and in their own mathematical ability were secondary outcome measures in the efficacy trial. There are no secondary teacher outcome measures in this effectiveness trial, as response rates were regarded to not be high in the efficacy trial and there are some issues with reporting confidence in teaching abilities. The effectiveness trial will explore teacher outcomes in the IPE.
	<b>Control condition</b>	N/A

## Appendix B: Baseline Early Numeracy (EN) Checklist

Appendix Table B2: Baseline Early Numeracy (EN) Checklist

Please indicate for [insert pupil's name] whether, based on what you have observed since they started Reception, they can do each of the following tasks with support or independently.		cannot do with support	can do with support	can do independently
1	Recognises numerals 1 to 5			
2	Counts up to 3 or 4 objects by saying one number for each item in order			
3	Subitises (recognises quantities without counting) up to 5			
4	Counts objects up to 10			
5	Counts objects beyond 10			
6	Counts out a smaller number of objects (up to 6) from a larger group - e.g. "give me four of the..."			
7	Selects the correct numeral to represent 1 to 5 objects			
8	Selects the correct numeral to represent 6 to 10 objects			
9	Estimates how many objects they can see and checks by counting them (up to 10)			
10	Uses the language of 'more than' and 'less than/fewer' to compare two sets of objects			
11	Finds the total number of items in two groups by counting all of them (up to 10)			
12	Finds one more or one less from a group of up to 5 objects			
13	Finds one more or one less from a group of up to 10 objects			

14	Records mathematical ideas using marks or objects they can interpret and explain			
15	Selects a particular named shape			
16	Can describe the relative position of objects or shapes such as 'behind' or 'next to'			
17	Uses familiar objects and common shapes to create and build models			
18	Continues, copies and creates repeating patterns			
19	Uses everyday language related to time			
20	Orders and sequences familiar events			