Randomised controlled trial evaluation of the White Rose Maths Reception Jigsaw

Evaluation report (updated with longitudinal follow-up)

May 2024

Main report originally published: October 2023

Sarah Tang, Eleanor Bradley, Kerry Martin, Gemma Schwendel, and Ben Styles

The Education Endowment Foundation is an independent charity dedicated to breaking the link between family income and education achievement. We support schools, nurseries and colleges to improve teaching and learning for 2 – 19-year-olds through better use of evidence.

We do this by:

- **Summarising evidence.** Reviewing the best available evidence on teaching and learning and presenting in an accessible way.
- **Finding new evidence.** Funding independent evaluations of programmes and approaches that aim to raise the attainment of children and young people from socio-economically disadvantaged backgrounds.
- **Putting evidence to use.** Supporting education practitioners, as well as policymakers and other organisations, to use evidence in ways that improve teaching and learning.

We were set-up in 2011 by the Sutton Trust partnership with Impetus with a founding £125m grant from the Department for Education. In 2022, we were re-endowed with an additional £137m, allowing us to continue our work until at least 2032.

For more information about the EEF or this report please contact:

Education Endowment Foundation
5th Floor, Millbank Tower
21–24 Millbank
SW1P 4QP

0207 802 1653

info@eefoundation.org.uk

www.educationendowmentfoundation.org.uk

Department for Education

What Works Network

Contents

About the evaluator

The project was independently evaluated by a team from the National Foundation for Educational Research (NFER). The trial director and principal investigator for this study was Dr Ben Styles, Head of NFERs Education Trials Unit. Helen Poet, Connie Rennie, and Sarah Tang led the evaluation team and the impact evaluation. Kerry Martin, Research Manager, led the implementation and process evaluation. They were supported by: David Sims as process evaluation director; Eleanor Bradley as process evaluation researcher; and Dr Joana Andrade, Chris Morton, and Gemma Schwendel as trial statisticians. Kathryn Hurd led the research operations with Guido Miani, Alison Hale, Asma Ullah, and Guvi Chohan as operations researchers.

Contact details:
National Foundation for Educational Research
The Mere,
Upton Park,
Slough,
Berkshire SL1 2DQ,
United Kingdom
Telephone: +44 (0)1753 574123
Email: s.tang@nfer.ac.uk

Acknowledgements

We are grateful to all the primary schools that participated in this evaluation. We are especially thankful to the school staff members who completed the baseline assessments and the staff and pupils who took part in the case study visits and interviews. We would like to give a special mention to the staff at White Rose Maths who have worked with us to support the evaluation whenever needed, namely Caroline Hamilton, Tony Staneff, Jane Brown, Jane Milner, and Kat Ellis.

We would like to thank the team at the Education Endowment Foundation, including Jennifer Stevenson, Celeste Cheung, Amy Clark, Maria Pomoni, and Ben Simuyandi.

ONS SRS publication status and statistical results

This output request has been granted publication-level clearance (as of 15 August 2023 and11th April 2024). All statistical results remain Crown Copyright.

This work contains statistical data from the Office for National Statistics (ONS), which is Crown Copyright. The use of the ONS statistical data in this work does not imply the endorsement of the ONS in relation to the interpretation or analysis of the statistical data. This work uses research datasets which may not exactly reproduce National Statistics aggregates. The analysis was carried out in the Secure Research Service, part of the Office for National Statistics.

Note on updated report

This report is an update of the original report published in October 2023. The revised version contains the results from the longitudinal follow-up and some additional exploratory analysis on the main trial data.

The longitudinal follow-up was planned from the outset and is referenced in the protocol (Poet *et al*, 2021) and the statistical analysis plan (Andrade *et al.*, 2022). The main additions to this report relating to the longitudinal follow-up can be found in the ‘Secondary outcomes’ section (where the assessment used is described), the ‘Longitudinal analysis’ section (where the statistical analysis is outlined), and the ‘Impact evaluation results’ section (where the results are presented. The ‘Participant flow including losses and exclusions’ diagram and subsequent tables showing the minimum detectable effect size and attrition have been updated to include information relating to the longitudinal follow-up. The findings have also been discussed in the ‘Conclusion’.

The second addition to the original report is some additional secondary analysis looking at the impact of Reception Jigsaw on the two separate mathematics scales in the Early Years Foundation Stage Profile. This was not prespecified in the protocol or statistical analysis plan and is therefore considered exploratory. The additions relating to the exploratory analysis can be found in the ‘Additional secondary analysis’ and ‘Secondary analysis’ sections.

Executive summary

The project

Reception Jigsaw is a professional development training package that aims to improve mathematics teaching quality and outcomes in reception classes. It is primarily aimed at reception teachers and reception teaching assistants (TAs). 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 involves five twilight Continuing Professional Development (CPD) sessions and five coaching sessions over an academic year, with gap tasks to be completed between CPD sessions. The tasks are based on content covered in the CPD session and are followed up during the coaching visit. Although the content is predominantly aimed at reception teachers, the training covers how the content can be extended into Year 1. The training is developed and delivered by White Rose Maths.

This project tested the impact of Reception Jigsaw on all children in reception classes (ages 4–5). Reception teachers, Year 1 teachers, and the setting mathematics lead attended the CPD sessions, which were face-to-face twilight sessions. The reception teachers also were expected to do the gap tasks and take part in the half-day coaching visits, which were optional for other staff (although the mathematics lead had to participate in part of the visit). Reception teachers received handout materials during the training and physical classroom resources. They were expected to use what they have learned in the training to inform their teaching and delivery of mathematics in the classroom.

Reception Jigsaw was evaluated using a two-arm randomised controlled trial. This was an efficacy trial in which 138 schools were randomised to either receive the intervention or continue with business as usual teaching. The primary outcome was performance on a maths test. A process evaluation involved interviews with teachers, observations of training sessions and coaching visits, and surveys given to teachers and TAs. Recruitment for the trial originally started in November 2019 and was expected to run in schools in Autumn Term 2020. Recruitment was paused in March 2020 due to the Covid-19 pandemic. Recruitment reopened in November 2020, programme delivery took place in the 2021/2022 academic year and the post-test took place in June 2022. Longitudinal follow-up assessments were undertaken in summer 2023 to assess the impact of Reception Jigsaw on mathematics attainment at the end of Year 1.

Table 1: Conclusions

Key conclusions	
1.	Children in Reception Jigsaw schools made the equivalent of one additional month of progress in maths, on average, compared to children in other schools. This result has a high-security rating. These results are not statistically significant. This means that the statistical evidence does not meet the threshold set by the evaluator to conclude that the true impact was non-zero
2.	Children eligible for free school meals made no additional progress in maths, on average, compared to children in other schools. This result has a lower security rating than the overall finding because of the smaller number of pupils
3.	There was evidence that the Reception Jigsaw programme led to reception teachers being more confident in teaching mathematics
4.	The programme was well received by schools. Reception teachers reported positively about the support they received throughout the Reception Jigsaw programme, finding it relevant to their teaching practice, high quality, and impactful in terms of the changes they made to their practice and environment
5.	Year 1 teachers also reported positively about the support, but less so compared to reception teachers. However, they could see that it would help them with supporting pupils to transition from reception to Year 1, and to build on knowledge pupils had gained in reception
6.	Longitudinal analysis found that children in Reception Jigsaw schools in reception made an average of one month's additional progress at the end of Year 1 compared to children in other schools. However, the statistical evidence does not meet the standard set by the evaluator to conclude that the true impact was non-zero.

The EEF security rating

These findings have a high-security rating. This was an efficacy trial, which tested whether the intervention worked under developer-led conditions. The trial was a well-designed and well-powered randomised controlled trial. Around 14% of

the pupils who started the trial were not included in the final analysis because their school did not provide test data. The pupils in Reception Jigsaw schools were similar to those in the comparison schools in terms of prior attainment.

Additional findings

Pupils in Reception Jigsaw schools made, on average, one additional month of progress than those in the control group equivalent. This is our best estimate of impact, which has a high-security rating. As with any study, there is always some uncertainty around the result: the possible impact of this programme also includes small negative effects of one month of less progress and positive effects of up to three months of additional progress. Given this confidence interval overlaps zero, the evaluator was unable to conclude this was a genuine effect. The longitudinal follow-up found that the additional month's progress made by pupils who were in Reception Jigsaw schools seen at the end of their reception year remained at the end of Year 1 but again there is not enough statistical evidence to be confident of a genuine non-zero effect as the confidence interval includes zero.

Teachers were expected to attend nine of the ten possible sessions (five twilight CPD sessions and five coaching sessions) to be judged as compliant to the programme. Around 65% of intervention pupils were taught by teachers who had complied with the programme requirement. There was evidence of a greater positive effect on mathematics attainment when pupils were taught by teachers who complied with this target. In this case, the effect size was estimated to be the equivalent to two additional months of progress compared to pupils in schools that did not receive Reception Jigsaw training. The majority of teachers taking part in Reception Jigsaw attended the CPD sessions, so differences in compliance likely reflect differences in attendance at coaching sessions.

Findings from the evaluation suggest that improved professional collaboration is another outcome of the intervention. Mathematics leads and senior leaders praised the inclusive nature of the training, which had allowed staff working in different year groups and at different levels to attend. They commented that it was an opportunity for staff to work with colleagues outside of their year group and had initiated professional discussions around mathematics teaching.

The Reception Jigsaw logic model expects that the support provided will lead to changes to the reception learning environment (output), changes to pedagogy (output), and that teachers and TAs will apply learning in the classroom (mediator). Findings from the endpoint survey, trainer reflection logs, and case study interviews suggest that most reception teachers did apply learning from the Reception Jigsaw, embedding activities into their mathematics teaching and continuous provision and adapting the classroom environment. All reception teachers who responded to the endpoint survey said they applied learning at least once a week, around two-thirds (68%) reported applying it in four to five mathematics lessons per week, and around a third (31%) applied learning in two to three lessons per week.

Cost

The average cost of Reception Jigsaw for one school is around £5,807, or £45.50 per pupil per year when averaged over three years. The cost to schools varies depending on distance from the WRM head office in Halifax, and the figure provided here is the upper estimate. Schools are expected to cover the cost of travel and accommodation where this is needed, and this has not been included in these calculations.

Impact

Table 2: Summary of impact on primary outcome

Outcome/Group	Effect size (95% confidence interval)	Estimated months' progress	EEF security rating	No. of pupils	EEF cost rating
Maths	0.08 (-0.07, 0.23)	1		952	£ £ £ £ £
Maths (FSM-eligible pupils)	-0.02 (-0.2, 0.2)	0	N/A	216	£ £ £ £ £

Introduction

Background

Mathematics ability has been shown to be an important determinant of economic, social, and health outcomes later in life (Davis-Kean *et al*., 2021; Ritchie and Bates, 2013). Attainment gaps in mathematics between those pupils from different socioeconomic backgrounds are evident at each stage of schooling. Attainment gaps appear early and often persist throughout a child's schooling career (Farquharson *et al.*, 2022).

As highlighted by Outhwaite *et al*. (2022a), there has been a gap between early reading attainment and maths attainment (Outhwaite *et al.,* 2022a). There is also some evidence that gaps in formal schooling due to Covid-19 disruptions may have had more of an impact on maths than reading for those in the initial stages of schooling (DfE, 2022a). Classroom practice in early years plays an important role in improving mathematics attainment (Outhwaite *et al.*, 2022b). Recent work surveying preschool practitioners (Costa *et al.*, 2021), found that practitioners were significantly more likely to focus on literacy rather than numeracy in their daily routines in the classroom and that mathematics activities were often more formal rather than play-based. The Effective Pre-school, Primary and Secondary Education (EPPSE) study (Taggart *et al.*, 2015) found that high-quality preschool provision (in particular one promoting early number concepts) was an important determinant of mathematics attainment at both age 11 and 14.

The Reception Jigsaw is a Continuing Professional Development (CPD) programme developed and delivered by White Rose Maths (WRM). The programme is primarily for reception teachers and involves five twilight CPD sessions and five coaching sessions over an academic year. The delivery model involves five cycles of a CPD session followed by a coaching visit around two weeks later, with a gap task (set at each CPD session) to be completed between sessions. The gap tasks are based on content covered in the CPD session and are followed up during the coaching visit.

WRM was created in 2017 to develop work being undertaken by Trinity multi-academy trust (MAT) around resources, schemes of work, and CPD for mathematics teachers. WRM provide mathematics resources, schemes of work, and training packages to schools both in the United Kingdom (UK) and internationally. In 2019, WRM reported that their schemes of work were being used in over 80% of primary schools in the UK and in 140 countries (WRM, n.d.). WRMs signature training programme is the Primary Jigsaw, which supports teachers from Years 1 to 6 in improving their mathematics teaching.

WRM developed the Reception Jigsaw in response to demand from schools for a CPD training package for mathematics aimed at reception teachers. Schools had also fed back to WRM that reception teachers often miss out on training opportunities. More broadly, it is recognised that early years teachers in particular often require support to develop their subject knowledge, knowledge of children's development, and pedagogical knowledge. Recent revisions to the early years curriculum and assessment have also increased demand for more CPD in early years teaching.

The early years curriculum in England is determined by the early years foundation stage (EYFS) framework, which sets out the education programmes for each of the seven areas of learning (DfE, 2021a). Within these seven areas, assessment against 17 Early Learning Goals (ELGs) takes place at the end of the reception year constituting the EYFS profile (EYFSP). Teachers are asked to assess whether each pupil has reached the expected level of development ('expected') or not ('emerging') for each of the 17 ELGs.[1]

Following a 2017 consultation of primary assessment in England, the Department for Education (DfE) reviewed the ELGs in order to ensure they were aligned to the Key Stage 1 curriculum, to make the descriptors for typical levels of development clearer and to explore whether the number of ELGs could be reduced (DfE, 2017). Mathematics and literacy in particular were flagged as needing attention in terms of reviewing and revising 'the mathematics and literacy ELGs to ensure that they support children to develop the right building blocks for learning at key stage 1' (DfE, 2017, p. 7). Revised ELGs were piloted in 2018 in 24 schools and an evaluation of the pilot was commissioned by the Education Endowment Foundation (EEF) ahead of a national rollout of revised ELGs (Husain *et al.*, 2019). The evaluation found that while the pilot schools viewed the draft ELGs positively, there was no consensus about whether reception children would be better prepared for Key Stage 1 as a result (Husain *et al*., 2019). The revised ELGs became statutory from September 2021 (DfE, 2022b).

[1] Prior to 2022 EYFSP there was a third level for assessment against ELGs in the EYFSP ('exceeding').

A key change in the revised ELGs for mathematics was the lack of reference to space, shape, and measure, with the two ELGs for mathematics looking at number and numerical patterns only. While this raised concern for some in the sector, others felt that the move allows practitioners to focus on number (specifically up to ten) developing a deeper understanding and more secure numeracy ahead of starting Key Stage 1 (James, 2022). Although no longer an ELG, space, shape, and measure are still part of the early years curriculum directly mentioned in the educational programme for mathematics (DfE, 2022b). The revised mathematics educational programme is much more detailed, articulating that pupils should have 'frequent and varied opportunities' to build and apply their understanding of counting, numbers to ten, and the relationships and patterns between them (DfE, 2022b). The educational programme also highlights the importance of confidence and developing a 'have a go' attitude with specific mention of talking to adults and peers about what they notice (DfE, 2022b). The Reception Jigsaw was designed with the new curriculum in mind.

Turning to the existing research literature around the role of CPD to improve mathematics teaching in the early years, a recent review commissioned by the EEF on early years and Key Stage 1 mathematics teaching (Hodgen *et al.*, 2020) concluded that while experimental evidence was lacking in the area of early years professional development, there was 'some evidence from expert-judgment-based reviews suggesting that coaching together with instructional feedback may be effective, particularly for manualised, or well-described, interventions' (Hodgen *et al.*, 2020, p. 64). The EEF guidance for practitioners based on this review of the evidence put forward five recommendations for improving mathematics in early years and Key Stage 1, with recommendation one citing developing practitioners:

- develop practitioners' understanding of how children learn mathematics;
- dedicate time for children to learn mathematics and integrate mathematics throughout the day;
- use manipulatives and representations to develop understanding;
- ensure that teaching builds on what children already know; and
- use high-quality targeted support to help all children learn mathematics (EEF, 2021).

Developing practitioner confidence alongside their skills may also have an important role in improving attainment in early years mathematics as more confident practitioners are likely to implement mathematics activities in the early years (Costa *et al.,* 2021; Knowles, 2017).

Prior to this trial, Reception Jigsaw had not been formally evaluated but had been piloted in a small number of schools. The content components are well grounded in existing evidence (see Appendix D) and the delivery model is based on WRMs signature training package—Primary Jigsaw. The evaluation of Reception Jigsaw was a two-arm cluster randomised controlled trial comparing the outcomes of pupils and teachers whose schools were assigned to take part in the Reception Jigsaw programme with those who did not. It was an efficacy trial as there has not been a trial of the programme to date and the developers delivered the intervention in the schools allocated to the intervention group. Efficacy trials are trials undertaken under developer-led conditions where there has not been a robust impact evaluation to determine causal links between the intervention and outcomes previously (see EEF, n.d.).

Intervention

WRMs Reception Jigsaw is a CPD programme primarily for reception teachers and reception teaching assistants (TAs). It involves five cycles of a twilight CPD session and a coaching visit. All ten sessions take place at the practitioners' school and the coaching, in particular, is tailored to the school's needs. The programme is delivered throughout the academic year. Details of the programme are given using the Template for Intervention Description and Replication (TIDieR) framework below.

Initial recruitment began in November 2019 but was then paused due to disruptions to schooling relating to the Covid-19 pandemic. The trial was subsequently delayed by a year with recruitment re-starting in November 2020. The programme ran in schools from November 2021 to May 2022. The developer (WRM) had planned to recruit new trainers in Essex and operate on a franchise/host school model to deliver the programme in this area, but this was not possible due to Covid-related restrictions. In a change from the original plan, trainers from Yorkshire travelled to Essex to deliver the training. This changed the delivery model in these schools as the twilight CPD session and the coaching visits were delivered on the same day (as opposed to two weeks apart).

There were also small changes to programme delivery from what was originally planned due to Covid restrictions. Some twilight CPD sessions used a hybrid approach where some practitioners attended in person and others remotely on Zoom. Two twilight sessions were held entirely on Zoom.

One further change to the original protocol was associated with testing. At design stage, Hodder's Progress in Understanding Mathematics Assessment (PUMA) was considered the most appropriate assessment for this trial as it had most recently been standardised to a UK population, the group nature of the assessment reduces costs and disruption in schools, and it had content alignment with the current national curriculum and EYFS framework (see protocol) (Poet *et al.*, 2021). However, the early years curriculum was revised with a new EYFS statutory framework being implemented in all schools from September 2021 (DfE, 2021a). In autumn 2021, after analysis of the revised PUMA assessment (New PUMA), it was felt that it did not align as well with the revised framework so other assessment options were considered in depth. WRM also highlighted that the nature of the assessment (pen and paper test without manipulatives) was not in line with their approach (and current broader thinking in early years pedagogy). A broad range of alternative assessments were considered, including the mathematics scales of the British Ability Scales 3 (BAS3) and online programmes using tablets, but nothing suitable was found to match the requirements of the trial. The New PUMA was retained as the primary outcome measure and an additional secondary measure was added, which was a subset of the New PUMA where questions covering topics not in the EYFS framework were removed (reduced PUMA – see 'Outcome measures' section).

TIDieR checklist

1. Brief name

White Rose Maths Reception Jigsaw

2. Why (rationale/theory)

WRMs Reception Jigsaw is a professional development training package that aims to improve mathematics teaching quality and outcomes in reception classes. Building on their popular CPD programme, the 'White Rose Maths Primary Jigsaw' (which covers Years 1–6), Trinity MAT have developed a training package for reception. WRM received a large demand for their primary training package and feedback from schools indicated that they felt that reception teachers often miss out on training opportunities. In response to this, WRM developed the Reception Jigsaw, aiming to improve the specialist mathematics pedagogical skills and knowledge of reception teachers and reception TAs. Other Key Stage 1 teaching staff were also invited to attend the sessions (particularly Year 1 teachers), to build on the delivery/teaching in reception. In particular, Year 1 teachers were invited to attend the training because the development of early number sense and early calculation strategies continues into Year 1. Although the content is predominantly aimed at reception teachers, the training covers how the content can be extended into Year 1. It is also useful for the Year 1 teachers to have an understanding of what has been covered in reception, in order to support those children who may be working below the requirements of the Key Stage 1 curriculum. Including both reception and Year 1 teachers (and in some schools, Year 2) provided the opportunity for discussion and comparison of how key learning points may be introduced in reception and in Year 1.

The Reception Jigsaw 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. Research shows that high-quality early numeracy education has the potential to have lasting positive effects that may help to narrow the gap in achievement throughout life. The Early Years Toolkit highlights the importance of professional development in supporting early numeracy approaches (EEF, n.d.). Key areas for improvement include: supporting practitioners' knowledge of mathematics; knowledge of children's development and development trajectories in mathematics; and understanding of the kinds of activities that support early mathematical learning. WRM have drawn on the What Works Clearinghouse 'Teaching Math to Young Children' review and practice recommendations (IES, 2013), and the National Centre for Excellence in the Teaching of Mathematics (NCETM) progressions for the early years. The evidence base for each of the five modules is shown in Appendix D.

3. Who (recipients)

Table 3 below shows, which members of staff take part in each element of the jigsaw; these people are the direct recipients of the training. Pupils are then recipients of the programme through being taught mathematics by a trained practitioner who puts into practice what they have learned through the CPD and coaching sessions.

Table 3: Participants of the elements of training

	Face-to-face twilight training	Gap tasks	Half-day coaching visit
Reception teachers	Essential	Essential	Essential (in 2+ form entry, participation in each coaching visit will be by one of the teachers, not both/all. Different reception teachers may take part in different coaching visits)
Reception teaching assistants	Recommended	Optional	Not expected to attend
Year 1 teachers	Essential	Optional	Not expected to attend
Year 1 teaching assistants	Recommended	Optional	Not expected to attend
Year 2 teachers and teaching assistants	Optional	Optional	Not expected to attend
Mathematics lead	Essential	Optional	Essential (although could be for just part of the visit)

4. What (materials)

The materials specific to the schools that participated in the intervention group:

- *Resources provided in training:* WRM provides dot plates and number cards for numbers zero to five to teachers in training session 1 and for larger numbers (6–10) in training session 3. WRM share a bank of resources and books with trainers to use during the training, depending on the needs of the school and what comes up in their discussions.
- *Gap tasks*: At the end of each twilight session the participants are set a 'gap task' related to the training given, to be completed before the next session (also see section on 'What (procedures, activities, or processes used)' below). The gap tasks are listed in the same booklet as the journal provided by WRM.
- *Journal to record individual reflections and progress:* This is a tool for reflection by those participating in the Reception Jigsaw twilight sessions. 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.
- *Videos demonstrating teachers modelling practice covered in the modules:* These were originally planned as part of the intervention. These videos were to be made available only to participant schools via a log-in. However, due to logistical challenges due to Covid-related restrictions, these videos did not form part of the intervention.

The materials available to all schools:

- *Online resources.* Publicly available to all schools on the WRM website, these include teacher guidance, schemes of learning, and interactive whiteboard presentations. Awareness of these resources may be higher in the intervention schools due to their ongoing participation in the training. However, the control group would also have had some awareness of the online resources as they were mentioned as part of the recruitment to the trial. High proportions of all trial schools stated they used WRM schemes of work in the baseline business as usual proforma.

5. What (procedures, activities, or processes used)

The training received by the schools that participated in the intervention group:

- **Five twilight in-depth training sessions** (each two hours), delivered face-to-face at each school. Each school receives their own training (i.e. not grouped with other schools)—for attendees, see section on 'Who (recipients)' above. In the trial, sessions were spaced across the reception year, between November 2021 and May 2022. Due to Covid-related restrictions, a few CPD sessions were delivered using a hybrid model, where some

practitioners were in school and others joined remotely (via Zoom). Two sessions out of all those delivered were delivered entirely remotely.

- *Sessions are delivered using slides*: Slides are the same across schools, but discussions and the sessions' focus are 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.
- Schools receive handouts summarising key messages from each session.
- Sessions are trainer-led from the front, with opportunities for discussions about some of the issues currently faced by the school/approaches taken.
- The training sessions include practical activities and games that schools could use in their own classrooms.
- **Five gap tasks**, to be completed between twilight sessions by reception teachers to encourage implementation and reflection of the learning in their teaching.
- Gap tasks are based on the content of the twilight 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 specialist leader in education (SLE) during the half-day coaching visit (see below).
- **Five half-day coaching visits** from an early years SLE.
- These sessions are expected to be attended by (at least one of) the reception teachers. The school mathematics lead also usually attends at least part of the coaching session. This is to reduce the burden on schools and the need for classroom cover. In two-form entry schools (or larger), the coaching sessions may be shared between the reception teachers.
- The aim of these visits is 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 visits are made after the gap tasks and are an opportunity to reflect on the gap task with the SLE. The model used in Essex varied from the standard approach due to challenges in recruiting trainers in the Essex area: Trainers from Yorkshire were used to deliver the programme in Essex. In order to reduce costs, trainers travelled to Essex five times, combining the coaching with the next CPD session (see 'Adaptations' section below).
- These sessions are more tailored than twilight sessions, but are still based on the topic covered in the previous twilight session. The school and SLE agree how to use the time—this could involve coaching, lesson observations, or reviewing the organisation of the classroom in relation to mathematics. Different schools may focus the time differently, for example some schools may opt to upskill a reception teacher to become an expert in mathematics teaching in the early years, while others may use the time to support newer or less confident reception teachers in their mathematics teaching.
- Twilight training sessions are completed outside of the school day whereas half-day sessions are held during the school day.

The Reception Jigsaw was delivered to each school between November 2021 to May 2022. The aim is that the same person (SLE/trainer) delivers all the sessions in a particular school—allowing for continuity and relationship building.

The five sessions/modules were:

- **Developing early number sense:** Focuses on counting principles, subitising, composition of number, comparison, and number relationships. The gap task is focused around building in activities to subitise through daily inputs and adaptations to the classroom provision.

- **Creating a mathematical classroom:** Themed around attitudes to learning and incorporating everyday mathematics through classroom routines. The module also covers the role of the adult in supporting learning and planning for adult-led activities, with the gap task revolving around planning a sequence of learning to include a balance of adult-led and independent play-based activities.

- **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 asks teachers to trial and observe the quality of talk in the classroom.

- **Reasoning and problem solving in the early years:** Focuses on classroom culture, developing reasoning, and problem solving (including odd and even, and doubling) through games and stories. The gap task is about trialling some of the suggested strategies for developing reasoning and problem solving in the classroom.

- **Exploring pattern and shape:** Discusses the importance of pattern spotting, progression through pattern, patterns through story, and looking at shapes. The gap tasks focus on choosing at least two areas of provision and considering how they could be enhanced to include opportunities for exploring pattern.

6. Who (implementers)

The Reception Jigsaw is delivered by expert trainers from WRM. Teachers in receipt of the training then deliver content to pupils over the course of the reception year.

Who are the trainers?

- For the trial, trainers were chosen selectively, both based on applications from potential trainers and through 'headhunting'. Some trainers were full-time employees of WRM. All trainers are qualified teachers and SLEs, and are required to have at least five years' of teaching experience as well as early years' experience. Some had been teaching for 25 years.

- Trainers were not necessarily mathematics specialists, but most tended to be. Trainers had a mix of expertise.

- The plan to deliver Reception Jigsaw in Essex involved operating a host school/franchise model, with locally-recruited WRM-trained trainers operating in that area. As mentioned above, recruitment challenges in Essex led to WRM trainers from Yorkshire delivering Reception Jigsaw in Essex (see 'Adaptations' section below).

How is training of trainers carried out?

- The core team at WRM runs the training. WRM make the people being trained ('trainers') aware of the high standards expected. The training of the trainers is delivered by one or more of the core WRM team. For this trial, the CPD sessions were all delivered by the WRM project lead. Additional training for the school support visits and developing quality early years provision and pedagogy was delivered by the WRM head of early years.

- The 'train the trainer' sessions follow the format of those successfully implemented on the Primary Jigsaw. Trainers have the opportunity to explore each session in depth and discuss the key pedagogy and research, which forms the basis of the training. They 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 twilight sessions in front of the WRM core team, and WRM provide feedback based on these sessions.

- In the trial, WRM started to recruit trainers during the 2019/2020 school year in line with the original delivery timetable. Trainers received some training prior to the start of delivery (in Summer Term 2020, and in Spring Term and Summer Term in 2021, depending on when they were recruited). The remaining training—about how to deliver each of the modules—was staggered through the 2021/2022 academic year to ensure that the delivery of the sessions was not too far removed from the training.

Quality assurance of the trainers' delivery

- Training sessions, delivered by every trainer, are observed by WRM twice a year. Based on their observations, WRM collate their feedback on different aspects of the training. The feedback is then analysed by the WRM team to identify and suggest specific areas of improvement for every trainer.
- As new trainers were recruited by WRM to deliver the Reception Jigsaw as part of the trial, WRM planned for all newly recruited trainers to deliver a 'practice' session before they delivered sessions in schools in the trial. WRM observed this practice session and provided feedback and guidance to the trainer as needed.
- After every session, feedback from attendees is collected by the trainer and analysed and used to inform both the course content/delivery across all trainers as well as the line management/development of the individual trainers.
- The trainers have fortnightly meetings with their line manager for supervision and (two-way) feedback.

How long does the training of the trainer last?

- There is one day of introductory training at the start, then for each module there is approximately one and a half days of training per module. The training is intensive and includes the research that the modules are based on. The trainers' delivery of the twilight sessions is observed (by the WRM team) and feedback is provided.
- Training is spaced out over the year and trainers are trained just before they are due to deliver the particular module.

7. How (mode of delivery)

The same trainer delivers all training to the same school. Each trainer goes to five to ten schools. The school has the trainer's email address so they can ask queries.

What is the balance between trainers following the content versus tailoring it?

Twilight sessions mostly follow the specified content, with coaching visits allowing more flexibility for tailoring. Trainers are free to add their own anecdotes (also see section on 'What (procedures, activities, or processes used)' above).

The participants use what they have learned in the training to change/inform their teaching and delivery of mathematics in the classroom. The trainers can provide support to do this, as part of the coaching, and if there are questions from the school/teachers then the trainers will support the school with additional requests.

As well as delivering the training and coaching, the trainers observe any changes that have occurred through the year. WRM are able to (lightly) monitor schools' participation in the project during each visit based on how engaged they are with the content as well as the extent to which the school has completed the gap task and reflection. During the trial, trainers completed detailed reports for each school visit, which were sent to the central WRM team. WRM also held regular team meetings and phone calls to discuss schools as needed.

8. Where (setting of the intervention)

The target schools were:

- primary schools in Yorkshire and surrounding areas (up to 75 miles from Halifax), and schools in Essex or those close to the Essex border in Outer London;
- schools that had not previously had more than two sessions of WRM Primary Jigsaw training or any sessions of Reception Jigsaw training;[2]
- schools who were not participating in any other early years trials by the EEF; and

[2] WRM plans to offer the Reception Jigsaw training sessions as either webinars or face-to-face training during the 2020/2021 academic year. The half-day school support visits will not be included. Any school accessing any of these sessions will be excluded from the trial.

- schools with stand-alone reception classes (i.e. not mixed reception/Year 1).

An adjustment was made to the eligibility criteria in January 2020 relating to the DFEs Early Years Professional Development Programme (EYPDP). Schools were able to participate in this trial *and* the EYPDP so long as the reception teachers in the school did not take part in the mathematics module of the scheme.

9. When and how much (dosage and duration)

As described above in section on 'What (procedures, activities, or processes used)', training consists of ten sessions—five twilight CPD sessions and five coaching sessions—held over an academic year. Under ideal conditions, training starts in July so that staff can use knowledge from the initial session to plan for the pupils starting their classes in September. In the trial this happened over a slightly shorter time frame than would normally happen in schools due to the data and testing requirements of the trial, from November 2021 to May 2022.

In terms of dosage for the pupils, while the Reception Jigsaw aims to feed into many aspects of the school day, 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)

The Reception Jigsaw training is delivered at each school alongside the coaching sessions. The Reception Jigsaw is delivered using a standard set of materials that WRM have developed, but the school-level delivery can be tailored to individual schools if there are particular areas they need support with.

The coaching (half-day) visits are highly tailored to each school (see above).

Teachers can tweak how they adapt their learning based on the level the children are at. The training package aims to develop teachers' knowledge of pedagogical approaches and developmental progressions for early maths so they can apply these in their own classrooms.

11. How well (planned)

Strategies to maximise adherence and fidelity

The half-day coaching sessions usually allow trainers to gauge engagement and an indication of whether schools are implementing the training. In the trial, trainers will feed this information back through school visit reports after each visit to the WRM project lead, who will have a follow-up phone call with the key contact at the school to discuss any issues where they arise. WRM project lead will keep a record of any issues and actions taken in response.

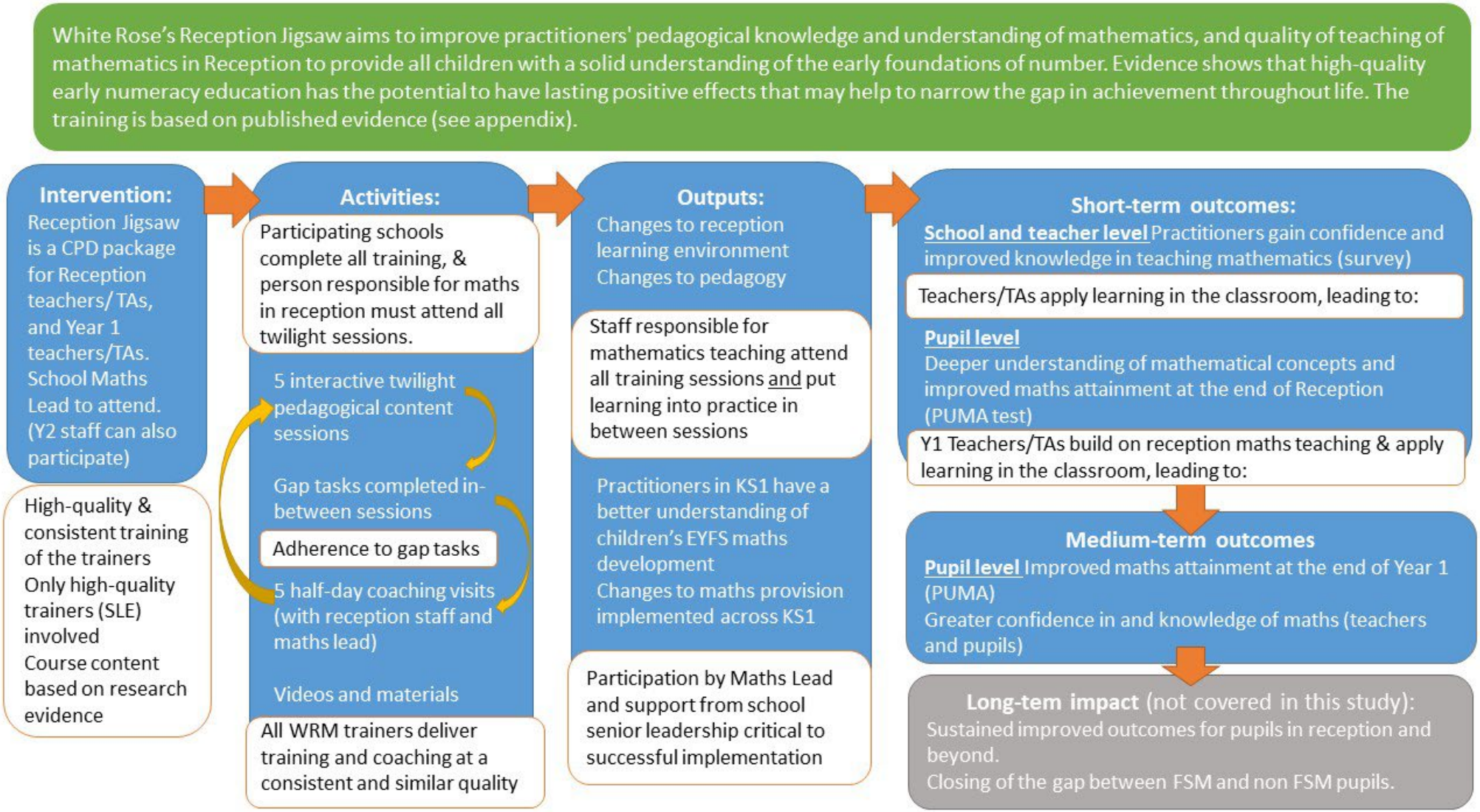
WRM collects informal records of where schools are in terms of progress and whether schools have improved over the year.

Logic model

Figure 1 shows the hypothesised logic model for the Reception Jigsaw. The hypothesised outputs, outcomes, causal mechanisms, and mediators are detailed. The impact evaluation explored whether there was an impact of accessing the Reception Jigsaw on practitioners' confidence in teaching mathematics, pupil attainment outcomes, and on practitioners' confidence in their own mathematics ability. Medium-term pupil-level outcomes will be assessed in Summer Term 2023 when the pupils in the trial will be in the last term of Year 1. The results of this analysis will be added to this report as an addendum in early 2024.

The outputs concerned with changes to the learning environment in reception and Year 1 (as well as changes to teachers' understanding of mathematic development) are captured in the implementation and process evaluation (IPE). Mediators around levels of compliance to the intervention (attendance at CPD sessions, coaching sessions, and completion of gap tasks) are also explored in the IPE, and the impacts of attendance compliance, specifically on pupils' mathematical outcomes, are estimated in the impact evaluation.

Figure 1: Hypothesised logic model (mediators shown in clear boxes)



Evaluation objectives

The evaluation research questions are given below. Further detail of the evaluation design can be found in the evaluation protocol (Poet *et al.*, 2021) and statistical analysis plan (SAP) (Andrade *et al.*, 2022).

Impact research question

Primary question

1. What is the impact of the WRM Reception Jigsaw intervention on reception children's mathematics attainment, as measured by New PUMA tests at the end of the reception year, compared to 'business as usual'?

Secondary questions

2. What is the impact of the WRM Reception Jigsaw intervention on pupils' mathematics attainment as measured by New PUMA tests at the end of Year 1, compared to 'business as usual'?
3. What is the impact of the WRM Reception Jigsaw intervention on pupils' score on the mathematics sections of the EYFSP, compared to 'business as usual'?
4. What is the impact of the WRM Reception Jigsaw on reception practitioners' and school mathematics leads' confidence to teach mathematics to children in their reception year, compared to 'business as usual'?
5. What is the impact of the WRM Reception Jigsaw on reception practitioners' and school mathematics leads' confidence in their maths ability, compared to 'business as usual'?[3]
6. Are effects on mathematics attainment in the reception year (as per research question 1) different for pupils eligible for free school meals (FSM)?

IPE research questions

1. Was the Reception Jigsaw delivered as intended in terms of dosage, nature, and quality?[4]
2. How well did the participants (teachers/support staff from reception and Key Stage 1, mathematics lead, senior leaders, and pupils) engage with the Jigsaw? Were there any implementation challenges faced? If so, what were they and to what extent were they overcome?
3. Was the quality of training, support, and intervention materials provided by the developer adequate? Was preparedness and confidence of staff delivering the intervention at the right level? If not, why?
4. To what extent do participants feel the intended outcomes of the programme are being achieved for children, practitioners, and the school? How are they being achieved?
5. What does business as usual consist of for the comparison group?

Ethics and trial registration

An ethical review was undertaken as part of the National Foundation for Educational Research (NFER) start-up meeting in October 2019 where consideration was given to consent and impact of the research on trial participants (practitioners and pupils). The evaluation was conducted in accordance with the NFER Code of Practice.

The trial was designed, conducted, and reported to Consolidated Standards Of Reporting Trials (CONSORT) standards and registered on the International Standard Randomised Controlled Trial Number (ISRCTN) registry (https://doi.org/10.1186/ISRCTN36328399).

Ethical agreement

Ethical agreement for participation within the trial was provided by the headteacher of the school when signing the memorandum of understanding (MoU). Ahead of any pupil data being shared, parents/carers were provided with a letter giving details about the trial. They were also provided with information about how they could withdraw their child from

[3] This research question was not explicitly stated in the protocol but was added when the SAP was drafted. Upon closer examination of the instrument used in Chen *et al.* (2014), it was felt that two of the three scales (measuring different constructs) were relevant in terms of outcomes in the logic model. One scale was 'confidence in teaching mathematics' and the other was 'confidence in one's own mathematics ability'.

[4] This relates to perceived quality, including aspects such as consistency of delivery against training materials, trainer responsiveness, and participant engagement levels.

testing and data processing if they had objections to this (see Appendix E). Practitioners completing surveys as part of the trial were also provided with full information about the research and how their responses and data were going to be used. Separate privacy notices were drawn up for pupils and school staff (see Appendix F) giving details of any personal data being collected as part of the trial and how they were being transferred, stored, used, and reported.

All data gathered during the trial were held in accordance with the data protection framework established by the Data Protection Act 2018 and the General Data Protection Regulation (GDPR) (European Union) 2016/679, and treated in the strictest confidence by the NFER, WRM, and the EEF. Our legal basis for gathering and using this data was legitimate interest through our work as a research organisation.

Data protection

Personal data was processed as part of the evaluation. Prior to the trial starting, a data sharing agreement between WRM, NFER, and the EEF was signed setting out the roles and responsibilities around data security and processing. This includes a description of the nature of the data being collected and how it will be shared, stored, protected, and reported by each party. NFER and WRM were the joint data controllers for the trial up until the data is passed to the EEF archive.

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 interest are overridden by the interests or fundamental rights and freedoms of the data subject which require protection of the personal data'.

We carried out a legitimate interest assessment, which demonstrated that the evaluation fulfilled one of NFERs core business purposes (undertaking research, evaluation, and information activities). It has broader societal benefits and will contribute to improving the lives of learners by providing evidence for schools when making decisions about mathematics CPD.

NFER and WRM provided a MoU to schools explaining the nature of the data being requested of schools and children, how it will be collected, and how it will be passed to and shared with NFER. An information sheet and withdrawal form were also distributed to parents who were due to start reception in September 2021 before any of their child's data was transferred to NFER (see Appendix E). The information sheet provided details about the project and any planned pupil data collection, as well as links to the privacy notice.

Transfer of personal data between WRM and NFER, and between schools and NFER, took place using NFERs secure online portal.

Pupil data was linked to the National Pupil Database (NPD) in order to collect pupil-level data on FSM eligibility and EYFSP outcomes. NFER provided the Data Sharing Team at the DfE with the pupil information, allowing a match to the NPD. After the matching process had taken place, NFER then analysed this data using the Secure Research Service (SRS) managed by the Office for National Statistics (ONS). NFER accessed the data for analysis remotely through the SRS secure online system. The SRS does not allow users to remove or copy data from its servers.

At the end of the EEF evaluations, all data is archived to allow for further secondary analysis. At this point, the EEF becomes the data controller and NFER is no longer responsible for the data and are no longer a data controller. After three months of the completion of the study, all of the matched data (i.e. to NPD) will be added to the EEF archive and 'de-identified' before being made available to researchers. The EEF archive is hosted by the ONS and managed by the Fischer Family Foundation (FFT) Education. Other research teams may use the de-identified data as part of subsequent research through the Approved Researcher Scheme. NFER will delete all data collected for the trial one year after the publication of this report.

For further information, privacy notices for the study are available in Appendix F.

Project team

Details of the project team and their roles and responsibilities, in terms of the evaluation, are given in Table 4.

Table 4: Project team roles and responsibilities

Name	Institute	Roles and responsibilities
Ben Styles	National Foundation for Educational Research (NFER)	Project director, responsible for leading the NFER team and project delivery
Helen Poet / Sarah Tang / Connie Rennie	NFER	Project manager, responsible for overseeing the day-to-day running of the trial and the implementation process evaluation (IPE)
David Sims	NFER	Process evaluation director, responsible for overseeing the development of IPE tools
Kathryn Hurd	NFER	Test and schools administration lead, responsible for overseeing recruitment, school contact, and testing
Guido Miani / Alison Hale / Asma Ullah / Guvi Chohan / Lydia Wallis	NFER	Operations researchers, responsible for coordinating school recruitment, school contact, and testing
Joana Andrade / Chris Morton / Gemma Schwendel	NFER	Statistician, responsible for statistical analysis
Kerry Martin	NFER	Researcher, responsible for leading IPE
Eleanor Bradley	NFER	Researcher, responsible for supporting IPE
Jane Brown	White Rose Maths (WRM)	Mathematics lead specialist, responsible for developing the intervention content, training the trainers, and leading the delivery team
Tony Staneff	WRM	Founder of WRM, responsible for overseeing the intervention
Caroline Hamilton	WRM	Director of WRM, responsible for overseeing the intervention
Kat Ellis	WRM	Head of early years foundation stage, responsible for leading the delivery team and school-to-school support training
Jane Rodda (name change from Milner)	WRM	The Education Endowment Foundation project coordinator, responsible for administration and school liaison

Methods

Trial design

Table 5: Trial design

Trial design, including number of arms		Two-arm, cluster randomised
Unit of randomisation		School
Stratification variable(s) **(if applicable)**		Geographic area (Yorkshire and Essex) Taking part in Mastering Number programme (Yes/No) (combined to give four strata: Essex_Yes, Essex_No, Yorkshire_Yes, Yorkshire_No)
Primary outcome	**Variable**	Mathematics attainment
	Measure **(instrument, scale, source)**	End of reception mathematics score, test total score, 0–30, New PUMA
Secondary outcome (1)	**Variable(s)**	Early years foundation stage profile (EYFSP)
	Measure(s) **(instrument, scale, source)**	EYFSP, taking the value one if a pupil meets or exceeds the two mathematics goals, and zero otherwise, National Pupil Database
Secondary outcomes (2, 3)	**Variable(s)**	• Practitioners' confidence to teach mathematics to reception pupils • Practitioners' confidence in own mathematics ability
	Measure(s) **(instrument, scale, source)**	• Practitioners' confidence to teach mathematics to reception pupils scale, 11–55, bespoke survey, adapted from the Early Math Beliefs and Confidence Survey (Chen *et al.,* 2014) • Practitioners' confidence in own mathematics ability scale, 9–45, bespoke survey, adapted from the Early Math Beliefs and Confidence Survey (Chen *et al*., 2014)
Secondary outcomes (4)	**Variable(s)**	Mathematics attainment
	Measure(s) **(instrument, scale, source)**	End of reception mathematics score, reduced test total score, 0–20, New PUMA

Secondary outcomes (5 – longitudinal outcome)	**Variable(s)**	Mathematics attainment
	Measure(s) **(instrument, scale, source)**	Year 1 mathematics score, test total score, 0–30, New PUMA – Summer 1
Baseline for primary outcome and secondary outcome (1, 4)	**Variable**	Mathematics ability
	Measure **(instrument, scale, source)**	Teacher assessment based on observation (Emerging Numeracy checklist), 10–60*, bespoke instrument
Baseline for secondary outcomes (2, 3)	**Variable**	• Practitioners' confidence to teach mathematics to reception pupils • Practitioners' confidence in own mathematics ability
	Measure **(instrument, scale, source)**	• Practitioners' confidence to teach mathematics to reception pupils scale, 11–55, bespoke survey, adapted from the Early Math Beliefs and Confidence Survey (Chen *et al.,* 2014) • Practitioners' confidence in own mathematics ability scale, 9–45, bespoke survey, adapted from the Early Math Beliefs and Confidence Survey (Chen *et al.,* 2014)

** The protocol states the scale for the Emerging Numeracy checklist as 0–20. The checklist is a 20-item list but as the instrument development progressed it was felt that a three-point scale for each answer was more appropriate than a binary (Yes/No) one.*

The trial was an efficacy trial using a two-arm cluster randomised controlled trial. A cluster design was the most appropriate as the Reception Jigsaw is a CPD programme aimed at reception practitioners in a school. The intervention was the Reception Jigsaw, which is a CPD programme aimed at reception staff developed and delivered by WRM. The control condition was 'business as usual', although it is important to note that the majority of randomised schools used WRM schemes of work in reception across both arms of the trial as part of their usual practice prior to the trial (91% of 54 intervention schools, 84% of 80 control schools).

In summer 2021, the Government launched a programme 'Mastering Number', which aims 'to secure firm foundations in the development of good number sense for all children from reception through to Year 1 and Year 2'.[5] In October 2021, during pupil-baseline data collection, trial schools were asked by email whether they had signed up to take part in the Mastering Number programme in the current academic year (2021/2022). This information was used as a stratifying variable in randomisation to ensure that there was balance across intervention and control arms in terms of schools taking part in the Mastering Number programme (see 'Randomisation' section below).

The evaluation measured the impact of Reception Jigsaw on pupils' mathematics attainment at the end of their reception year in school. Mathematics attainment was measured using the New PUMA (primary outcome) and the EYFSP (secondary outcome 1). An additional secondary attainment outcome (4) was included in the SAP that was not included in the protocol. On closer examination of the New PUMA by WRM, it was felt that some of the items included in the assessment were not in line with the revised early years curriculum. The project team (WRM, NFER, and the EEF) decided to include a secondary outcome that included a reduced number of items from the New PUMA (see 'Outcome measures' section below). Other secondary outcomes (2, 3) considered: the impact on reception practitioners (teachers

[5] https://www.ncetm.org.uk/maths-hubs-projects/mastering-number/

and TAs) and school mathematics leads; measuring the impact the programme had on their confidence to teach mathematics to reception pupils; and confidence in their own mathematics ability.

In autumn 2022, the decision was made by EEF to proceed with the longitudinal follow up as per EEF guidance (EEF, 2019) and the trial protocol (Poet *et al.*, 2021). In March 2023, all schools were asked to confirm their pupil data (such as name, date of birth and whether they had withdrawn from the trial). We provided schools with a pupil data form prepopulated with the details of the pupils who took part in the trial previously. Schools were asked to confirm whether these pupils were still able to take part or if they had since left the school or been withdrawn from the assessment. This enabled the research team to track the pupils. Schools had a similar opportunity to mark a pupil as having left the school or withdrawn when the registers are sent out with the assessments. All schools taking part in the trial were asked to complete a second New PUMA Summer 1 (for Year 1 pupils) with the cohort of pupils who had taken the original impact assessment at the end of their reception year in summer 2022. This second (longitudinal) assessment took place in summer 2023 at the end of Year 1.

WRM recruited schools from Yorkshire and Essex, however fewer schools were recruited from Essex than originally planned due to challenges in recruiting trainers and schools in Essex (see 'Randomisation' section below). Financial incentives were paid to the 135 schools that completed the pupil-level endpoint assessment. Intervention schools (n=55) received £250 and control schools (n=80) received £750 in Autumn Term 2022. Further financial incentives were paid to those schools that completed the longitudinal follow-up assessments to the sum of £200 per school.

Participant selection

School eligibility

Eligible schools were any state schools in Yorkshire or Essex with at least one reception class that had not taken part in any Reception Jigsaw training in the past. More specifically, the eligibility criteria for schools were:

- all state primary schools situated within a 75-mile radius of Halifax, and schools in Essex or those close to the Essex border in Outer London;
- schools that have not previously had more than two sessions of WRM Primary Jigsaw training or any sessions of Reception Jigsaw training;[6]
- schools who are not participating in any other early years trials by the EEF in 2021/2022;
- schools with stand-alone reception classes (i.e. not mixed reception//Year 1); and
- schools may participate in this trial *and* the DfE EYPDP so long as the reception teachers in the school do not take part in the mathematics module of the scheme.

These criteria were explained to the schools by WRM during recruitment. The number of schools in the intervention group was 55 as this was the number of schools that WRM had the capacity to deliver to during the intervention period. Of the 55 intervention schools, 35 were planned for the North-East and 20 for Essex, as WRM are based in Yorkshire. The aim was to recruit 85 schools from Yorkshire and surrounding areas (35 intervention, 50 control) and 50 from Essex and surrounding areas (20 intervention, 30 control), although fewer schools were recruited in Essex than planned. This was at least due in part to not being able to hold as many face-to-face recruitment events in Essex due to restrictions on travel related to Covid-19.

Within the school, all reception teachers, Year 1 teachers, and school mathematics lead are expected to take part in the intervention. Other school staff such as reception TAs and Year 1 TAs are encouraged to attend the training whereas Year 2 classroom staff can attend but are not the intended focus of the Reception Jigsaw (see 'Intervention' section). The intervention is a CPD intervention delivered to the school mathematics lead and all reception and Year 1 teachers, who then use the teaching approaches with all their classes.

[6] WRM offered the Reception Jigsaw training sessions as either webinars or face-to-face training during the 2020/2021 academic year (although the half-day school coaching visits were not included). Any school accessing any of these sessions were excluded from the trial.

School-level pupil sampling

For the trial, in order to reduce burden on teachers and costs, a sample of 20 pupils per school was used for the pupil-level assessments (at baseline and endpoint) as NFER test administrators can accomplish 20 tests in a day's visit when using small-group assessments. If schools had fewer than 20 pupils all the pupils would be in the sample. Initial plans involved NFER analysts using the pupil data provided by the schools in summer/early autumn 2021 to randomly select 20 pupils per school, who would then go on to be the trial pupils. It was expected that all pupils in the reception classes of intervention schools would receive the intervention (via their teachers), but the assessments would just take place on the 20 sampled pupils.

During the design stage of the trial, the evaluation team assumed that 25% of pupils participating in the trial would be eligible for FSM,[7] which led to a rough estimation of a cluster size of five FSM pupils per school. The minimum detectable effect size (MDES) for the FSM subgroup was calculated using this cluster size estimate (see Table 10).

However, in a subsequent analysis based on the historical figures for proportion of FSM pupils in the schools that signed up for the trial, the trial's statistician concluded that the average proportion of FSM-eligible pupils was overestimated in the original design. Therefore, a random sample of 20 pupils per school would not achieve an average FSM cluster size of five FSM pupils. In fact, in order to ensure a cluster size of five FSM pupils per school (and the corresponding MDES described in the protocol), it would have been necessary to randomly select 28 rather than 20 pupils per school.

Based on the updated calculation, the evaluation team adopted the following sampling scheme:

- We randomly selected an initial sample of 28 pupils per school, the randomisation being stratified to reflect the reported FSM proportion of the individual schools (the proportion of reported FSM was calculated using the Yes/No responses for each individual school—ignoring missing data). Of these 28 pupils, 20 constituted the pupils to be included in the measurement for the trial and eight pupils were standbys if replacement was necessary.

- In each school, the pupils explicitly identified as eligible for FSM in the initial sample were automatically included in the sample of 20 pupils selected for the trial. The remaining pupils in the initial sample were then randomly allocated to the trial sample or assigned to a group of replacements of up to eight children per school.

- The pupils in the replacement groups were treated as standbys, to be included in the trial sample in case the pupils in the trial samples of their schools could not take part in the trial (e.g. if they did not start at the school) or were withdrawn from it (i.e. by their teacher or parent/guardian). This replacement process could only happen prior to randomisation.

The R code used to draw the pupils is included in Appendix G. All the calculations were performed in R 4.0.3.

Recruitment

This trial originally began in July 2019. Recruitment started in November 2019 with a view to run the intervention in schools in Autumn Term 2020. As WRM had existing links with a large number of schools, they led the recruitment of schools to the trial. Recruitment was paused in March 2020 as schools were closed to the majority of pupils in response to the Covid-19 pandemic. As of March 2020, WRM had contacted 532 schools of which 98 had signed MoUs. The project team (WRM, NFER, and the EEF) decided to delay the project by a year due to disruption and pressures in schools relating to the pandemic. Recruitment reopened in November 2020. WRM approached schools that had already signed an MoU to ascertain whether they would like to take part in the rescheduled trial. They also re-approached schools that had been approached during the first recruitment period (pre-Covid-19) but had not previously signed up as well as approaching new schools.

By May 2021, WRM had recruited 160 schools (145 recruited to the trial and 15 constituted a waitlist), having approached a total of 1,004 schools. Of the 145 schools that signed up to take part in the trial, 82 had previously signed MoUs to take part in the original trial. Between May 2021 and randomisation in October 2021, 19 schools either withdrew

[7] The figure of 25.0% of pupils eligible for FSM is still considerably higher than the national proportion that stands at 20.8% as of June 2021, having gone up from 17.3% in 2020 (https://explore-education-statistics.service.gov.uk/find-statistics/school-pupils-and-their-characteristics).

or dropped out of the trial, so the 15 schools on the waitlist were invited to take part in the trial and agreed to do so. A further three recruited schools were identified as being independent schools and thus ineligible to take part in the trial. A total of 138 schools were randomised. While the sample size calculations at design stage suggested an achieved sample of 129 schools was required (see 'Sample size' section below), it was decided that it would be desirable to increase this number slightly given the uncertainty around working with schools during the Covid-19 pandemic and recovery period, and their engagement in research during such a time.

Outcome measures

Baseline measures

In the first half of Autumn Term 2021, the first statutory national administration of the NFERs new Reception Baseline Assessment (RBA) took place. The RBA data is not available to researchers via the NPD (NFER, 2019) and therefore it was not possible to use this as the baseline measure. As an alternative, NFER researchers created an Emerging Numeracy (EN) checklist with expert input from WRM early years specialists, which was used as a baseline measure for the pupil-level outcomes (primary and secondary 1 and 4).

The EN checklist was developed in consultation with colleagues in the Centre for Assessment at NFER and early years specialists at WRM and comprised 20 tasks based on the Early Years Outcomes and the ELGs (DfE, 2021a; DfE, 2021b). The EN checklist is available in Appendix H. Teachers were asked to complete the EN checklist after they had completed the RBA.[8] For each of the 20 tasks, teachers were asked to signal whether each pupil, from what they had observed, could not do the task, could do it with support, or could do it independently. Responses were coded as 1, 2, 3 thus creating a scale that ranged from 20 to 60. To ensure that the EN checklist was adequately reliable, the Cronbach's alpha index was calculated for the group of 20 variables, and was found to be 0.95. A Cronbach's alpha greater than 0.7 is generally considered adequately reliable (Nunnally and Bernstein, 1994).

The baseline measures for the practitioner-level outcomes, namely 'practitioners' confidence to teach mathematics to reception pupils' and 'practitioners' confidence in own mathematics ability' were the same instrument as the outcome measure.

Primary outcome

As a follow-up test we used the reception version (summer) of the New PUMA**,** a standardised test developed by Rising Stars (part of Hodder Education). At the time of the initial trial design, the PUMA was considered the most appropriate assessment choice as it had been most recently standardised to a UK population, it could be administered in small groups (as opposed to one to one) and it was aligned to the early years curriculum. In 2020, after the evaluation was designed, an edition of the 2014 PUMA was released (i.e. New PUMA). There was no change in the content between the New PUMA and the previous edition at reception level, despite curriculum changes in this area; so, the alignment to the curriculum was considered not as strong. In late 2021 and early 2022, other assessment options were considered in depth but a more suitable option was not found. An additional outcome measure (reduced New PUMA) was added in response to the lack of curriculum alignment (see 'Outcomes and analysis' section below). The range for the New PUMA (total score) was 0–30, with a lower score indicating lower achievement.

Test administrators were used to administer the assessments in schools. The test administrators received additional training due to the very young age of the pupils being assessed. The recommended time for the assessment was 40 minutes with a maximum of five pupils. Tests (and accompanying documents such as attendance registers) were despatched from NFER directly to the test administrators via secure courier. Completed test scripts were returned by the test administrators also using secure couriers. Scripts were marked by NFER staff and results returned to schools via a secure portal. Markers were blinded to group allocation although the analyst was not.

Secondary outcomes

1. EYFSP

As a follow-up measure for the first secondary analysis, an aggregation of the two mathematics ELGs in the EYFSP was used, using data from the NPD. This outcome data was accessed and analysed using the SRS (see 'Data protection' section above). The EYFSP was included as a secondary outcome as it provides another measure of

[8] Not immediately afterwards, due to requirements of the administration of the RBA.

mathematical attainment (the hypothesised short-term pupil-level outcome) with no additional burden on schools as it is a statutory assessment. The EYFSP is completed by the pupil's reception teacher during the Summer Term of the reception year. Teachers are asked to submit an assessment against 17 ELGs (two of which relate to mathematics) to state whether each pupil is 'meeting the level of development expected at the end of the EYFSP (expected)' or 'not yet reaching this level (emerging)' (DfE, 2022b, p.8). Teachers were not blinded to group allocation, but the EYFSP is a statutory assessment, so this was considered unlikely to be a source of bias.

The latest version of the EYFSP has only two mathematics goals, G11 (Number) and G12 (Numerical Patterns); a pupil is recorded as either 'Emerging' or 'Expected' in each of these. We computed EYFSP as a binary measure, taking the value one if a pupil met the expected level for both mathematics goals, and zero otherwise.

2. Confidence in teaching mathematics / 3. Confidence in own mathematics ability

The other secondary outcomes are the impact of the intervention on practitioners' confidence to teach mathematics in the reception year and practitioners' confidence in their own mathematics ability. Increases in practitioners' confidence in teaching mathematics is the initial hypothesised short-term outcome described in the logic model so this outcome was included as a secondary outcome. A practitioner-level outcome measure is important as the Reception Jigsaw is a CPD programme. Changes must first occur in the practitioners' practice before any changes can be expected in the pupil outcomes. We used a pre- and post-survey of teachers, TAs, and the school mathematics lead. The survey was based on the survey developed by Chen *et al.* (2014) and used recently in the Maths Champions trial (Robinson-Smith *et al*., 2018), with minor adaptations to reflect the setting (mainly related to terminology, i.e. to say reception/school rather than nursery). The third scale in Chen *et al.* (2014) measured teacher beliefs about preschoolers and early maths and was not included as it was not aligned to the logic model. The survey was administered at the same time as the pre-and post-primary outcome activities. The scale for practitioners' confidence in teaching mathematics ranged from 11–55, with a lower score representing lower confidence. The scale for practitioners' confidence in their own mathematics ability ranged from 9–45, with again a lower score representing lower confidence.

4. New PUMA—reduced scale

As for the primary outcome, this secondary outcome used pupils' scores from the New PUMA, but in this case only 20 of the 30 items contributed to the total score. The scale was 0–20. It was decided after consultation with WRM that ten items did not align with the current reception mathematics curriculum, upon which the Reception Jigsaw training is based. These ten items, which were related to totalling money or using teen numbers, were not included in the secondary measure. As this is a bespoke scale that has not been validated in previous studies, its inclusion in the analysis was conditional on it demonstrating a Cronbach's alpha of 0.7 or above as this is the accepted threshold for a reliable scale (Nunnally and Bernstein, 1994). The Cronbach's alpha for this reduced New PUMA scale was 0.8.

5. New PUMA (Year 1) – longitudinal analysis

The New PUMA for Year 1 (Summer 1) pupils was used for the longitudinal follow-up measure. The range for the New PUMA (total score) was 0–30, with a lower score indicating lower achievement.

As with the original follow-up assessment, test administrators were employed to administer the assessments in schools. Where possible the same test administrators were used in the same schools as for the original follow-up. This was to provide consistency for the schools and to build on the positive feedback received from schools after the main follow-up. Tests (and accompanying documents such as attendance registers) were despatched from NFER directly to the test administrators via secure courier. Test administrators were blinded as to whether schools were in the intervention or the control group. However, teachers may have disclosed this information to the test administrators during the assessment. A time limit of one hour was recommended for the delivery of the assessment with each test administrator working with a maximum of five pupils at any one time. Completed test scripts were returned by the test administrators using secure couriers. Scripts were marked by NFER staff and results returned to schools via a secure online portal. Markers were blinded to group allocation although the analyst was not.

Sample size

In order to determine the number of schools needed to adequately power the trial, estimates of parameters were used to complete sample size calculations using a bespoke Excel spreadsheet. For the sample size calculations at design stage, analysts based these estimates on parameters from the EEF Maths Champions evaluation (Robinson-Smith *et*

al., 2018), which used the Centre for Evaluation and Monitoring's ASPECTS assessment. This evaluation had an intracluster correlation coefficient (ICC) of 0.17 and a pre-/post-test correlation of 0.59 at analysis stage. It was anticipated that using the EN checklist at baseline (as opposed to a more formally developed assessment) would result in a lower correlation; so, at design stage, analysts assumed a correlation of 0.4 for the estimated pre-/post-test correlation.

An estimate of cluster size was also needed. In order to reduce costs and to minimise the burden on teachers (who were required to complete the pupil-level baseline assessment), only 20 pupils per school were included in the sample (see 'Participant selection' section). A sample of 20 pupils per school was chosen since this is the number of small-group tests that an NFER test administrator can accomplish in a day's visit. Using these assumptions and after discussions with the developer concerning their capacity to deliver to more than 50 schools, a sample size of 55 intervention schools and 74 control schools was agreed (total n=129). This design has 80% power to detect an effect size of 0.22. These are achieved sample sizes, namely, schools with both baseline and follow-up data.

As shown in Figure 2, 138 schools were randomised as higher levels of drop out than usual were anticipated due to Covid-related pressures on schools. At randomisation, the MDES was calculated to be 0.21 due to the slightly higher sample size. Three schools did not submit follow-up data (see Figure 2) but the MDES remained 0.21 at the analysis stage. The pre-/post-test correlation was higher than predicted at 0.59, suggesting that the EN checklist was a reasonable predictor of later mathematics outcomes. The ICC was also higher at analysis stage (0.24) than predicted at the outset (see Table 10) suggesting schools in the trial are more segregated in terms of numeracy attainment than anticipated. Although the ICC and the pre-/post-test correlation are different from what was predicted at design stage, the direction of the variance means that their effects balance out and the desired MDES was achieved.

At design stage it was assumed that the sampling frame of schools would be likely to include schools with a higher-than-average proportion of pupils eligible for FSM, as pupils with a higher level of disadvantage (signalled by eligibility for FSM) are of particular interest to the EEF. The initial sample size calculations assumed 25% of pupils in the recruited schools would be eligible for FSM, which is considerably higher than the national average (20.8% in June 2021, National Statistics, 2022).

The sample of schools recruited had a lower average proportion of pupils eligible for FSM[9] than anticipated, so a sampling approach (see 'Participant selection' section) was used in an attempt to boost the proportion of FSM pupils in the 20 selected per school. This approach was successful, as at randomisation there was an average cluster size of five FSM pupils per school/cluster, as can be seen from Table 10. Prior to analysis, the FSM variable used in sample size calculations was a school-reported binary FSM indicator, which was collected directly from schools in the Summer Term and early Autumn Term 2021. At analysis stage, the FSM variable was drawn from the NPD and was the EVERFSM6 variable.[10] This change in variable led to an FSM average cluster size of 5.1 in the analysis stage, and the target MDES for the FSM subgroup analysis (0.27) was achieved despite a slightly reduced sample of 125 schools. In 13[11] of the 138 randomised schools there were no FSM pupils in the sampled group of 20 pupils (as measured by EVERFSM6 in the NPD), and these schools were therefore not included in the subgroup analysis.

Randomisation

Under the original design, geographical area was the sole stratifier in the randomisation in order to ensure that WRM was able to deliver the intervention across the two areas. However, during the recruitment stage of the trial the team was made aware that the Maths Hubs network[12] was recruiting schools for the Mastering Number[13] programme, an initiative launched by the DfE that aims to provide training to teachers in order to promote 'firm foundations in the development of good number sense for all children from reception through to Year 1 and Year 2' (NCETM, 2021).

After consulting with the participating schools, the evaluation team concluded that a number of schools that were recruited for the Reception Jigsaw trial had also signed up for Mastering Number. Given the similarities between the two programmes in terms of objectives and methodology, the evaluation team felt it was advisable to ensure that the participation on Mastering Number is balanced across intervention and control schools. It was necessary to control for

[9] As measured by historical published school-level FSM data.
[10] EVERFSM6 variable flags whether a pupil has been in receipt of FSM at any point in the last six years. As these pupils are in their first year of schooling, we would expect the EVERFSM6 and FSM variable for the reception year to be identical.
[11] Two of the 13 schools had missing FSM data for some of the sampled pupils.
[12] See https://www.ncetm.org.uk/maths-hubs/.
[13] See https://www.ncetm.org.uk/maths-hubs-projects/mastering-number/.

possible instances of contamination (teachers in control schools participating in Mastering Number and receiving similar training to teachers in intervention schools) and also to account for confounding or even synergies between the two interventions (the average effects detected in the trial may be the result of the Mastering Number interventions implemented in intervention schools rather than those of Reception Jigsaw).

Bearing in mind the circumstances described above, the evaluation team decided to update the stratified randomisation design described in the protocol to a stratified randomisation that comprised of four strata: Yorkshire schools that signed up for the Mastering Number programme; Yorkshire schools that did not sign up for the Mastering Number programme (or unknown); Essex schools that signed up for the Mastering Number programme; and Essex schools that did not sign up for the Mastering Number programme (or unknown). An email was sent to all recruited schools in Autumn Term 2021 (prior to randomisation) in order to ascertain whether they had signed up to take part in the Mastering Number programme. This data was then used to create the stratification blocks in the randomisation. It should be noted that schools could sign up for Mastering Number after randomisation (or after we asked schools but before randomisation).

As per protocol, a total of 55 schools were randomised as intervention schools, with the remaining schools being assigned, by default, to the control group. Randomisation took place after baseline assessments had been completed, using the school Masterfile, which listed information about each school, and R to perform the stratified randomisation (see 'Randomisation code' in Appendix J). The number of intervention schools to be randomly selected out of each stratum was determined so that the representativeness of each block in the samples of intervention and control schools mirrors that of the overall sample of schools participating in the trial. As mentioned in the 'Trial design' section above, the number of schools recruited in Essex was lower than planned and this was therefore reflected in the numbers in Table 5. Only nine schools in Essex received Reception Jigsaw as part in the trial compared to the 20 schools originally planned.

Table 6: Stratified randomisation

Type of school	Yorkshire + Signed up for Mastering Number	Yorkshire + Did not sign up for Mastering Number / unknown	Essex + Signed up for Mastering Number	Essex + Did not sign up for Mastering Number / unknown	Total
Participating, number of schools (%)	39 (28.3%)	75 (54.3%)	8 (5.8%)	16 (11.6%)	138 (100%)
Intervention, number of schools (%)	16 (29.1%)	30 (54.5%)	3 (5.5%)	6 (10.9%)	55 (100%)
Control, number of schools (%)	23 (27.7%)	45 (54.2%)	5 (6.0%)	10 (12.0%)	83 (100%)

The R code used to perform the stratified randomisation is included in Appendix J. All the calculations were performed in R 4.0.3.

Statistical analysis

The primary and secondary analyses followed the EEF 2018 guidelines (EEF, 2018) and assumed intention-to-treat (ITT).

Primary analysis

A multilevel random intercepts model with two levels (school and pupil) was used to account for cluster randomisation. The primary analysis investigated whether a school's participation in the Reception Jigsaw programme had an effect on their pupils' mathematics attainment by the end of their reception year. This was determined by fitting a model with mathematics attainment at follow-up, as measured by New PUMA scores, as the dependent variable.

To control for prior ability, pupil-level EN scores assessed at baseline were included in the model as a covariate. The model also contained a dummy variable for stratum (school region vs school participation in the Mastering Number programme) to reflect the stratified randomisation.

The two-level random intercepts model is given by:

$$\mathrm{PUMA}_{ij} = \beta_0 + u_{0j} + \beta_1 \mathrm{intervention}_j + \beta_2 \mathrm{baseline\ EN}_{ij} + \beta_3 \mathrm{stratum}_j + \epsilon_{ij}$$

Where PUMA_{ij} is the New PUMA score of pupil i in school j, $\beta_0 + u_{0j}$ is the random intercept in school j, $\mathrm{intervention}_j$ is the school-level intervention/control dummy variable, $\mathrm{baseline\ EN}_{ij}$ is the baseline EN score of pupil i in school j, and $\mathrm{stratum}_j$ is a dummy variable for the randomisation stratum of school j.

The model was run in R (version 4.1.2) using the package 'nlme' (Pinheiro *et al*., 2021).

Secondary outcome analysis 1

The first secondary outcome analysis evaluated whether a school being assigned to participate in the Reception Jigsaw had an effect on pupils meeting the two ELGs relating to mathematics. For this purpose, we fitted a multilevel logistic regression model whose dependent variable was the binary EYFSP measure described in the subsection 'Secondary outcomes, 1. EYFSP' under 'Outcomes measure' section above.

The two-level random intercepts regression model/logistic regression is given by:

$$\mathrm{EYFSP}_{ij} = \beta_0 + u_{0j} + \beta_1 \mathrm{intervention}_j + \beta_2 \mathrm{baseline\ EN}_{ij} + \beta_3 \mathrm{stratum}_j$$

Where the dependent variable EYFSP_{ij} is the binary indicator (one if meeting expected level in both goals, zero otherwise) for pupil i in school j, $\beta_0 + u_{0j}$ is the random intercept in school j and $\mathrm{intervention}_j$ is the school-level intervention/control dummy variable. The baseline EN score of pupil i in school j is $\mathrm{baseline\ EN}_{ij}$, and $\mathrm{stratum}_j$ is a dummy variable for the randomisation stratum of school j.

The analysis was run in R (version 4.1.2) using the package 'lme4' (Bates *et al*., 2015) for a multilevel logistic regression model.

Additional secondary analysis

After reviewing the results of the main impact evaluation, WRM felt that it would be useful to consider the two EYFSP mathematics goals (Number and Numerical Patterns) separately as they felt that participation in Reception Jigsaw was more likely to have an impact on the Number outcome than the Numerical Patterns outcome. The model remained the same as the one used for the main EYFSP analysis other than the dependent variable was binary for achieving the expected level for a single mathematical goal rather than across both. Separate regressions were run for each of the goals – Number and Numerical Patterns. As this analysis was not prespecified in the statistical analysis plan it is considered exploratory.

Secondary outcome analysis 2 and 3

The second and third secondary outcomes analyses assessed if teachers' confidence to teach mathematics to reception pupils and confidence in their own mathematics ability were affected by their school's assignment to participate in the Reception Jigsaw programme. For this effect, models of practitioners' confidence in teaching mathematics and confidence in their mathematics ability at follow-up were fitted. Both aspects of practitioners' confidence were measured at baseline and follow-up via the adapted bespoke survey described in the subsection 'Secondary outcomes' under 'Outcome measures' section above. In accordance with the EEF 2018 guidelines, (EEF, 2018) the analysis models included a baseline measure of practitioners' confidence and, taking into account the clustered randomisation design, was run as a two-level (teacher and school) random intercepts models.

The two-level random intercepts model for the secondary outcome analysis 2 is given by:

$$\mathrm{PCTM}_{ij} = \beta_0 + u_{0j} + \beta_1 \mathrm{intervention}_j + \beta_2 \mathrm{baseline\ PCTM}_{ij} + \beta_3 \mathrm{stratum}_j + \epsilon_{ij}$$

Where $\mathrm{PCTM_{ij}}$ and $\mathrm{baseline\ PCTM_{ij}}$ are, respectively, the practitioners' confidence in teaching mathematics scores derived from the survey taken by teacher i in school j at baseline and follow-up.

And the two-level random intercepts model for the secondary outcome analysis 3 by:

$$\mathrm{PCMA_{ij}} = \beta_0 + \mathrm{u_{0j}} + \beta_1 \mathrm{intervention_j} + \beta_2 \mathrm{baseline\ PCMA_{ij}} + \beta_3 \mathrm{stratum_j} + \epsilon_{ij}$$

Where $\mathrm{PCMA_{ij}}$ and $\mathrm{baseline\ PCMA_{ij}}$ are, respectively, the scores derived from the practitioners' confidence in teaching mathematics survey taken by teacher i in school j at baseline and follow-up. In both models, $\beta_0 + \mathrm{u_{0j}}$ is the random intercept in school j, $\mathrm{intervention_j}$ is the school-level intervention/control dummy variable, and $\mathrm{stratum_j}$ is a dummy variable for the randomisation stratum of school j.

The model was run in R (version 4.1.2) using the package 'nlme' (Pinheiro *et al*., 2021).

Secondary outcome analysis 4

The model used for the secondary outcome analysis 4 was identical to the primary analysis model in all respects, except that only 20 selected New PUMA items were summed to calculate the dependent variable, rather than all 30 of them.

Analysis in the presence of non-compliance

WRM collected attendance registers at all training sessions, so it was possible to link pupils to their reception teachers' attendance data. Thus, compliance for the intervention was defined at pupil level, in terms of the number of training sessions (twilight sessions and coaching visits) completed by a pupil's reception teacher.

The evaluation team had initially conceptualised the compliance measures as a school-level measure based on the number of training sessions completed by each school. However, the team later revised this, as it was felt that school-level measures would not adequately reflect the experience of teachers and pupils, as well as the implementation of the intervention. While individual teacher attendance at the twilight sessions should be high, the nature of the coaching sessions is such that we would not expect all individuals to attend all coaching sessions. Furthermore, it was considered likely that the vast majority of the intervention schools would complete the full set of ten training sessions and therefore school-level measures would overlap intervention/control group allocation to a high degree.

As outlined in the SAP, it was considered that a pupil had been taught by a reception teacher who complied with the intervention if the teacher participated in nine or more sessions out of the ten available (five twilight in-depth training sessions and five half-day coaching sessions). This binary measure formed an optimal compliance indicator in the complier average causal effect (CACE) analysis. In addition, a separate CACE analysis using a pseudo-continuous dosage measure of compliance (0–10 sessions completed by the teacher) was run. There is a small group of children in the trial's sample that were taught by more than one teacher at reception. As only one reception teacher in a two-form entry school was required to receive coaching on each visit, the binary compliance measure would then be disproportionately low for larger schools. Therefore, we did not define compliance measures for these children or include their data in the CACE analyses.

To evaluate if there is an association between teachers having completed the intervention and mathematics attainment of their pupils, we adopted the instrumental variables (IVs) methodology prescribed by the EEF 2018 guidelines: we ran an IV regression (ivreg) by two-stage least squares model, with group allocation as the IV (EEF, 2018).

The same approach was used to investigate the presence of an association between dosage (number of training sessions attended by a teacher) and pupil mathematics attainment. Both models were fitted using the function ivreg from the R package 'ivreg' (Fox *et al.*, 2021) and the estimation of causal effects was completed resorting to the functions contained on the 'ivpack' package (Jiang and Small, 2014).

The analyses were run in R (version 4.1.2).

Missing data analysis

A low level of pupil attrition was expected within the context of the Reception Jigsaw trial due to the nature of the intervention (classroom based) and as test administrators were used to collect the endpoint assessment data. Despite this, we planned to address the issue of missing data if the proportion of participants with missing primary outcome data exceeded 5%. As per Table 11 (attrition data table), 14% of participants were missing their primary outcome data so a missing data analysis was conducted. We investigated patterns of missingness of the primary outcome variable by means of a two-level (pupil and school) logistic model, where the outcome is missingness and baseline EN, school region (Essex or Yorkshire), whether the school signed up to Mastering Number, and randomisation group indicators are covariates. Additionally, variables that we believed may be associated with missingness, but which were not included in the primary analysis, were also included as covariates. These were the proportion of pupils eligible for FSM, whether the school was urban or rural, whether the school was an academy or not, and its latest Office for Standards in Education, Children's Services and Skills (Ofsted) rating. As a result of this model (see Table 21) and following the roadmap from the EEF 2018 analysis guidance, (EEF, 2018)[14] a sensitivity analysis for the missing data was not carried out as the only variable associated with missingness was the EN checklist, which was included as a covariate in the model.

Subgroup analyses

As specified in the protocol, a subgroup analysis was run to investigate possible differential effects of the Reception Jigsaw on the attainment of children eligible for FSM. EVERFSM6 collected from the NPD was used as the identifier for the main subgroup analysis.

As discussed below in the 'Randomisation' section, Mastering Number, a programme with similar characteristics to the Reception Jigsaw, was implemented in 47 of the 138 participating schools. To better inform the interpretation of the trial's results and investigate the existence of interaction and confounding effects a subgroup analysis considering school participation in Mastering Number was included as an additional subgroup analysis.

The analyses were approached in two distinct ways: (i) by running models with interaction terms (i.e. models that include both the subgroup indicator and the product of the subgroup indicator and randomisation group); and (ii) by running separate primary outcome models on:

- just the FSM-eligible pupils; and
- schools that signed up for Mastering Number and on schools that did not sign up for the programme.

Both approaches conform to the EEF 2018 guidelines (EEF, 2018).

The multilevel random intercepts model with interaction terms for the FSM subgroup analysis is given by:

$$\mathrm{PUMA_{ij}} = \beta_0 + \mathrm{u_{0j}} + \beta_1 \mathrm{intervention_j} + \beta_2 \mathrm{baseline\ EN_{ij}} + \beta_3 \mathrm{FSM_{ij}} +$$

$$+\beta_4 \mathrm{FSM_{ij}} * \mathrm{intervention_j} + \beta_5 \mathrm{stratum_j} + \epsilon_{ij}$$

With $\mathrm{FSM_{ij}}$ being a dummy variable for pupil i in school j's FSM eligibility status and the remaining variables as described in the 'Primary analysis' subsection above.

And the two multilevel level random intercepts models with interaction terms for the participation/non-participation in Mastering Number subgroup analysis is given by:

$$\mathrm{PUMA_{ij}} = \beta_0 + \mathrm{u_{0j}} + \beta_1 \mathrm{intervention_j} + \beta_2 \mathrm{baseline\ EN_{ij}} + \beta_3 \mathrm{MN_j} +$$

$$+\beta_4 \mathrm{MN_j} * \mathrm{intervention_j} + \beta_5 \mathrm{stratum_j} + \epsilon_{ij}$$

[14] Work was completed under the expectation that there will be will no missing values among the models' covariates under MAR (missing at random), and that it will be possible to obtain valid estimates by including covariates predictive of non-response in the substantive models. The models' interpretation is conditional on these covariates being included. Covariates in the model were stratum variable and the EN checklist. Pupils were not included in the trial if the EN checklist was not completed so there were no missing data in the model's covariates.

Where MN_j is the indicator of whether school j has signed up for the Mastering Number programme or not, and the remaining variables as described above in the 'Primary analysis' subsection above.

Power analyses were performed to determine if the FSM subgroup analyses were underpowered. In accordance with the EEF 2018 guidelines (EEF 2018), underpowered subgroup analyses were reported as exploratory.

Estimation of effect sizes

Effect sizes have been calculated according to the formula:

$$\boldsymbol{g = \frac{\bar{o}i - \bar{o}c}{s^*}}$$

Where the numerators have been calculated as the coefficients of the intervention group from the regression models, and the denominator as the unconditional total variance from the corresponding models without covariates. The effect sizes thus computed are equivalent to Hedges' g.

For the first secondary outcome, results are reported as an odds ratio (OR) calculated according to the formula:

$$OR = \exp(\beta_1)$$

Where β_1 is the coefficient of the intervention/control dummy variable of the logistic regression model.

Estimation of ICC

ICCs were estimated from the variance of the random intercept and residual variance of multilevel models by means of the formula:

$$ICC = \frac{\sigma^2_{intercepts}}{\sigma^2_{intercepts} + \sigma^2_{residuals}}$$

Pre-test ICCs were computed considering random intercepts two-level (school and pupil) models with no covariates, and post-test ICCs were derived from the primary ITT and secondary ITT models described above.

Longitudinal analysis

During autumn term 2022, the decision was taken by EEF to proceed with the longitudinal follow-up following the criteria set out in the EEF longitudinal guidance (EEF, 2019). All the schools that were randomised were asked to re-administer the New PUMA (for Year 1) with the cohort of pupils who had taken part in the trial. These results were analysed using a separate multilevel model identical to the one specified in the primary analysis, except in this case the dependent variable was mathematics attainment as measured by the version of PUMA adopted for Year 1, rather than reception. As with the primary analysis the EN scores assessed at baseline are used to control for prior ability.

We completed a secondary analysis using the longitudinal follow-up (New PUMA - Year 1) data. This analysis repeated the primary longitudinal analysis but included dummy variables to take into account whether a school had taken part in the Primary Jigsaw during 2021/22 or 2022/23 (56 schools) or Reception Jigsaw in 2022/23 (control schools only – 8 schools) as participation in Jigsaw outside of the trial could bias the results if this in not accounted for (Schwendel, 2023). Missing case and compliance analyses were also completed using the longitudinal follow-up data.

Subgroup analyses were also undertaken using the subgroups as defined in the main analysis – namely FSM pupils and those in schools participating in Mastering Number - following EEF guidance (EEF, 2019) and the longitudinal statistical analysis plan (Schwendel, 2023). The membership of the subgroups is identical to those in the main analysis. The FSM measure uses the FSM status of pupils when they were in reception and may not reflect their FSM status in Year 1.

IPE

Research methods

The IPE is intended to complement the impact evaluation by providing information on how implementation of the Reception Jigsaw affected the outcomes of the trial.

The IPE investigated the following research questions:

1. Was the Reception Jigsaw delivered as intended in terms of dosage, nature, and quality?[15]
2. How well did the participants (teachers/support staff from reception and Key Stage 1, mathematics lead, senior leaders, and also pupils) engage with the Reception Jigsaw? Were there any implementation challenges faced? If so, what were they and to what extent were they overcome?
3. Was the quality of training, support, and intervention materials provided by the developer adequate? Was preparedness and confidence of staff delivering the intervention at the right level? If not, why?
4. To what extent do participants feel the intended outcomes of the programme are being achieved for children, practitioners, and the school? How are they being achieved?
5. What does 'business as usual' consist of for the comparison group?

A range of data collection methods were utilised in order to answer the IPE research questions. An overview of the IPE methods presented in Table 7 below.

[15] This relates to perceived quality, including aspects such as consistency of delivery against training materials, trainer responsiveness, and participant engagement levels.

Table 7: IPE methods overview

Research methods	Data collection methods (intended participants)	Participants / data sources (type, number)	Data analysis methods	IPE research question (RQ) addressed	Implementation / logic model relevance
Business as usual	Electronic proforma (baseline) (All schools) Online proforma (endpoint) (All control schools)	All schools (baseline) Control schools (endpoint)	Descriptive statistics; correlations	RQ5	Business as usual
Practitioner survey	Endpoint practitioner survey (online) (All intervention schools)	Intervention schools Long survey – reception teachers, reception teaching assistants (TAs), and mathematics leads Short survey – Year 1 teachers and headteachers	Descriptive statistics; correlations	RQ1, RQ2, RQ3, RQ4, RQ5	Fidelity, Dosage, Quality, Reach, Responsiveness, Programme differentiation, Monitoring control / comparison, Adaptation, Costs
Attendance register	School training and school visit attendance electronic log (All intervention schools)	All intervention schools	Frequency counts	RQ1, RQ2	Fidelity
Training observations	Structured observations (Intended n=5)	White Rose Maths developer team training observations (n=5)	Deductive coding; within-case analysis; cross-case analysis	RQ1, RQ2, RQ3	Fidelity, Dosage, Quality, Reach, Responsiveness, Adaptation, Costs
Trainer reflections	Electronic proforma (Intended n=5)	All Reception Jigsaw trainers (n=9)	Descriptive analysis	RQ1, RQ2, RQ4	Fidelity, Dosage, Reach, Adaptation
Case studies (longitudinal)	Semi-structured one to one face-to-face / telephone / online interviews (Intended n=4 schools, three timepoints for three staff, one timepoint for headteacher) Structured observations (Intended n=8)	Four intervention schools, with interviews at three time points (baseline, midpoint, and endpoint) with reception teachers, Year 1 teachers, mathematics leads, and at endpoint with headteachers or deputy headteachers Observations of training at baseline and school visit at endpoint	Inductive / deductive coding; thematic analysis; within-case analysis / triangulation; cross-case analysis	RQ1, RQ2, RQ3, RQ4, RQ5	Fidelity, Dosage, Quality, Reach, Responsiveness, Programme differentiation, Adaptation, Costs
Case studies (best practice)	Semi-structured one to one telephone / online interviews (Intended n=2 schools, four staff, one time point)	Two intervention schools at one time point (endpoint) with reception teachers, Year 1 teachers, mathematics coordinator, and headteacher	Inductive coding; thematic analysis; within-case analysis / triangulation; cross-case analysis	RQ1, RQ2, RQ3, RQ4, RQ5	Fidelity, Dosage, Quality, Reach, Responsiveness, Programme differentiation, Adaptation, Costs
Interviews with the developers	Semi-structured online interviews (Intended n=0)	Two White Rose Maths developer team members	Qualitative analysis	RQ1, RQ2, RQ3, RQ5	Context; Fidelity; Adaptation; Responsiveness, Costs

Data collection

Business as usual proforma/survey

In June 2021, teachers from control and intervention schools were asked to complete a baseline 'business as usual' proforma (an Excel spreadsheet). The proforma asked schools about numeracy pedagogy in reception and about the mathematics-related CPD for reception staff that took place in a typical year. This provided the research team with an understanding of usual practice across all schools. Of the 138 randomised schools, 134 schools (intervention n=54 and control n=80)[16] completed a baseline 'business as usual' proforma.

An endpoint proforma (an online survey) was sent to control schools at the end of the 2021/2022 academic year. This was an adapted version of the 'business as usual' questions asked at baseline. It covered any changes to numeracy pedagogy in reception since baseline and about the mathematics-related CPD that took place for reception and Key Stage 1 staff during the trial period. Overall, 57 teachers (from 57 control schools) completed an endpoint 'business as usual' proforma.

Practitioner survey

Questions about participation in the Reception Jigsaw were added to the endpoint practitioner survey for the secondary outcome on confidence teaching mathematics (long version). This was administered online to reception teachers, reception TAs, and mathematics leads in intervention schools in Summer Term 2022. The IPE-related elements of this survey explored staff engagement with and responsiveness to the intervention; particularly take up of the techniques. It also covered perceptions of the quality of training, support, and intervention materials, extent of achievement of intended outcomes and comparisons to 'business as usual' practice. Overall, 140 responses to the (long) endpoint practitioner survey from 55 intervention schools were received. This was made up of 70 reception teachers, 38 reception TAs, and 32 mathematics leads.

In order to reduce the burden on participating schools, a separate, short version of the online practitioner survey was administered (in May 2022) to Year 1 teachers and headteachers. It explored Year 1 teachers' engagement and responsiveness to the intervention, headteacher views relating to any changes at a whole-school level as a result of the intervention, and the associated costs. Overall, we received 47 responses to the short endpoint practitioner survey from 37 intervention schools. This was made up of 17 Year 1 teachers and 30 headteachers.

Attendance register

WRM offered Reception Jigsaw training to intervention schools between November 2021 and June 2022. This training was planned to be delivered as five in-depth twilight training sessions (each two hours), delivered face-to-face at each school. The developer team set out 'essential', 'recommended', and 'optional' participants to the training sessions. All reception teachers, Year 1 teachers, and the mathematics lead in each school were required to attend ('essential' participants). It was recommended that reception and Year 1 TAs also attend, and for Year 2 teachers/TAs (and other school staff) attendance was optional. The Reception Jigsaw also included five half-day coaching visits to schools. It was essential that reception teachers take part in those visits[17] and that mathematics leads also attend (although this could be for just part of the visit). Other school staff typically would not be expected to attend.

The developer collected a record of the attendance of all staff at each training session and school visit and logged it in an Excel spreadsheet. This data was collated and shared securely with NFER via a secure online portal. Training attendance data was analysed to measure compliance.

Trainer reflections

In order to further understand the support schools received and how and when the Reception Jigsaw was being implemented, trainers completed a reflections log (an Excel spreadsheet developed jointly by NFER and the developer) after each school visit they delivered. They were asked to record their perceptions of teacher engagement with the

[16] In total, 150 'business as usual' baseline proformas were received. Some of these were from schools, which did not go on to be randomised and have not been analysed. Where more than one proforma was received from the same school, one was randomly selected, and open responses were combined.
[17] In 2+ form entry schools, it was intended that participation in each coaching visit involve only one reception teacher (not all). Different reception teachers could take part in different coaching visits.

intervention, including completion of gap tasks and implementation of the learning in classrooms. Trainers were also asked to indicate the type of support they provided during each visit (e.g. revisiting content from the training sessions, more in-depth support towards developing in-school expertise, and/or support for implementation in the school). Questions were close ended, such as in the form of Likert scales, to reduce burden.

Trainer reflections data was collated by WRM, along with the information about attendance, and shared securely with NFER via a secure online portal. Completed reflections logs were received from all nine Reception Jigsaw trainers for each of their intervention schools.

Case studies

A series of case studies were conducted by the research team in order to explore implementation factors in relation to all assumptions of the logic model in greater depth.

There were four longitudinal case studies operating over the course of the intervention. The rationale behind these case studies was twofold. First, we anticipated a fair amount of turbulence in early years due to changes to statutory testing, which may have impacted on engagement with this intervention. Second, this intervention could have an impact at whole-school level, meaning there are various interdependencies that could have an effect on impact that would otherwise be difficult to capture (e.g. strength and style of leadership, school policy, amount and type of TA support, and the level of mathematics lead support and engagement). The focus of these case studies was on dimensions of fidelity, dosage, quality, reach, responsiveness, programme differentiation, adaptation, and costs. It was also intended that the longitudinal case studies would enable the research team to gain a deeper understanding of the school visits and how closely aligned they are with the school's perceived needs. The case study design is such that both a training session and a school visit was observed in each school.

The research team adopted a purposive sampling approach to select four schools from the intervention group. This allowed us to focus on particular characteristics of intervention schools that are of interest to help us answer the research questions. This included the need to achieve a geographical spread[18] and include schools who had/had not signed up to the Mastering Number programme. We also wanted to select schools receiving training from different WRM trainers to enable us to observe any potential differences in delivery. Using these characteristics we created a short list, from which four schools were approached to take part. All four schools agreed to participate.

Interviews with staff at longitudinal case study schools took place at three time points (baseline: November–December 2021, midpoint: March 2022, and endpoint: May–June 2022). In each school we sought to interview: a reception teacher; a Year 1 teacher; the mathematics lead (at all three time points); and, at the endpoint only, the headteacher. At baseline and endpoint these interviews typically took place face-to-face when we visited the school to observe the Reception Jigsaw training or school visit. Where staff members were unavailable on the day of the visit, we conducted the interviews remotely via video or telephone call at a later date. All interviews at midpoint were conducted remotely. In some of the longitudinal case study schools, staff were slow to respond to requests for interviews (particularly at the midpoint stage). We adapted our reminder strategy and increased the number of reminder phone calls and emails to interviewees to increase the response rate. We also offered greater flexibility around the timeline for interviews, as well as an opportunity to provide written feedback to interview questions via email to facilitate participation. Despite this, we were unable to carry out interviews with all of the required staff in some of the longitudinal case study schools. Where individuals provided a reason for their non-participation, this was either because they did not have time, due to capacity issues, or because they did not have direct involvement with the intervention. Table 8 sets out the number of longitudinal case study interviews achieved.

[18] There were three schools from the Yorkshire randomisation block and one from the Essex block.

Table 8: Interviews conducted in longitudinal case study schools

Longitudinal case study school	Baseline	Midpoint	Endpoint
School 1	Interviews x 3 with: • Reception teacher • Year 1 teacher • Mathematics lead	No interviews completed	Interviews x 3 with: • Reception teacher • Year 1 teacher • Mathematics lead
School 2	Interviews x 3 with: • Reception teacher • Year 1 teacher • Mathematics lead	Interviews x 2 with: • Reception teacher • Mathematics lead	Interviews x 3 with: • Reception teacher • Year 1 teacher • Mathematics lead
School 3	Interviews x 1 with: • Reception teacher	Interview x 1 with: • Reception teacher	Interviews x 3 with: • Reception teacher • Mathematics lead • Deputy headteacher
School 4	Interviews x 4 with: • Reception teacher (x 2) • Year 1 teacher • Mathematics lead	Interviews x 4 with: • Reception teacher (x 2) • Year 1 teacher • Mathematics lead	Interviews x 4 with: • Reception teacher • Year 1 teacher • Mathematics lead • Deputy headteacher

Observations of Reception Jigsaw training delivery (twilight session 1) took place in all four longitudinal case study schools in November 2021. An observation schedule was designed by NFER and populated by a member of the research team at each of the sessions they attended. These observations focused on how the training was delivered by the four trainers, any differences in quality, engagement, style of delivery or adaptation of content, and staff engagement and response.

Observations of coaching visits (coaching session five) took place in all four longitudinal case study schools in May 2022. An observation schedule was designed by NFER and populated by a member of the research team at each coaching visit they attended. These observations focused on the form and nature of the support provided, staff engagement, and implementation of the learning in classrooms.

In addition to the longitudinal case studies, we conducted two best practice case studies. These were carried out in two schools at a single timepoint towards the end of the intervention window (Spring/Summer Terms 2022). The purpose of these studies was to explore how these schools implemented the training to enable us to understand an ideal environment for the intervention. Towards the end of the delivery period (March 2022), WRM provided four nominations for best practice case study schools. The research team selected two to approach in the first instance. These two schools were chosen to offer further geographical spread compared to the longitudinal case studies. Both agreed to participate in the research. In both schools, we interviewed a reception teacher, a Year 1 teacher, the mathematics lead, and the headteacher.

Training observations

The training of the WRM early years specialists (early years SLEs) was observed by the research team to enable a better understanding of the Reception Jigsaw intervention and how it should be delivered, and to observe the levels of engagement and interaction among trainees. Observation of the first 'training of the trainers' session, delivered by the development team, took place in person in September 2021. Subsequent training sessions (Sessions 2–5) were observed remotely (via Zoom). An observation schedule was designed by NFER and populated by a member of the research team at each of the training sessions they attended.

Interviews with the developer team

In August 2022 (towards the end of the trial), the research team conducted online interviews with two members of the WRM developer team to support our understanding of delivery and the support provided to teachers, as well as the

costs of the intervention as it was delivered in the evaluation. Developers also provided secondary data in the form of feedback from trainers.

Data collection instruments

The research team developed the following data collection instruments: baseline and endpoint proformas; endpoint practitioner surveys; school staff and developer interview schedules; and observation schedules. The teacher survey included two of the three scales from Chen *et al.,* (2014) in order to assess any impacts of the Reception Jigsaw on teacher confidence in mathematics and in teaching mathematics (see 'Outcome measures' section and 'Outcomes and analysis' section below). The developer also contributed to the design of the proforma and trainer reflection logs to ensure they captured data on the key components of the intervention. The research instruments were developed to meet the specific requirements of the trial, although they follow a similar format to others used in the EEF trials conducted by NFER. The research team collected all data from surveys, interviews, and observations. WRM collected and collated the attendance register and trainer reflections data. This information was shared securely with NFER.

Rationale for the data collection methods utilised

NFER researchers chose the range of data collection methods outlined above as they offered both breadth and depth to the IPE. The logic model was used to help prioritise data collection to focus on the key features of the intervention and the assumptions underpinning it. The surveys provide an efficient way of measuring implementation and fidelity and inform our understanding of usual practice across a large number of schools. However, as these are self-reported measures, observational evidence of the delivery was also sought. Combined with additional qualitative information gained from interviews with school staff, these data collection sources yielded further insight into the perceived impacts of the Reception Jigsaw. As noted in the 'Limitations and lessons learned' subsection below, the IPE is limited by the small number of observations and interviews conducted and the relatively low response rate from Year 1 teachers and headteachers to the endpoint survey, and any insights should be viewed as tentative.

Analysis

The IPE data available for analysis is summarised as:

- observations of four Reception Jigsaw training sessions;
- observations of four Reception Jigsaw school visits;
- baseline interviews with 11 teachers from four longitudinal case study schools;
- midpoint interviews with seven teachers from four longitudinal case study schools;
- endpoint interviews with 13 teachers from four longitudinal case study schools;
- interviews with eight teachers from two best practice case study schools;
- business as usual baseline proformas from 134 teachers in intervention (n=54) and control (n=80) schools;
- business as usual endpoint proforma from 57 schools (control schools only);
- endpoint practitioner surveys (long and short) from 187 practitioners in 55 intervention schools; and
- interviews with two members of the developer team.

The research team summarised qualitative data from observations and interviews with case study schools and developers in a grid representing data sources mapped against research questions (e.g. fidelity, quality of delivery, and adaptation). The analysis used both inductive and deductive approaches (i.e. seeking to identify patterns in the data as well as to test the assumptions in the logic model) to enable detailed analysis of the available data.

Proforma and survey responses were analysed to explore the school- and teacher-level impacts specified in the logic model, namely any differences between control and intervention groups at endpoint in relation to practitioners' confidence in mathematics and in teaching maths, and application of learning from the intervention in the classroom. Quantitative data from teacher surveys was analysed using R. This analysis was conducted with a full audit trail and quality-assured by a senior statistician at NFER. The attendance data was collated and summarised using Microsoft Excel.

The research team triangulated the IPE qualitative and quantitative data sources in order to cross-validate the results and support the interpretation of the findings.

Costs

As part of the cost evaluation, cost estimates in terms of both time and financial resource were collected from a range of sources. The main estimates of the costs of taking part in the Reception Jigsaw are based on how the intervention was delivered in the trial. All costs (time and financial) of taking part in the Reception Jigsaw are presented as compared to schools, which are not taking part in the programme. Business as usual costs (financial only) were collected and reported but are not included in the main cost estimate.

Time

Estimates of the amount of time required to fulfil the requirements of the Reception Jigsaw (the training, gap task, coaching visits, and implementing strategies/approaches in the classroom) were gathered. Estimates for tasks where there was no or little variation foreseen between schools were not collected from all schools, whereas estimates for tasks where a larger range of times was expected, were asked to all intervention teachers in the endpoint survey.

- *Training time:* Time expectations for twilight sessions, coaching visits, and gap tasks (collected from WRM), attendance (collected from attendance registers).
- *Implementation time:* Additional time spent by reception teachers, reception TAs, mathematics leads on implementing the learning from Reception Jigsaw outside of the twilight sessions, coaching visits, and gap tasks (collected from the endpoint survey).

Financial

Following the EEF guidance on cost evaluation, costs were considered in terms of pre-requisite costs, start-up costs, or recurring costs (EEF, 2019). The programme costs to schools for receiving the Reception Jigsaw as part of the trial were taken from the MoU schools signed. Other financial costs related to taking part in the Reception Jigsaw were collected via the endpoint surveys (both long and short versions) and case study interviews.

The cost per pupil per year was calculated assuming that Reception Jigsaw programme was followed in intervention schools for three academic years, following the EEF cost guidance (EEF, 2019). Three years is used as an estimate for the average number of years a teacher is likely to stay in the same school after having the training.[19] Costs across years are adjusted to take account of inflation. The number of pupils who were considered to have received the intervention were all those in the reception classes of schools in the intervention group. This data was gathered from schools in Summer Term 2021 and early Autumn Term 2021 when schools first provided their pupil data. The Reception Jigsaw training was also attended by teachers and TAs from Year 1 and Year 2 so, there are likely to be broader effects beyond the reception year group, however the focus of the programme is reception practitioners and pupils.

Market costs

The programme costs for the Reception Jigsaw if undertaken outside of the trial are higher for schools. In order to provide an estimate of costs to schools under market conditions, the market programme costs were taken into account to provide a market cost estimate per pupil per year. This is reported separately to the main per pupil cost estimate.

Business as usual costs

A baseline proforma was used in Summer Term 2021 to gather data from both control and intervention schools on any mathematics-related CPD normally provided to reception teachers and estimates of the costs of this. The CPD undertaken by reception and Key Stage 1 teachers and/or TAs in control schools was monitored using an endpoint proforma completed in Summer Term 2022. Costs associated with any mathematics-related CPD undertaken during the same academic year of the trial (2021/2022) in schools in the control group were collected.

[19] As stated in the EEF cost guidance (EEF, 2019), teachers complete an average of 6.7 years in one school (Allen *et al.*, 2012) so, three years is used as an estimate for the length of time a teacher will have left in a particular school before they move.

Timeline

Table 9: Timeline

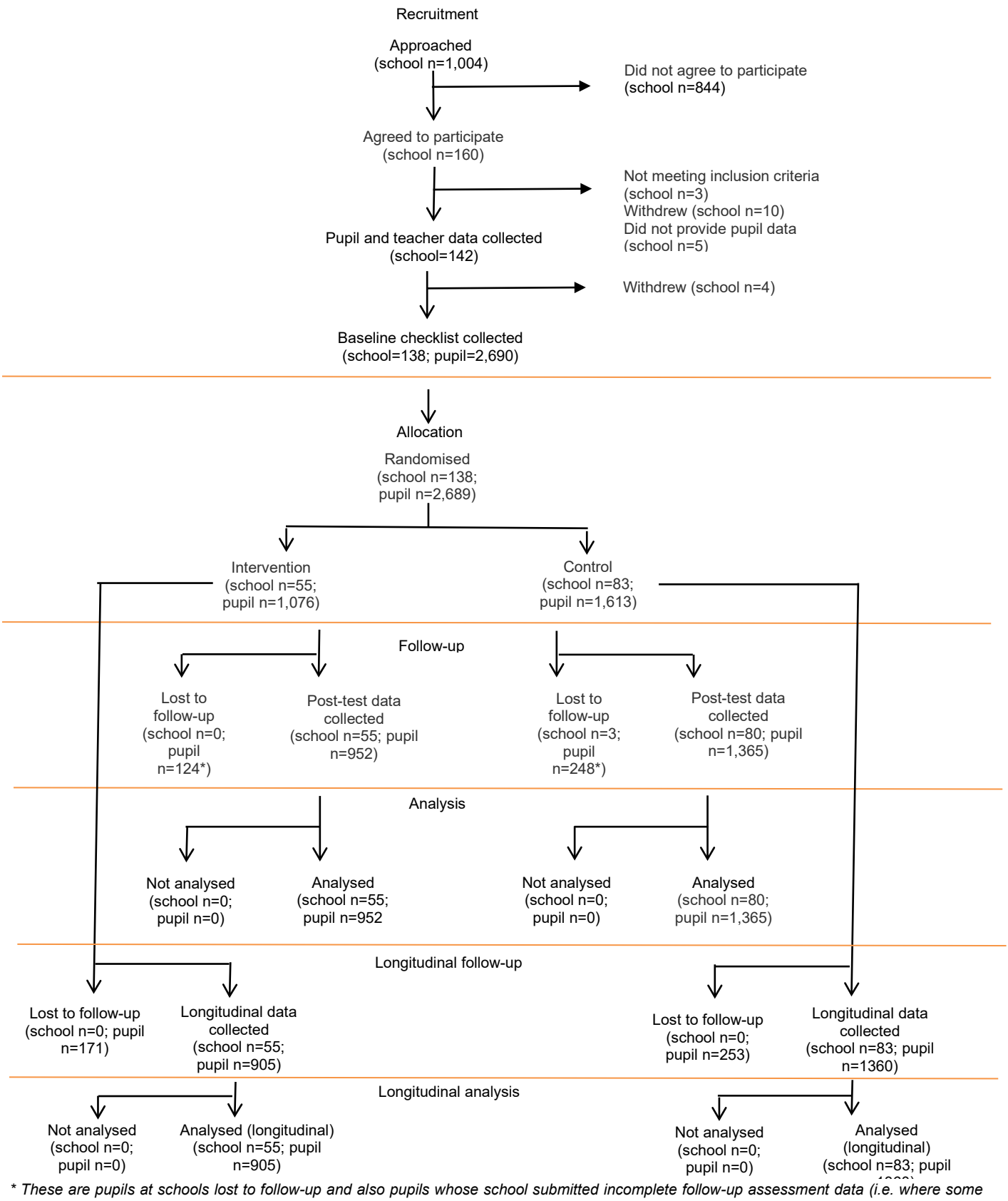
Dates	Activity	Staff responsible / leading
July 2019	• Project set-up meetings	Helen Poet
October 2019 – December 2019	• Protocol writing	Helen Poet, Ben Styles
November 2019	• IDEA workshop	Helen Poet
November 2019 – May 2020	• School recruitment (original)	WRM, Guido Miani, Kathryn Hurd
March 2020	• **Project paused due to school closures (Covid-19)**	**–**
November 2020 – May 2021	• School recruitment—recontact schools recruited in 2019/2020 + top-up recruitment to required numbers	WRM, Alison Hale, Kathryn Hurd
June 2021 – September 2021	• Pupil data collection, including FSM eligibility (requested from schools after receipt of MoU) • Teacher data collection • Business as usual proforma in all schools (baseline)	WRM, Alison Hale, Asma Ulla, Kathryn Hurd
September 2021 – March 2022	• Observation of training of the trainers	WRM, Kerry Martin, Eleanor Bradley
September / October 2021	• RBA • Schools to complete observational checklist after their school has completed the RBA • Practitioner survey in all schools (baseline)	Alison Hale, Kathryn Hurd
October 2021	• Email to schools to ask if they were taking part in the Mastering Number programme	Asma Ulla, Alison Hale, Kathryn Hurd
October 2021	• Randomisation	Ben Styles, Joana Andrade
November 2021	• Intervention commences	WRM
November 2021 – December 2021	• First set of longitudinal case study visits (including observations)	Kerry Martin, Eleanor Bradley
January 2022 – March 2022	• Second set of longitudinal case study interviews	Kerry Martin, Eleanor Bradley
May 2022 – June 2022	• Third set of longitudinal case study visits • Best practice case study interviews (selected by developer)	WRM, Kerry Martin, Eleanor Bradley
May 2022	• Intervention ends	WRM
June 2022	• Reception pupils sit New PUMA endpoint test, NFER administrators in schools	Alison Hale, Kathryn Hurd, Connie Rennie
June 2022 – July 2022	• Practitioner survey in all schools (endpoint) • Business as usual proforma in control schools (endpoint)	Alison Hale, Kathryn Hurd
June 2022 – July 2022	• Schools submit pupils' EYFSP data to the DfE	Schools

Dates	Activity	Staff responsible / leading
August 2022	• NFER thank schools for taking part in the trial and feedback New PUMA results to schools	Guvi Chohan, Kathryn Hurd
September 2022	• Incentives paid to schools	WRM
August 2022 – November 2022	• Analysis of outcomes for reception: New PUMA (primary outcome) • Analysis of practitioner confidence (secondary outcome)	Ben Styles, Gemma Schwendel, Sarah Tang
October 2022 – January 2023	• Report writing	Eleanor Bradley, Kerry Martin, Gemma Schwendel, David Sims, Sarah Tang, Ben Styles
December 2022	• Schools sent feedback form regarding involvement in trial and informed of the longitudinal follow-up planned for next summer	Guvi Chobhan, Kathryn Hurd
January 2023	• Analysis of the EYFSP	Gemma Schwendel, Ben Styles
June 2023 – July 2023	• Year 1 pupils sit New PUMA longitudinal follow-up test, NFER administrators in schools	Kathryn Hurd, Guvi Choban, Lydia Wallis
August 2023	• Analysis of outcomes for Year 1: New PUMA (longitudinal follow-up)	Gemma Schwendel, Ben Styles
September 2023	• Publication of final report • Analysis of outcomes for Year 1: New PUMA (longitudinal follow-up) • Incentives paid to schools	Gemma Schwendel, Ben Styles Kathryn Hurd, Lydia Wallis
October 2023	• Update report with longitudinal findings	Sarah Tang, Gemma Schwendel, Ben Styles

Impact evaluation results

Participant flow including losses and exclusions

Figure 2: Participant flow diagram



** These are pupils at schools lost to follow-up and also pupils whose school submitted incomplete follow-up assessment data (i.e. where some pupil endpoint assessment data was missing).*

WRM recruited 145 schools to take part in the trial, with an additional 15 schools added to a waitlist. Of the original 145 schools, it transpired that three were independent schools and were thus not eligible to take part in the trial. During the second half of the Summer Term and summer holidays 2021, a decision was made to include the 15 waitlist schools in the trial, so a total of 160 schools were recruited to the trial (with signed MoUs). This was in response to school withdrawals and concern around potential higher numbers dropping out due to Covid-related pressures and DfE's launch of the new Mastering Number programme (see 'Background' section above). In Summer Term 2021, schools were asked to submit data to NFER on the pupils due to start reception in the following academic year via the NFER secure online portal. Data on the pupil names, date of birth, unique pupil number (UPN), FSM status, class name, and teacher name were requested. After this, in the first half of Autumn Term 2021, schools were asked to complete the EN checklist for 20 of their reception pupils in order to continue in the trial. Completion of the EN checklists was required in order to be put forward for randomisation and to continue in the trial. In the end, 138 schools were randomised. Of the 19 schools that did not go on to be randomised, five schools stated that this was due to staff changes or a change in the headteacher (one of these schools also mentioned the Covid pandemic), one school wanted consent to be as opt-in, one school stated unforeseen circumstances, a further seven schools did not give a reason, and five schools did not respond to requests for pupil data.[20]

The minimum detectable effect size at each stage of the evaluation is shown in Table 11. The evaluation was designed to have an MDES of 0.22 and a slightly lower MDES was achieved at all stages of the trial, including the longitudinal follow-up. The trial was not powered to detect an effect for FSM pupils but the aim of assessing five pupils eligible for FSM per school was achieved. The intra-cluster correlation (ICC) and the pre/post-test correlation were higher than predicted.

Table 10: Minimum detectable effect size (MDES) at different stages

		Protocol		Randomisation		Analysis		Longitudinal follow-up	
		Overall	FSM	Overall	FSM	Overall	FSM	Overall	FSM
MDES		0.22	0.27	0.21	0.26	0.21	0.265	0.20	0.226
Pre-/post-test correlations	**Level 1 (pupil)**	0.4	0.4	0.4	0.4	0.59	0.54	0.54	0.47
Intracluster correlations	**Level 2 (school)**	0.17	0.17	0.17	0.17	0.24	0.23	0.20	0.10
Alpha		0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05
Power		0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8
One-sided or two-sided?		Two-sided	Two-sided	Two-sided	Two-sided	Two-sided	Two-sided	Two-sided	Two-sided
Average cluster size		20	5	19.49	5	17.0	5.1	19.3	5.1
Number of schools	**Intervention**	55	55	55	55	55	47	55	47

[20] These five schools who did not respond to requests for data could not remain in the trial but did not actively withdraw.

	Control	74	74	83	83	80	78	83	78
	Total:	129	129	138	138	135	125	138	125
Number of pupils	**Intervention**	1,100	275	1,076	275	952	216	905	253
	Control	1,480	370	1,613	415	1,365	321	1360	385
	Total:	2,580	645	2,689	690	2,317	537	2265	638

FSM=free school meals.

Attrition

Table 11 shows the pupil-level attrition from randomisation to analysis. At a school level, three control schools from the 83 randomised schools did not respond to requests from the evaluation team to supply possible dates for endpoint testing availability and therefore, did not submit any of their endpoint pupil assessment data (see Figure 2). All 55 intervention schools completed at least some of their endpoint New PUMA data. In summer 2023, pupils from all schools that were randomised (138 schools) completed the longitudinal follow-up assessment. Attrition was therefore zero for the longitudinal follow-up at school-level. At pupil-level, attrition was 15.8 per cent for the longitudinal follow-up, slightly higher than the post-test where it was 13.8 per cent.

Table 11: Pupil-level attrition from the trial (primary outcome)

		Intervention	Control	Total
Number of pupils	Randomised	1,076	1,613	2,689
	Analysed	952	1,365	2,317
Pupil attrition (from randomisation to main analysis)	Number	124	249	373
	Percentage	11.5%	15.4%	13.8%
Pupil attrition (from randomisation to longitudinal analysis)	Number	905	1360	2265
	Percentage	15.9%	15.7%	15.8%

Pupil and school characteristics

Balance at baseline results

The balance at baseline for the two groups have been analysed both as randomised and analysed against a number of characteristics:

- proportion of FSM-eligible pupils within the school;
- if the school is rural or urban;
- type of school governance; and
- the latest Ofsted rating.

Analysis of the data for all randomised pupils reveals that there are notable imbalances at baseline for proportion of FSM-eligible pupils, type of school governance, and latest Ofsted rating. There are more pupils in schools with higher proportions of FSM-eligible pupils within the control group, with just over half (52%) of these pupils falling within the top two quintiles. Around one in five (21%) of pupils within the control group attend a school whose latest Ofsted rating was 'Outstanding', notably higher than the equivalent proportion within the intervention group (14%). Around 56% of pupils within the control group attend an academy or free school compared to 38% of pupils within the intervention group. By contrast, there is very little difference between the two groups in the proportions who attend urban schools (87% and 88% for the control and intervention groups, respectively). We see a similar pattern of imbalance and notable differences among the analysed pupils' data (see Table 13 for more details). The differences seen between the two groups were already present at randomisation, so we are confident that they are not indicative of any systematic bias.

The distribution of the baseline EN scores within the two groups are broadly consistent (see histogram in Appendix K), with little difference in the mean scores (41.4 and 40.5 for intervention and control, respectively).

Despite the imbalances in pupil characteristics at baseline, the effect size of grouping on the baseline EN score is small for both randomised and analysed (0.04 and 0.06, respectively).

Table 12: Baseline characteristics of groups as randomised

School level (categorical)	**National-level mean**	**Intervention group**		**Control group**		
		n/N (missing)	**%**	**n/N (missing)**	**%**	
% FSM eligibility national quintile 2018/2019						
Lowest 20%		121/1,038 (38)	15.0%	156/1,533 (80)	7.9%	
2nd lowest 20%		301/1,038 (38)	25.8%	268/1,533 (80)	19.6%	
Middle 20%		319/1,038 (38)	13.4%	139/1,533 (80)	20.8%	
2nd highest 20%		399/1,038 (38)	19.2%	199/1,533 (80)	26.0%	
Highest 20%		393/1,038 (38)	26.6%	276/1,533 (80)	25.6%	
Overall Ofsted rating 2018/2019						
Outstanding		140/1,038 (38)	13.5%	318/1,493 (120)	21.3%	
Good		726/1,038 (38)	69.9%	976/1,493 (120)	65.4%	
Requires Improvement		152/1,038 (38)	14.6%	160/1,493 (120)	10.7%	
Inadequate		20/1,038 (38)	1.9%	39/1,493 (120)	2.6%	
School type						
Academy / FSM		400/1,058 (18)	37.8%	886/1,593 (20)	55.6%	
Maintained		658/1,058 (18)	62.2%	707/1,593 (20)	44.4%	
Urban/rural						
Rural		127/1,058 (18)	12.0%	212/1,593 (20)	13.3%	
Urban		931/1,058 (18)	88.0%	1381/1,593 (20)	86.7%	
Pupil level (continuous)		**n/N (missing)**	**Mean (SD)**	**n/N (missing)**	**Mean (SD)**	**Effect size**
Baseline EN	Not available at the national level: EN was a bespoke tool used for this project	1,076/1,076 (0)	40.5 (10.8)	1,613/1,613 (0)	40.0 (11.0)	0.04 (-0.12,0.21)

EN=Emerging Numeracy; FSM=free school meals; Ofsted=Office for Standards in Education, Children's Services and Skills; SD=standard deviation.

Table 13: Baseline characteristics of groups as analysed

School level (categorical)	**National-level mean**	**Intervention group**		**Control group**		
		n/N (missing)	**%**	**n/N (missing)**	**%**	
% FSM eligibility national quintile 2018/2019						
Lowest 20%		144/918 (34)	15.7%	107/1,296 (69)	8.3%	
2nd lowest 20%		247/918 (34)	26.9%	248/1,296 (69)	19.1%	
Middle 20%		122/918 (34)	13.3%	285/1,296 (69)	22.0%	
2nd highest 20%		167/918 (34)	18.2%	319/1,296 (69)	24.6%	
Highest 20%		238/918 (34)	25.9%	337/1,296 (69)	26.0%	
Overall Ofsted rating 2018/2019						
Outstanding		125/918 (34)	13.6%	271/1,260 (105)	21.5%	
Good		651/918 (34)	70.9%	826/1,260 (105)	65.6%	
Requires Improvement		128/918 (34)	13.9%	129/1,260 (105)	10.2%	
Inadequate		14/918 (34)	1.5%	34/1,260 (105)	2.7%	
School type						
Academy / FSM		345/934 (18)	36.9%	757/1,347 (18)	56.2%	
Maintained		589/934 (18)	63.1%	590/1,347 (18)	43.8%	
Urban/rural						
Rural		113/934 (18)	12.1%	191/1,347 (18)	14.2%	
Urban		821/934 (18)	87.9%	1,156/1,347 (18)	85.8%	
Pupil level (continuous)		**n/N (missing)**	**Mean (SD)**	**n/N (missing)**	**Mean (SD)**	**Effect size**
Baseline EN	Not available at the national level: EN was a bespoke tool used for this project	952/952 (0)	41.4 (10.4)	1,365/1,365 (0)	40.5 (10.8)	0.06 (-0.1,0.2)

EN=Emerging Numeracy; FSM=free school meals; Ofsted=Office for Standards in Education, Children's Services and Skills; SD=standard deviation.

Outcomes and analysis

Primary analysis

The mean New PUMA score is 14.2 overall, with a small difference between the means for the intervention and control groups (14.6 and 14.0, respectively). Hedges' g is positive at 0.08 (confidence interval [CI]: -0.07, 0.23) but the CI contains zero, with a p-value of 0.3. Pupils with teachers that had access to the Reception Jigsaw did make a small amount of additional progress in mathematics (estimated around one month) compared to pupils with teachers who continued with business as usual; however, the statistical evidence does not meet the threshold set by the evaluator to conclude that the true impact was non-zero. As shown in Figure 3 below, the distribution of the New PUMA total score is broadly Gaussian, with little difference in the distribution between the intervention and control groups.

Figure 3: Histogram of primary outcome, split by grouping

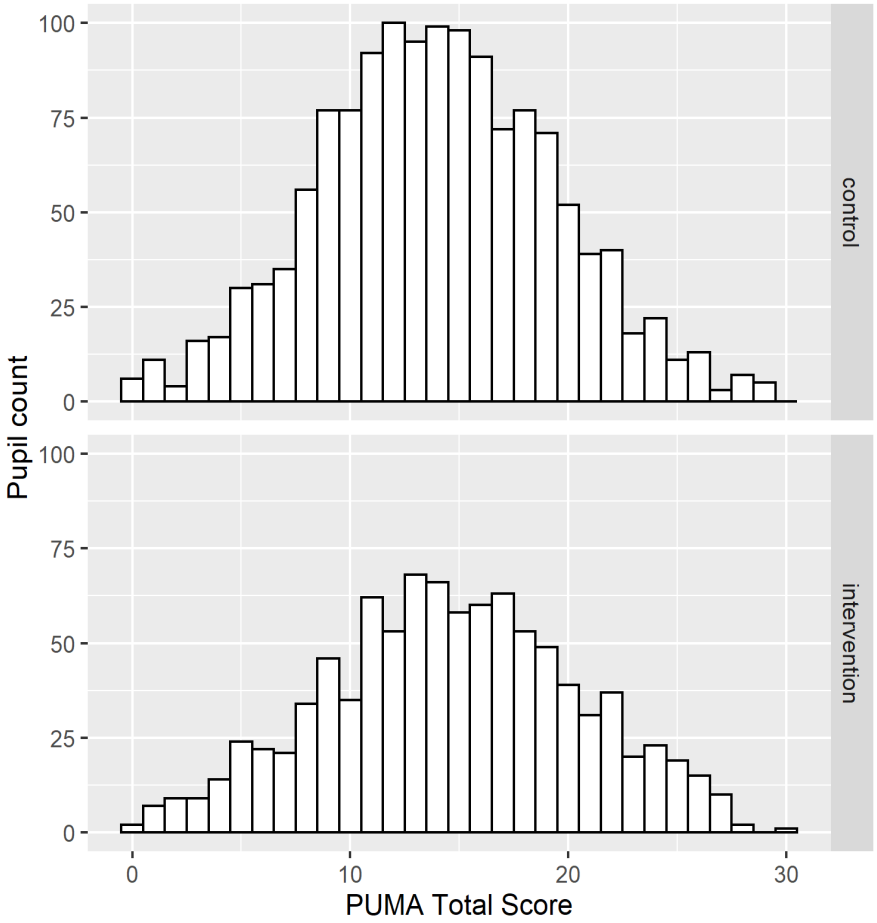


Table 14: Primary analysis

	Unadjusted means				Effect size			
	Intervention group		Control group					
Outcome	**N (missing)**	**Mean (95% CI)**	**n (missing)**	**Mean (95% CI)**	**Total n (intervention, control)**	**ICC (full)**	**Hedges' g (95% CI)**	**P-value**
PUMA Total Score	952 (124)	14.6 (14.3, 15.0)	1,365 (248)	14.0 (13.7, 14.3)	2,317 (952; 1365)	0.24	0.08 (-0.07, 0.23)	0.3

Secondary analysis

Secondary outcome 1 (EYFSP)

Overall, just over three-quarters (76.1%) of pupils met the required standard for mathematics attainment at the end of the reception year. Slightly more pupils within the intervention group met the standard compared to the control group (76.4% and 75.8%, respectively), but as the OR[21] CI overlaps unity, we cannot conclude that this was a genuine effect (Table 15).

Table 15: Secondary analysis 1 (EYFSP)

	Proportion				Effect size		
	Intervention group		Control group				
Outcome	N (missing)	Proportion (95% CI)	n (missing)	Proportion (95% CI)	Total N (intervention, control)	OR (95% CI)	P-value
EYFSP	1,053 (23)	76.4% (73.7%, 77.9%)	1,591 (22)	75.8% (73.9%, 79.0%)	2,644 (1,053, 1,591)	0.93 (0.64,1.36)	0.72

CI=confidence interval; EYFSP=early years foundation stage profile; OR=odds ratio.

Secondary analysis 1 (EYFSP additional analysis – Number and Numerical Patterns goals separately)

Exploratory analysis was carried out on the EYFSP where the logistic model was run using the two mathematics goals (Number and Numerical Patterns) separately. Following the main EYFSP analysis, a slightly higher proportion of the intervention group achieved the expected level in both goals as shown in Table 16. The OR for both spans 1 so there is not enough statistical evidence to conclude that this is a genuine effect.

Table 16 Secondary analysis 1 (EYFSP additional analysis – Number and Numerical Patterns goals separately)

	Proportion				Effect size		
	Intervention group		Control group				
Outcome	N (missing)	Proportion (95% CI)	n (missing)	Proportion (95% CI)	Total N (intervention, control)	OR (95% CI)	P-value
EYFSP - Number	1,053 (23)	78.2% (75.7%, 80.7%)	1,591 (22)	77.9% (75.8%, 79.9%)	2,644 (1,053, 1,591)	0.92 (0.62,1.37)	0.68
EYFSP – Numerical Patterns	1,053 (23)	77.7% (75.2%, 80.2%)	1,591 (22)	77.1% (75.0%, 79.1%)	2,644 (1,053, 1,591)	0.94 (0.65, 1.37)	0.75

[21] An Odds Ratio (OR) shows the odds that a pupil will achieve the expected level in the EYFSP if they have been in a school taking part in Reception Jigsaw (an intervention school) compared to the odds that a pupils will achieve it if they were not in a school taking part in Reception Jigsaw (a control school).

Secondary outcome 2 (confidence in teaching mathematics)

Reception Jigsaw did have an impact on practitioners' confidence in teaching mathematics. Practitioners included in this survey were reception teachers, reception TAs, and school mathematics leads. The difference in mean scores on the 'Confidence in teaching mathematics' scale is 3.8 points and the effect size (Hedges' g) is 0.58 (CI: 0.36, 0.79) (Table 17). As the CI does not include zero we can conclude that there is a positive impact of the Reception Jigsaw programme on practitioner confidence in teaching mathematics. Figure 4 shows the difference between the outcome measure of the two groups. It is important to highlight that there is a high level of attrition in the responses. There is a slightly higher level of attrition in the control group (47%) compared to the intervention group (42%). High levels of attrition could be a source of bias where practitioners with particular characteristics are more likely to respond to the survey. Here, it is reassuring that attrition is reasonably balanced across the two groups. Further exploratory analysis shows there is a statistically significant difference in mean scores by respondent's role, with reception teachers scoring more highly on average than either of the other two groups.

Figure 4: Histogram of secondary analysis 2 (confidence in teaching mathematics), split by grouping

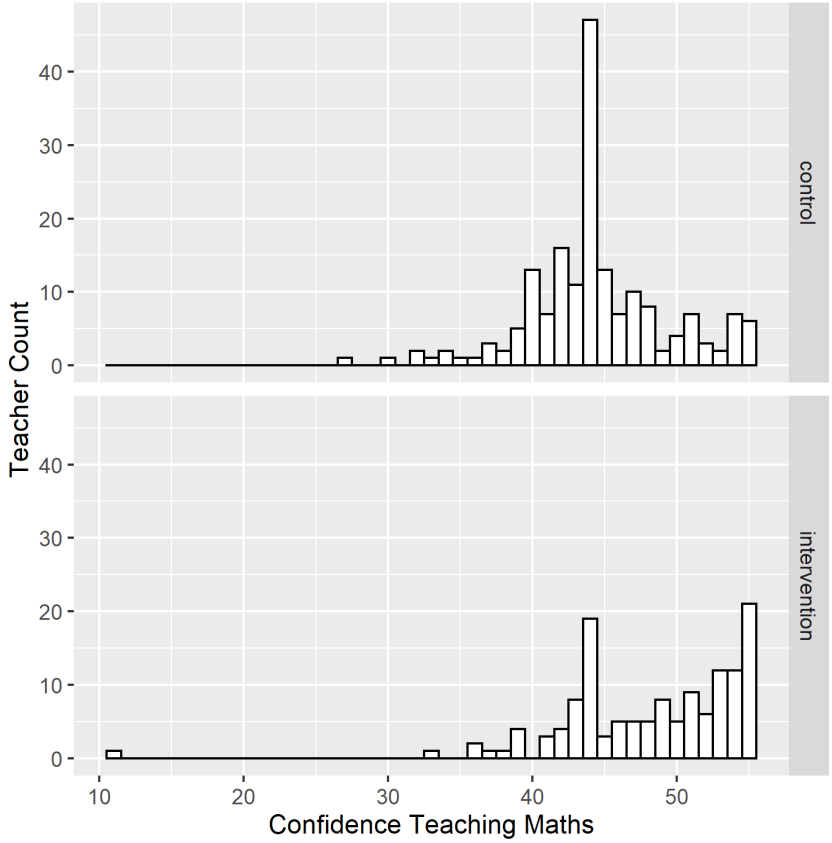


Table 17: Secondary analysis 2 (confidence in teaching mathematics)

	Unadjusted means				Effect size			
	Intervention group		Control group					
Outcome	**n (missing)**	**Mean (95% CI)**	**N (missing)**	**Mean (95% CI)**	**Total n (intervention, control)**	**ICC (full)**	**Hedges' g (95% CI)**	**P-value**
Confidence in teaching mathematics	135 (99)	48.2 (47.2, 49.3)	182 (160)	44.4 (43.7, 45.1)	317 (135, 182)	0.10	0.58 (0.36, 0.79)	0.000

Secondary outcome 3 (confidence in mathematical ability)

There is insufficient evidence to draw a conclusion as to whether the Reception Jigsaw has an impact on practitioners' confidence in their own mathematical ability. While the histograms (Figure 5) appear to show a small positive difference between them, there is only a very small difference in the means of the intervention and control groups.

Figure 5: Histogram of secondary analysis 3 (confidence in mathematical ability), split by grouping

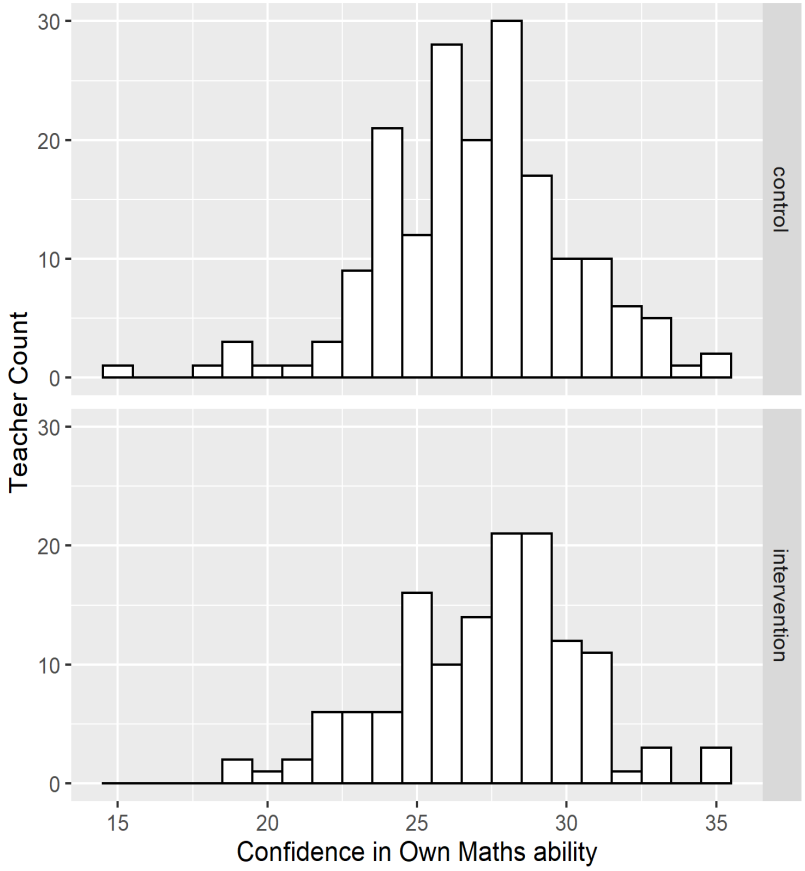


Table 18: Secondary analysis 3 (confidence in mathematical ability)

	Unadjusted means				Effect size			
	Intervention group		Control group					
Outcome	**n (missing)**	**Mean (95% CI)**	**N (missing)**	**Mean (95% CI)**	**Total n (intervention, control)**	**ICC (full)**	**Hedges' g (95% CI)**	**P-value**
Confidence in mathematical ability	135 (99)	27.3 (26.8, 27.8)	181 (161)	27.0 (26.5, 27.4)	316 (135, 181)	0.00	0.04 (-0.15, 0.22)	0.7

Secondary outcome 4 (reduced New PUMA)

Reflecting the results for the primary New PUMA outcome, while the pupils in the intervention group did make additional progress (estimated to be less than a month) compared to the control group, there is not enough statistical evidence from this alone to draw conclusions of the Reception Jigsaw's impact on pupils' mathematics attainment. There is a small positive difference between the means of the intervention and control groups (12.0 and 11.5, respectively) but as the CI for the effect size contains zero, we cannot conclude that this difference did not occur by chance (Table 19).

Figure 6: Histogram of secondary analysis 4 (PUMA 20), split by grouping

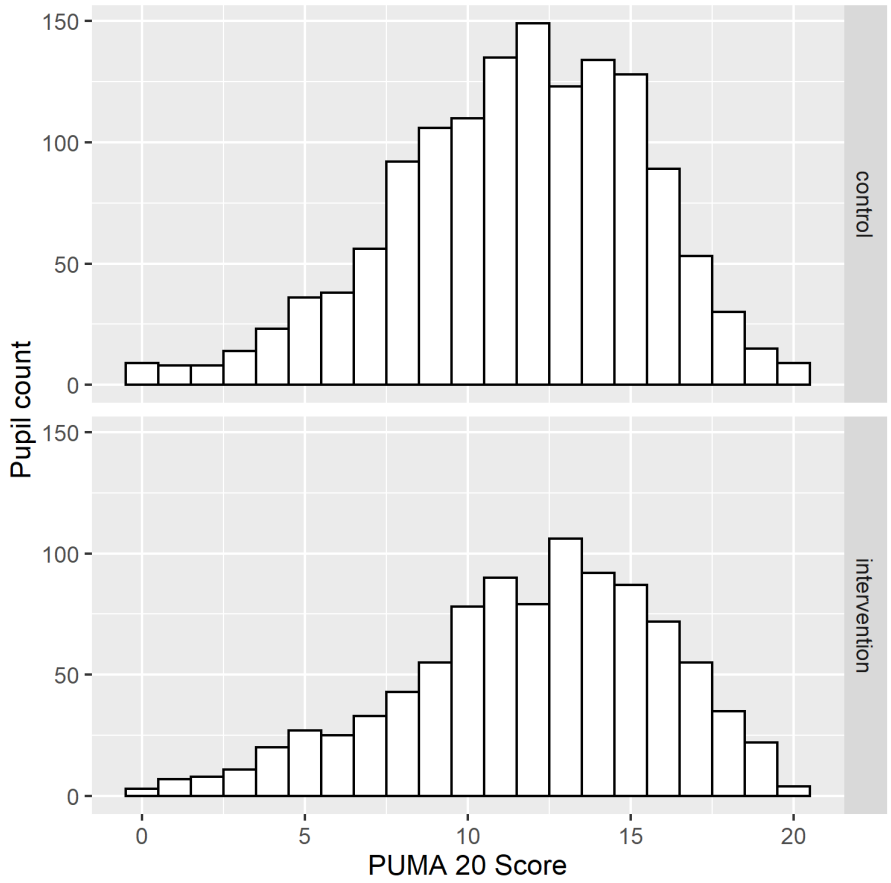


Table 19: Secondary analysis 4 (reduced New PUMA)

	Unadjusted means				Effect size			
	Intervention group		Control group					
Outcome	**n (missing)**	**Mean (95% CI)**	**n (missing)**	**Mean (95% CI)**	**Total n (intervention, control)**	**ICC (full)**	**Hedges' g (95% CI)**	**P-value**
Reduced PUMA (PUMA 20)	952 (124)	12.0	1,365 (248)	11.5	2317 (952; 1,365)	0.22	0.07 (-0.08, 0.21)	0.4

CI=confidence interval; ICC=intracluster correlation coefficient; PUMA=Progress in Understanding Mathematics Assessment.

Analysis in the presence of non-compliance

An analysis was completed on the outcomes of pupils who were taught by a teacher who was considered to have complied with the programme requirements. Teachers were expected to have attended nine of the ten possible sessions (five twilight CPD sessions and five coaching sessions) to be judged as compliant to the programme. Around 65% of intervention pupils (695) were taught by teachers who had complied with the programme requirement (see the 'Compliance' section for more detail on compliance patterns). Table 20 shows the impact of the Reception Jigsaw on mathematical attainment when pupils are taught by teachers who had attended at least nine out of ten of the sessions. There is a positive effect size (0.13) for these pupils in terms of their mathematical ability, with the whole CI lying above zero, suggesting that the Reception Jigsaw programme has a small beneficial effect on mathematical attainment among pupils whose teachers attended most of the training. The p-value is 0.02 suggesting there is likely to be a difference in outcomes between those pupils taught by a teacher who attended at least nine sessions and those who were not. As

described in the SAP (Andrade *et al.*, 2022) a pseudo-continuous compliance measure was also used. The findings are presented in Table 21 and are in line with the binary measure.

Table 20: Analysis in the presence of non-compliance (binary compliance measure)

Model stage	Total n (intervention, control)	Predictor	Standardised effect size (95% CI)	P-value
Stage 1: Compliance indicator regressed on intervention status*	2,270 (917, 1,353)	Intervention status	1.5 (1.4, 1.5)	0.00
Stage 2: PUMA score regressed on compliance indicator*	2,270 (917, 1,353)	Compliance indicator	0.13 (0.03, 0.24)	0.02

**Additionally, baseline EN score and the geographical/mastering numeracy stratification variable were included at both stages.*
CI=confidence interval; EN=Emerging Numeracy; PUMA=Progress in Understanding Mathematics Assessment.

Table 21: Analysis in the presence of non-compliance (pseudo-numeric compliance measure)

Model stage	Total n (intervention, control)	Predictor	Raw coefficients (95% CI)	P-value
Stage 1: Compliance indicator regressed on intervention status*	2,270 (917, 1,353)	Intervention status	8.13 (7.98, 8.27)	0.00
Stage 2: PUMA score regressed on compliance indicator*	2,270 (917, 1,353)	Compliance indicator	0.06 (0.01, 0.11)	0.02

**Additionally, baseline EN score and the geographical/mastering numeracy stratification variable were included at both stages.*
CI=confidence interval; EN=Emerging Numeracy; PUMA=Progress in Understanding Mathematics Assessment.

Missing data analysis

Table 22: Multilevel logistic regression where the outcome is whether a pupil is missing their New PUMA Total Score (1= 'yes', 0= 'no')

Variable	N (%) covariate missing	OR (95% CI)	P-value
Pupil level			
Baseline EN score	0 (0%)	0.95 (0.94, 0.97)	0.00
Stratification block: Region	0 (0%)	0.88 (0.47, 1.63)	0.68
Stratification block: MN participation	0 (0%)	0.87 (0.54, 1.39)	0.56
School level			
% FSM	118 (4%)	1.08 (0.88, 1.33)	0.47
Latest Ofsted rating	158 (6%)	1.10 (0.74, 1.62)	0.64
Urban or rural	38 (1%)	1.17 (0.57, 2.38)	0.67
Type of school	38 (1%)	1.24 (0.75, 2.06)	0.41

CI=confidence interval; EN=Emerging Numeracy; FSM=free school meals; MN=Mastering Number; Ofsted=Office for Standards in Education, Children's Services and Skills; OR=odds ratio.

Missing data analysis was completed to investigate whether there was any pattern to the missing data, which may require further sensitivity analyses. The only variable that was associated with missing outcomes was the baseline EN score. The results in Table 22 show that a lower EN score is related to missing the New PUMA outcome (p-value = 0.00). No further sensitivity analyses are needed as the EN baseline is included in the substantive model as a covariate.

Subgroup analyses

Two subgroup analyses were completed. The first looked at those pupils who were reported as being eligible for FSM. The second looked at those whose school had signed up to take part in the Mastering Number programme during the 2021/2022 academic year as it was felt there may be some interaction (positive or negative) between the two programmes. This data was collected prior to randomisation and used for the stratification. There is no impact of Reception Jigsaw on mathematical outcomes for either subgroup. Neither of the interaction models show different impacts of group allocation for the subgroups. It is worth noting that FSM rates for pupils in reception classes are systematically lower than for other year groups as this is the first year that pupils can claim this benefit[22] (National Statistics, 2022). This is a characteristic of the NPD dataset rather than a reflection of the true status of these pupils. The FSM subgroup analysis is underpowered (see Table 10) and therefore should be considered exploratory.

It is worth noting that schools could have signed up (or dropped out) of the Mastering Number programme after responding to email gathering data on Mastering Number participation. We have evidence that this did happen in some schools as we asked a question about Mastering Number participation in the endpoint survey. Of the 69 schools, which responded to this question in the endpoint survey, 28 schools had reported to having signed up to Mastering Number when asked at the beginning of the year and 41 schools had said either they did not know or that they had not signed up to Mastering Number. Of the 28 responding schools that had signed up for Mastering Number, four schools did not end up taking part in it. Of the 41 responding schools that did not state at the beginning of the academic year they were planning on taking part in Mastering Number, 13 schools reported that they had taken part in the programme when asked at the end of the year.

Table 23: The main analysis model (PUMA Total Score) restricted to particular subgroups of pupils

	Unadjusted means				Effect size			
	Intervention group		Control group					
Subgroup	N (missing)	Mean (95% CI)	n (missing)	Mean (95% CI)	Total n (intervention, control)	ICC (full)	Hedges' g (95% CI)	P-value
FSM-eligible pupils	216 (37)	11.9	321 (64)	12.2	537 (216, 321)	0.23	-0.02 (-0.2, 0.2)	0.9
Schools taking part in the Mastering Number programme	312 (54)	14.3 (13.7, 15.0)	486 (69)	14.2 (13.7, 14.7)	798 (312, 486)	0.27	0.0 (-0.3, 0.2)	0.9

CI=confidence interval; EN=Emerging Numeracy; FSM=free school meals; PUMA=Progress in Understanding Mathematics Assessment.

[22] As universal FSM are in place for this year group, the benefits of signing up for FSM are not immediate for the parents and the school is often the driving force in recruiting parents to apply for FSM, as pupils eligible for FSM attract higher per-pupil funding allocations for the school.

Longitudinal follow-up

Secondary outcome 5a (mathematics attainment at end of Year 1)[23]

There is a small difference between the means of the New PUMA (Year 1) score for the intervention and control groups (14.5 and 13.8, respectively). Hedges' g is positive at 0.09 (confidence interval [CI]: -0.05, 0.24) but the CI contains zero, with a p-value of 0.2. Pupils with teachers that had access to the Reception Jigsaw did make a small amount of additional progress in mathematics (estimated around one month) compared to pupils with teachers who continued with business as usual; however, the statistical evidence does not meet the threshold set by the evaluator to conclude that the true impact was non-zero. As shown in Figure 7 below, the distribution of the New PUMA (Year 1) total score is broadly Gaussian, with little difference in the distribution between the intervention and control groups.

Figure 7: Histogram of secondary outcome 5a (mathematics attainment at end of Year 1), split by grouping

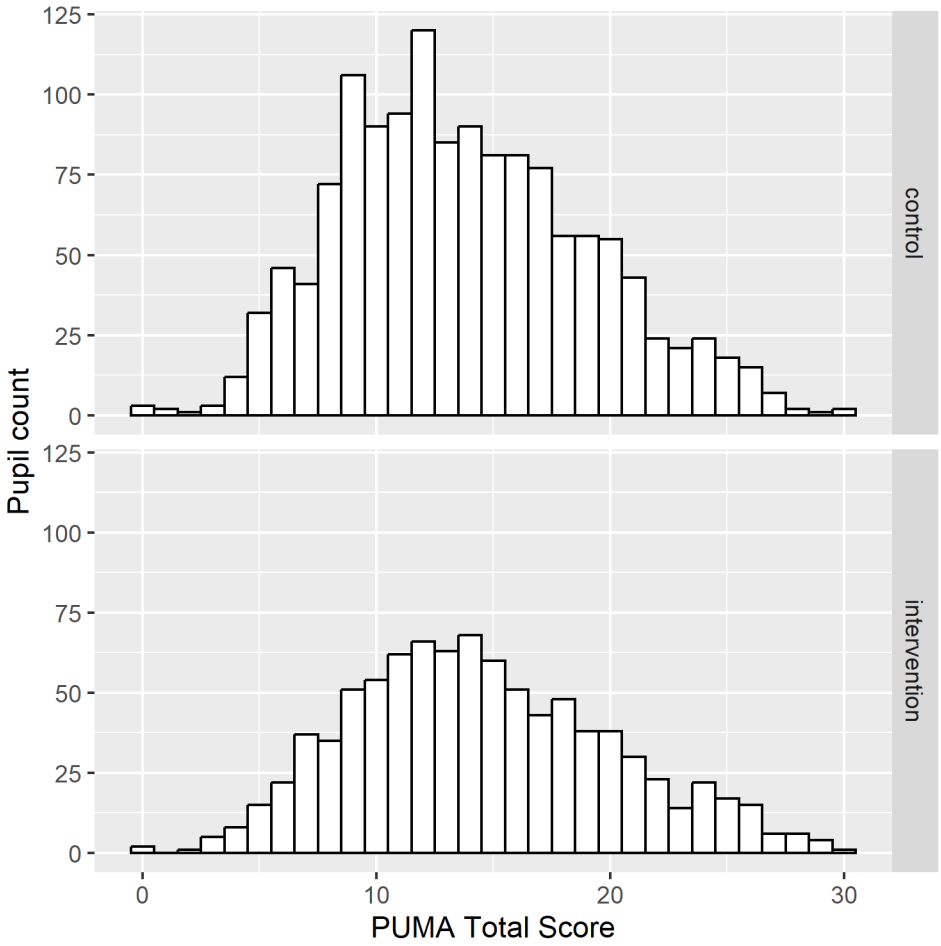


Table 24: Secondary analysis 5a (mathematics attainment at end of Year 1)

	Unadjusted means				Effect size			
	Intervention group		Control group					
Outcome	**N (missing)**	**Mean (95% CI)**	**n (missing)**	**Mean (95% CI)**	**Total n (intervention, control)**	**ICC (full)**	**Hedges' g (95% CI)**	**P-value**
PUMA (Year 1)	905	14.5	1360	13.8	2265	0.20	0.09	0.20

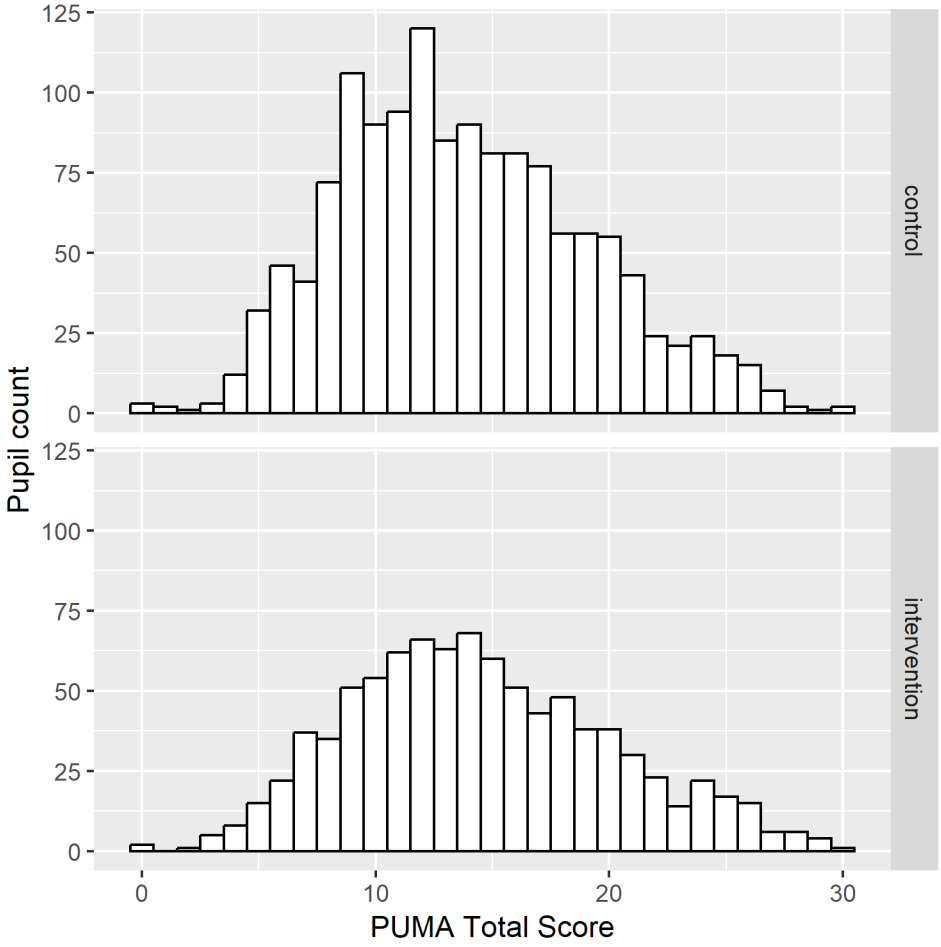
[23] This is the primary analysis in the longitudinal SAP (Schwendel, 2023).

Total Score	(171)	(14.1, 14.8)	(253)	(13.5, 14.0)	(905, 1360)		(-0.05, 0.24)	

Secondary outcome 5b (mathematics attainment at end of Year 1 taking into account further participation in Jigsaw)[24]

Further longitudinal analysis was run using the same model as for the main longitudinal analysis but with the additional of two dummy covariates which indicate whether a school has taken part in Reception Jigsaw or Primary Jigsaw in the academic year 2022/2023 (control schools) or Primary Jigsaw (intervention schools) in either 2021/22 or 2022/2023 outside of the trial. This was completed to take into account any confounding effects related to participating in other Jigsaw programmes. The results mirror the main longitudinal analysis with a small positive effect size. However, the statistical evidence does not meet the threshold set by the evaluator to conclude that the true impact was non-zero as the confidence interval includes zero.

Figure 8: Histogram of secondary outcome 5a (mathematics attainment at end of Year 1 taking into account further participation in Jigsaw), split by grouping



[24] This is the secondary analysis in the longitudinal SAP (Schendel, 2023).

Table 25: Secondary analysis 5b (mathematics attainment at end of Year 1 taking into account further participation in Jigsaw)

	Unadjusted means				Effect size			
	Intervention group		Control group					
Outcome	**N (missing)**	**Mean (95% CI)**	**n (missing)**	**Mean (95% CI)**	**Total n (intervention, control)**	**ICC (full)**	**Hedges' g (95% CI)**	**P-value**
PUMA (Year 1) Total Score	905 (171)	14.5 (14.1, 14.8)	1360 (253)	13.8 (13.5, 14.0)	2265 (905, 1360)	0.20	0.05 (-0.10, 0.21)	0.49

Analysis in the presence of non-compliance (longitudinal follow-up)

Mirroring the main phase, an analysis was completed on the longitudinal outcomes of pupils who were taught by a teacher who was considered to have complied with the programme requirements. Teachers were expected to have attended nine of the ten possible sessions (five twilight CPD sessions and five coaching sessions) to be judged as compliant to the programme. There is a positive effect size (0.17) for pupils who had been taught in their reception year by teachers who had attended at least nine out of ten Reception Jigsaw sessions. The confidence interval does not include zero so we conclude that there is enough statistical evidence to show that the Reception Jigsaw programme has a small beneficial effect on mathematical attainment among pupils whose teachers attended most of the training that is maintained into their second year at school. The findings using the pseudo-continuous compliance measure are presented in Table 27 and are in line with the binary measure.

Table 26: Analysis in the presence of non-compliance (binary compliance measure - longitudinal follow-up)

Model stage	**Total n (intervention, control)**	**Predictor**	**Standardised effect size (95% CI)**	**P-value**
Stage 1: Compliance indicator regressed on intervention status*	2212 (873, 1339)	Intervention status	1.46 (1.41, 1.52)	0.00
Stage 2: PUMA score regressed on compliance indicator*	2212 (873, 1339)	Compliance indicator	0.17 (0.06, 0.28)	0.00

**Additionally, baseline EN score and the geographical/mastering numeracy stratification variable were included at both stages.*

CI=confidence interval; EN=Emerging Numeracy; PUMA=Progress in Understanding Mathematics Assessment.

Table 27: Analysis in the presence of non-compliance (pseudo-numeric compliance measure - longitudinal follow-up)

Model stage	Total n (intervention, control)	Predictor	Raw coefficients (95% CI)	P-value
Stage 1: Compliance indicator regressed on intervention status*	2212 (873, 1339)	Intervention status	8.12 (7.97, 8.27)	0.00
Stage 2: PUMA score regressed on compliance indicator*	2212 (873, 1339)	Compliance indicator	0.07 (0.025, 0.12)	0.00

**Additionally, baseline EN score and the geographical/mastering numeracy stratification variable were included at both stages.*
CI=confidence interval; EN=Emerging Numeracy; PUMA=Progress in Understanding Mathematics Assessment.

Missing data analysis (longitudinal follow-up)

Missing case analysis was undertaken on the longitudinal data to check whether there was any pattern to the missingness. As with the main analysis, the only variable which associated with missing outcome values was the baseline assessment (the EN score). As this variable was included in all the models, no further analysis was needed.

Table 28: Multilevel logistic regression where the outcome is whether a pupil is missing their New PUMA Total Score (1= 'yes', 0= 'no')

Variable	N (%) covariate missing	OR (95% CI)	P-value
Pupil level			
Baseline EN score	0 (0%)	0.95 (0.94, 0.96)	0.00
Stratification block: Region	0 (0%)	0.86 (0.57, 1.28)	0.44
Stratification block: MN participation	0 (0%)	0.85 (0.63, 1.16)	0.30
School level			
% FSM	118 (4%)	1.14 (0.99, 1.31)	0.07
Latest Ofsted rating	158 (6%)	1.01 (0.78, 1.30)	0.94
Urban or rural	38 (1%)	0.10 (0.62, 1.61)	0.98
Type of school	38 (1%)	0.91 (0.66, 1.25)	0.56

CI=confidence interval; EN=Emerging Numeracy; FSM=free school meals; MN=Mastering Number; Ofsted=Office for Standards in Education, Children's Services and Skills; OR=odds ratio.

Subgroup analysis (mathematics attainment at end of Year 1)

As with the main analysis, two subgroup analyses were completed using the longitudinal follow-up assessment data (New PUMA (Year 1)). The first looked at those pupils who were reported as being eligible for FSM in reception. The second looked at those whose school had signed up to take part in the Mastering Number programme during the 2021/2022 academic year. This data was collected prior to randomisation and used for the stratification. There is no impact of Reception Jigsaw on mathematical outcomes for either subgroup. Neither of the interaction models show different impacts of group allocation for the subgroups (results not reported). The FSM subgroup analysis is underpowered (see Table 10) and therefore should be considered exploratory.

Table 29: The main analysis model (PUMA (Year 1) Total Score) restricted to particular subgroups of pupils

	Unadjusted means				Effect size			
	Intervention group		Control group					
Subgroup	**N (missing)**	**Mean (95% CI)**	**n (missing)**	**Mean (95% CI)**	**Total n (intervention, control)**	**ICC (full)**	**Hedges' g (95% CI)**	**P-value**
FSM-eligible pupils	197 (56)	11.9 (11.3, 12.6)	292 (93)	12.0 (11.5, 12.6)	489 (197, 292)	0.10	0.03 (-0.17, 0.22)	0.78
Schools taking part in the Mastering Number programme	305 (61)	13.8 (13.2, 14.4)	470 (85)	13.4 (13.2, 14.1)	775 (305, 470)	0.26	0.04 (-0.24, 0.32)	0.80

Implementation and process evaluation

The IPE findings are reported under the five IPE research questions using the relevant IPE dimensions. Mediators are discussed throughout.

1. Was the Reception Jigsaw delivered as intended in terms of dosage, nature, and quality?

Fidelity

To ascertain fidelity, the evaluation team carried out: observations of all five 'train the trainer' sessions delivered by WRM; observed the first CPD session and fifth half-day coaching visit in the four longitudinal case study schools; interviewed the Reception Jigsaw developer team; and analysed findings from the trainer reflection logs. The findings from these observations, the interview, and analysis of reflection logs, suggest that trainers delivered the Reception Jigsaw with high fidelity to the recommended approach.

To what extent does the delivery of the Reception Jigsaw training sessions adhere to the intended approach?

The Reception Jigsaw CPD sessions are highly manualised. Trainers receive a set of PowerPoint presentations from which they deliver the content and showcase examples of Reception Jigsaw activities in practice. In addition, teachers are provided with the same set of resources, which support the practical elements of the sessions (e.g. Numicon and counters), and reception teachers receive a separate set of resources to keep, which they can use in their classroom to deliver the activities. Having observed the 'train the trainer' sessions, in which trainers receive comprehensive training on the CPD session content by going through the training presentations slide by slide, the evaluation team were able to see how trainers' delivery of the first CPD session in case study schools mirrored that of the developer in the training. They also picked up on the level of consistency between trainers. The developer team reported that they had carried out their own internal quality assurance of trainers' delivery, observing one of each trainer's CPD sessions. The developer team reported being very happy with the quality and consistency of delivery across the nine trainers.

How consistent is the delivery of Reception Jigsaw training across the schools?

The CPD sessions allow for 'adaptation with fidelity', meaning that trainers can place greater emphasis on some elements of the content over others depending upon teachers' and schools' needs, and to make the content relevant for different staff and year groups. Responsiveness to a mixed audience within schools was clear to the evaluation team during their observations. One trainer provided examples to Key Stage 2 teachers of how to make activities more challenging, and suggested ways of making activities more age-appropriate for older children who are not yet secure in their early mathematics skills. In all CPD sessions the evaluation team observed, trainers also provided ideas for adapting activities for children with special educational needs and disability (SEND) and for higher ability children who required an additional level of challenge. All these adaptations to the CPD sessions are permissible.

The evaluation team observed minor differences in trainers' delivery. For example, one trainer dedicated more time to teachers engaging with the practical activities than other trainers, while another took more time to explain in greater depth the concepts presented in the training. One trainer included more research while presenting the content (additional to that already included in the presentation) while another drew more upon their own experiences of teaching reception when explaining concepts and activities to teachers. These differences are likely to be the result of trainers' levels of experience delivering CPD and their recency in the classroom themselves. These differences in emphasis did not affect the fidelity with which the training was delivered in the case study schools.

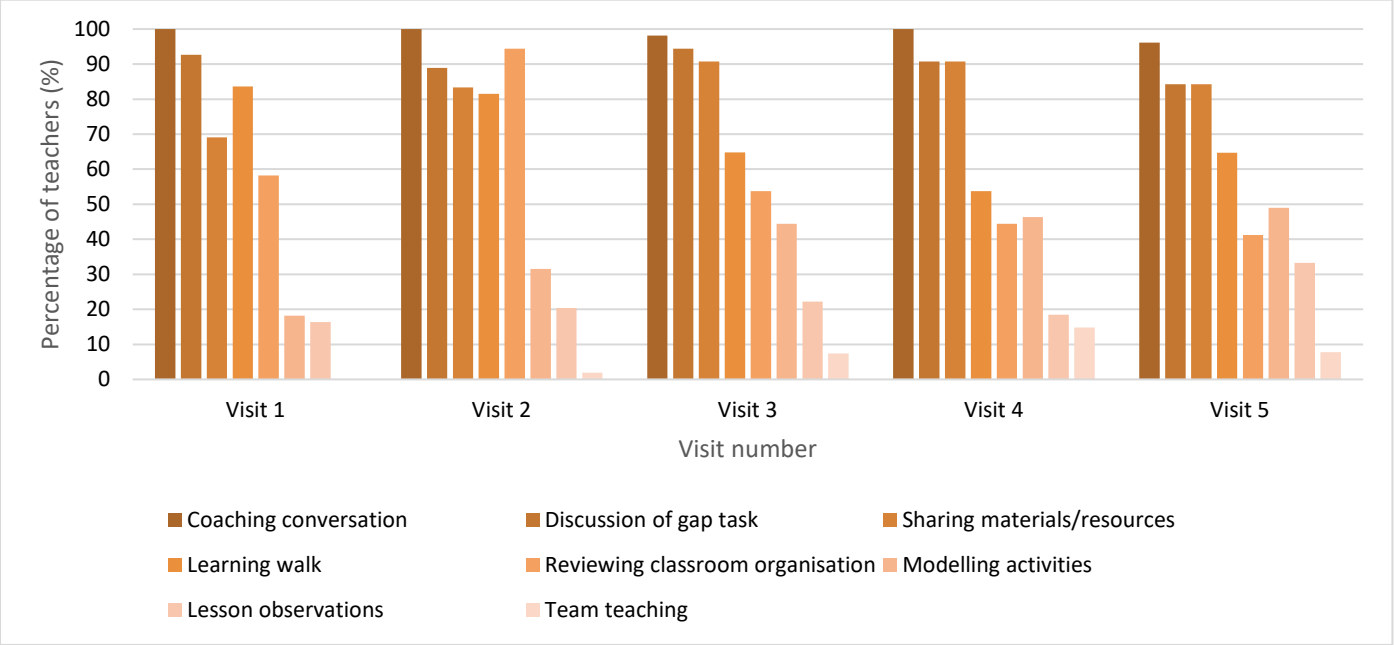
To what extent does delivery of the half-day coaching visits adhere to the intended approach?

Compared to the CPD sessions, delivery of support at the half-day coaching visits is bespoke, tailored to meet the needs of the reception teachers and their provision, while also accounting for the context in which they are working. Coaching visits included but were not limited to: trainers observing classroom practice and the provision to identify areas for improvement; reviewing the gap task; providing support with planning; and supporting teachers to plan for sustaining the learning and impacts for the future.

Analysis of the trainer reflection logs gave the evaluation team an insight into the types of support trainers provided to reception teachers during each of the five visits. Figure 9 illustrates the percentage of teachers who received each type of support during each of the half-day coaching visits, as recorded by the trainers in the reflection log. Coaching

conversations and a discussion of the gap task[25] set at the last CPD session featured at the majority of coaching visits. Across all visits, team teaching was the least common form of support.

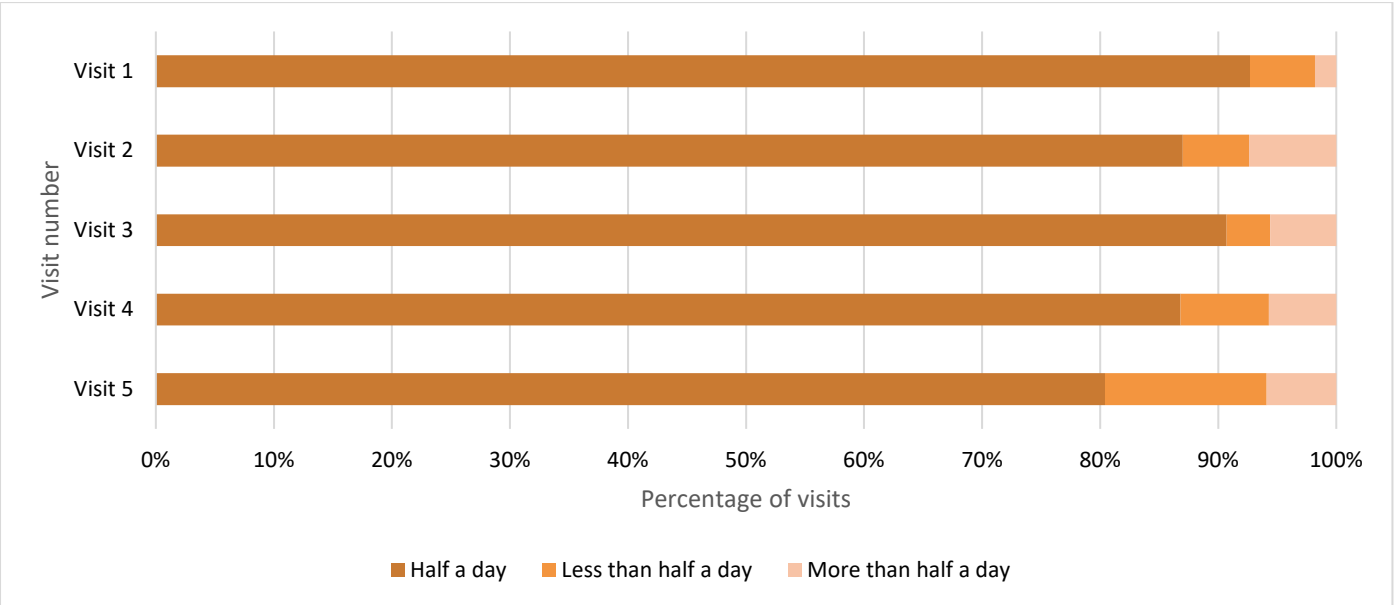
Figure 9: Elements of support delivered at the half-day coaching visits



Notes: Data from trainer reflection logs: Which of the following support elements did you provide during school visit… (School visit 1 N=55; school visit 2, 3, 4 N=54; school visit 5 N=51). Items from the trainer reflection log reordered to be presented in descending order.

The evaluation team asked trainers to record the duration of each of the visits in the reflection logs. Most visits lasted half a day, suggesting high fidelity to the intended approach. Across all visits, only a small proportion were less than or more than half a day. Figure 10 illustrates these findings.

Figure 10: Duration of half-day coaching visits



Notes: Data from the trainer reflection logs: How long was the school visit…? (School visit 1 N=55; school visit 2, 3, 4 N=54; school visit 5 N=51).

[25] Gap tasks were set for reception teachers. They ensured teachers implemented a new activity in either their teaching or provision, related to the CPD content.

Data from the trainer reflection logs show that one school only received coaching visit one, and three additional schools did not receive visit five. The developer explained that the school who only received coaching visit one was one of the hardest schools to engage in the Reception Jigsaw. Due to several postponements in three schools, trainers combined visit four and five for these schools so that all visits were completed before pupils completed the endpoint assessment.

Given the level of flexibility with the coaching visits, the evaluation team asked the developer team of any non-permissible adaptations to this element of the support. They reported that in one school, reception teachers did not have time to engage with the coaching visit and instead, the school wanted this support to be directed at the Year 1 teacher and their classroom. Due to the focus of the Reception Jigsaw being reception teachers and pupils, this adaptation could not be allowed.

Delivery of training and coaching at a consistent and similar quality is a mediator in the logic model. Several factors contributed to the successful delivery with fidelity of the twilight sessions and coaching visits. The developer team reported that the trainers were consistent throughout, with only a few instances of the team having to cover one another due to illness. All the trainers were experienced in teaching early years mathematics. Of the nine trainers, four trainers were experienced in delivering mathematics CPD, one trainer had one year's experience and four trainers were new. Other factors include the quality assurance carried out by the developer team (mentioned above) and the comprehensive training and support trainers received (see 'Quality' section below).

How was delivery adapted and impacted by Covid-19?

Adaptations to delivery resulting from the Covid-19 situation are discussed below in the 'Adaptations' section of this report. However, any adaptations did not affect the fidelity with which the programme was delivered.

Adaptations

What adaptations were made to the Reception Jigsaw training delivery model?

The developers planned to recruit a team of trainers from across Yorkshire and Essex so that trainers would deliver Reception Jigsaw within their region. However, the recruitment advert for trainers in Essex was released the week before the country entered a national lockdown due to Covid-19, and the developers did not recruit any Essex-based trainers. Instead, six of the nine Yorkshire-based trainers travelled to Essex to deliver the Reception Jigsaw to these schools.

This led to differences in the delivery model between the two regions. Intervention schools in Yorkshire had a CPD session every month and their coaching visit would fall in the middle of these, around two weeks after their CPD session. Essex schools still had a CPD session every month; however, due to the distance trainers had to travel, they received the coaching visit linked to the previous CPD the day before the next CPD session. The developer reported that trainers added in a phone call to Essex schools halfway between each CPD/visit to ensure they still had regular communication.

Did trainers adapt their delivery of the Reception Jigsaw CPD sessions and/or half-day coaching visits?

The evaluation team asked the developers what, if any, adaptations trainers had made to their delivery of the Reception Jigsaw CPD sessions and/or coaching visits. As discussed above in the 'Fidelity' section of this report, their responses suggest high fidelity to the intended training approach and only a few, permissible adaptations were made.

The developers reported that trainers had to adapt their delivery of some of the Reception Jigsaw CPD sessions to accommodate the social distancing restrictions in place in schools due to the Covid-19 pandemic.[26] Several trainers delivered hybrid sessions, where they delivered in person to reception teachers and reception TAs, and other staff accessed the session via Zoom. The developer team reported that of all the CPD sessions, only two took place entirely virtually, via Zoom, because the school could not accommodate the trainers. Many schools had to postpone CPD sessions and coaching visits, which delayed delivery of the Reception Jigsaw programme. In several schools, trainers combined the later coaching visits to ensure schools received all the support before the summer holidays. The developer reported that one Essex school, where engagement with the Reception Jigsaw was low, received four of the five CPD sessions due to repeated cancellations.

[26] Delivery of the Reception Jigsaw began in November 2021, when schools (and the rest of England) were experiencing another rise in Covid-19 cases due to the new Omicron variant. In schools, teachers and pupils had to remain within year-group 'bubbles'.

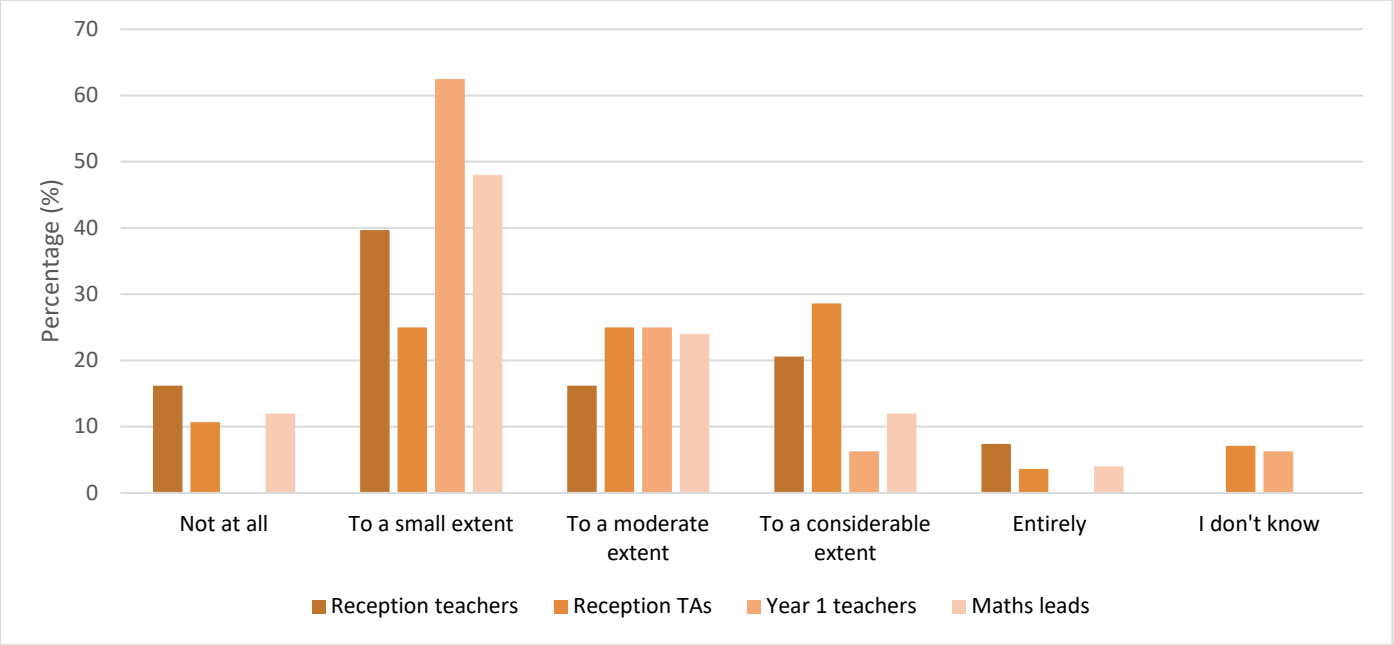
The evaluation team asked the developers if trainers made any adaptations to the content of the CPD sessions. As delivery of the Reception Jigsaw progressed and reception teachers incorporated activities from the CPD sessions into their practice, trainers photographed these to share with other teachers. The developers added these photographs into the CPD session presentations. They reported that teachers found it helpful to see the activities in action because it gave them ideas for how they could recreate it in their provision, which supported their delivery.

During the CPD session observations in case study schools, the evaluation team identified that there was evidence of trainers 'adapting with fidelity' (i.e. making adaptations allowed by the programme), giving greater emphasis to areas of the content where teachers required more support, or making suggestions for how to adapt activities to suit older children or children with additional needs. The observations did not identify any other adaptations from the recommended approach.

What adaptations are made during delivery by teachers, and why?

The Reception Jigsaw CPD sessions provide teachers with ideas they are encouraged to adapt to meet the needs of their pupils and their classroom. The endpoint surveys in Summer Term 2022 asked reception teachers, reception TAs, Year 1 teachers, and mathematics leads to comment on the extent to which they had adapted the Reception Jigsaw. As shown in Figure 11 below, most respondents reported adapting the strategies. Across reception teachers, Year 1 teachers and mathematics leads, the largest proportions reported adapting the strategies 'to a small extent'. The largest proportion of reception TAs adapted the strategies 'to a considerable extent'.

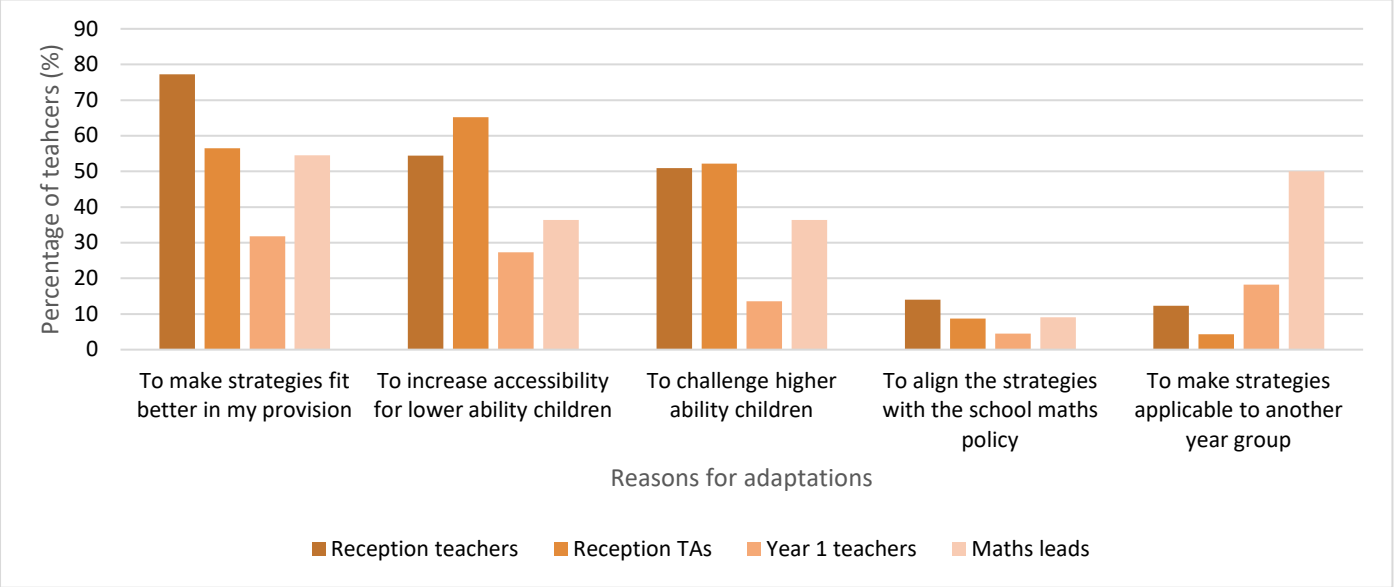
Figure 11: Extent to which teachers adapted the Reception Jigsaw



Notes: Data from the endpoint survey: To what extent have you adapted the Reception Jigsaw strategies from what the trainer presented? (Reception teacher N=68; reception TA N=28; Year 1 teacher N=16; mathematics lead N=25). Percentages have been rounded so may not sum to 100. TA=teaching assistant.

The survey asked respondents to comment on why they adapted the Reception Jigsaw. Figure 12 below details the percentage of respondents who reported each reason for adapting the programme. The largest proportions of reception teachers, Year 1 teachers, and mathematics leads reported adapting the Reception Jigsaw to better fit their provision. This can be expected because trainers encouraged teachers to adapt the activities and strategies to suit their pupils and their classroom. It can also be expected that around half of mathematics leads reported adapting strategies for another year group. The developer team said that many mathematics leads taught in Key Stage 2, meaning activities would require adaptation to suit older children with more advanced mathematics skills. Around two-thirds of reception TAs reported adapting the Reception Jigsaw to increase accessibility for lower ability children. This finding may be explained by the fact that often, TAs work with small groups of pupils who are not working at age-related expectations to provide additional support.

Figure 12: Reasons why teachers adapted the Reception Jigsaw



Notes: Data from the endpoint survey: Why did you adapt the Reception Jigsaw strategies? (Reception teacher N=57; reception TA N=23; Year 1 teacher N=15; mathematics lead N=22). Items reordered to present most frequently to least frequently reported adaptation by reception teachers. TA=teaching assistant.

During the best practice and longitudinal case study interviews with reception and Year 1 teachers, the evaluation team asked teachers whether they had adapted the Reception Jigsaw and if so, their reason for this. Their responses echo findings from the survey. Several reception teachers and Year 1 teachers reported adapting some of the strategies to pitch them at a lower level for children working below age-related expectations and for children with SEND who required additional support.

> *The Year 1 teachers discussed as a group how the activities could be differentiated. The ideas are solid and you just might need to use different numbers or make other links to what they have learning in Year 1, but all the tasks and ideas are useable and adaptable…A lot of the time I'd be thinking of how the activities would work with particular children in mind because I have children with EHCPs [Education Health and Care Plans] and their level of understanding isn't as solid.* (Year 1 teacher, best practice case study school)

Teachers also reported adapting activities for higher ability children who required extension tasks, for example using larger numbers, posing trickier calculations, or asking them to make more complicated patterns. One reception teacher commented that despite making some adaptations to suit pupils' abilities, overall, the Reception Jigsaw programme was suitable for all children. She said:

> *It's really inclusive. With the number sense, it doesn't exclude anyone who might be less able and for those who are more confident. [The trainer] always gives us opportunities to extend them and dig deeper.* (Reception teacher, longitudinal case study school)

These findings suggest that, other than permissible adaptations, teachers implemented the Reception Jigsaw programme with high fidelity.

2. How well did the participants (teachers/support staff from reception and Key Stage 1, mathematics coordinator, senior leaders, and then also pupils) engage with the Reception Jigsaw? Were there any implementation challenges faced? If so, what were they and to what extent were they overcome?

Compliance

The Reception Jigsaw training programme consisted of five two-hour CPD sessions and five half-day coaching visits, delivered between November 2021 and June 2022. Compliance to the intervention is indicated by reception teachers participating in at least nine of the ten sessions. Reception Jigsaw trainers completed attendance registers, which were used to measure how many of the sessions teachers, TAs, mathematics leads, and senior leaders participated in. A

review of the attendance logs showed that across the 55 schools, trainers logged attendance against the names of 92 reception teachers whose pupils went on to complete the New PUMA assessment (this does not include EYFS leads). Of the 92, 54 reception teachers (59%) from 44 schools complied (i.e. attended nine out of the ten sessions).[27] Attendance to the CPD sessions was higher than for the coaching visits with 77 reception teachers attending at least four CPD sessions (84%) and 54 reception teachers attending at least four coaching visits (58%). This disparity is partly driven by schools where there was more than one reception teacher. In these schools a lower percentage of reception teachers attended at least four coaching visits compared to schools with only one reception teacher. Compliance at the school level was low, largely due to low attendance by mathematics leads at the half-day coaching visits. The reasons for this are further discussed below.

Attendance at CPD sessions

A review of the attendance logs and interviews with the developer team demonstrated that Reception Jigsaw CPD sessions were generally well attended by reception teachers, mathematics leads, and Year 1 teachers. Attendance of reception TAs, other staff, and senior leaders was more variable as would be expected due to attendance being recommended rather than required for these practitioners. As reported by the developers and noted by trainers in the logs, gaps in attendance of reception teachers and mathematics leads tended to be due to staff turnover or absence (e.g. due to maternity leave). Where there was turnover of reception teachers, it is clear from the attendance logs that the replacement teacher picked up the Reception Jigsaw training where their predecessor left, suggesting that pupils continued to be taught by a Reception Jigsaw trained teacher (although this scenario does not translate into compliance). Attendance of reception teachers at all the CPD twilight sessions is a mediator in the logic model. Participation in the programme by school mathematics leads is also highlighted as a mediator in the logic model.

The developer team reported that where reception teachers joined intervention schools part way through Reception Jigsaw delivery, trainers spent time during the coaching visits revisiting key pedagogy and messages they had missed from earlier sessions. The developers said this was essential for teachers to implement gap tasks with their pupils. However, one trainer reported that missing the earlier sessions would have had a detrimental impact on teachers' understanding and delivery. They reported that CPD session one and two are: *'extremely important in terms of subject knowledge and pedagogy'* (Developer team). Teachers also missed the discussions that took place in the sessions relating to implementation. The developers reported that revisiting the content in this way during the visit could not be as thorough or as effective as attending the session in full. This suggests that although trainers did their best to support teachers who had missed sessions, their overall understanding of the Reception Jigsaw and therefore delivery in the classroom would still have been compromised.

Interviews with participants in case study schools provide further insight into these findings. Almost all teachers highly regarded the CPD sessions, which can explain the levels of attendance. They reported that even when teachers could not attend in person, for example, when at home self-isolating with Covid-19, they still made the time to attend the sessions virtually. Reception TAs could not always attend CPD sessions because they were after school, but senior leaders and mathematics leads from all six case study schools reported that their schools offered to pay reception TAs to facilitate their attendance. Further details on teachers' response to the Reception Jigsaw can be found below in the 'Quality' and 'Responsiveness' sections of this report.

Attendance to coaching visits

A review of the attendance logs and interviews with the developer team suggest that attendance to the half-day coaching visits was highest among reception teachers and variable among mathematics leads. The interview with the developer team provides further insight into attendance at the coaching visits. They too reported that while attendance of reception teachers was good, many mathematics leads struggled to attend all of the coaching visits. Many mathematics leads taught in upper Key Stage 2 and securing cover and release time when faced with capacity pressures due to the Covid-19 pandemic was challenging. The developer team felt that, upon reflection, mathematics leads' attendance at the visits was not essential and, in the future, they would not be required to attend. The developers reported that visits had still been successful where trainers had only met with reception teachers and had then fed back to mathematics leads about what the visit had involved, either at the CPD sessions or via email.

[27] This percentage is slightly different from the compliance percentage reported in the impact section (65%). This is primarily because this measure was calculated at teacher level and the impact compliance measure was calculated at the pupil level.

These findings demonstrate that at the reception teacher level, although just less than two-thirds of reception teachers complied with the intervention (impacted by absences and staff turnover), reception teachers highly valued the support they received from the CPD sessions and coaching visits they attended. School-level compliance was impacted by low attendance of mathematics leads to the coaching visits, however comments from the developers suggest that this did not impact the implementation of the intervention.

Dosage

To what extent do reception teachers and reception TAs apply learning from the Reception Jigsaw?

The Reception Jigsaw logic model expects that the support provided will lead to changes to the reception learning environment (output), changes to pedagogy (output), and that teachers and TAs will apply learning in the classroom (mediator). Findings from the endpoint survey, trainer reflection logs, and case study interviews suggest that most reception teachers did apply learning from the Reception Jigsaw, embedding activities into their mathematics teaching and continuous provision and adapting the classroom environment.

The endpoint survey in Summer Term 2022 asked reception teachers and reception TAs in the intervention group how often they applied learning from the Reception Jigsaw in their mathematics teaching. All reception teachers who responded applied learning at least once a week, around two-thirds (68%) reported applying it in four to five mathematics lessons per week, and around one-third (31%) applied learning in two to three lessons per week.

Findings from the endpoint survey suggest that application of the Reception Jigsaw was slightly less frequent among reception TAs compared to reception teachers, however the majority reported that they applied learning at least once a week. Similar proportions applied it in four to five mathematics lessons per week (30%) and in two to three mathematics lessons per week (37%).

Interviews with seven reception teachers in the six longitudinal and best practice case study schools provide further insight into teachers' use of the Reception Jigsaw. The first interviews with teachers in the four longitudinal case study schools took place after their first CPD session in Autumn Term 2021. The evaluation team asked teachers how they thought they may be able to implement Reception Jigsaw in their classroom. Teachers reported that the CPD sessions provided them with lots of practical ideas, which they were able to easily incorporate into their delivery straight after the CPD session.

> *The ideas, the resources and how to use them is really great, I can see how we can use it and I've already put some of it into next week's planning, I got it written down last night while it was still fresh.* (Reception teacher, longitudinal case study school)

In the final interviews in Summer Term 2022, all reception teachers (from the longitudinal and best practice case study schools) reported that they used the Reception Jigsaw strategies and activities everyday as part of their direct teaching time, in continuous provision and in everyday routines:

> *All of the activities taught in the training have been implemented in maths lessons and the skills transferred into provision when children are working independently.* (Reception teacher, best practice case study school)

The programme had become an embedded part of their practice, which they felt had enhanced their lessons and mathematics provision. As discussed below in the 'Quality' section of this report, support from the Reception Jigsaw trainer during the half-day coaching visits had led to teachers making changes to their provision. The most significant change, as reflected in the following quote, was the inclusion of mathematics activities throughout all areas of the provision:

> *[The trainer] helped us to incorporate maths into all areas of the provision rather than just the maths corner so we're encouraging mathematical thinking throughout the provision.* (Reception teacher, best practice case study school)

This had enabled teachers to apply Reception Jigsaw activities to other subjects, and activities in the school day such as registration, increasing the dosage children had of Reception Jigsaw. One reception teacher commented:

> *The strategies are very easily implemented, they don't take much planning. The only things that have taken a bit of time has been some of the classroom environment changes, we've just done these when*

> *we've had a bit of time, whereas the practice related strategies are easily incorporated into our day-to-day practice.* (Reception teacher, longitudinal case study school)

The evaluation team asked trainers to complete a reflection log after each half-day coaching visit. Trainers reported that most teachers (between 87% and 93%) applied learning 'to some extent' or 'to a large extent' in the classroom after each CPD session.

To what extent do Year 1 teachers apply learning from the Reception Jigsaw?

The Reception Jigsaw logic model expects that the support provided through the Reception Jigsaw will lead to: Key Stage 1 practitioners having a better understanding of children's EYFS mathematics development (output); changes to mathematics provision (output); and that Year 1 teachers and Year 1 TAs will build on reception mathematics teaching and apply learning in their classrooms (mediator). Findings from the endpoint survey and case study interviews suggest that Year 1 teachers did apply learning from the Reception Jigsaw; however, they did so less frequently, and with less ease, compared to reception teachers.

The endpoint survey in Summer Term 2022 asked Year 1 teachers in the intervention group how often they applied learning from the Reception Jigsaw in their mathematics teaching. The majority of teachers applied it in at least one lesson per week, 12% applied it in four to five mathematics lessons per week, just over half (53%) applied it in two to three mathematics lessons per week, and just less than a fifth (18%) in one lesson per week.

Interviews with four Year 1 teachers in four of the six case study schools (longitudinal and best practice) provide further insight on teachers' use of the Reception Jigsaw. All four Year 1 teachers reported implementing Reception Jigsaw activities into their mathematics teaching practice but to varying extents. Some reported that it had been difficult to implement all the activities and suggestions from the CPD sessions because they are following the national curriculum, which is prescriptive and delivered through structured mathematics lessons, and they do not always have continuous provision[28] like reception classes do. Despite this, they had adapted some activities so they could be incorporated into mathematics lessons and support specific groups of pupils, as discussed below in the 'Reach' section of this report. One senior leader commented:

> *In the training, there were elements that were beyond the expectations of EYFS so they were able to take that on board. They've used the practical elements within their planned curriculum to develop the children's knowledge and making sure it's secure.* (Deputy Headteacher, longitudinal case study school)

Why did teachers not apply learning from the Reception Jigsaw?

Only a very small proportion of reception TAs (7%) reported through the endpoint survey that they never used the Reception Jigsaw strategies. They went onto explain that they did not deliver the taught mathematics input but supported pupils with accessing and completing mathematics activities during their time in provision. A larger proportion of mathematics leads (19%) reported never applying learning from the Reception Jigsaw. This finding can be expected because many mathematics leads taught in Key Stage 2, so the activities and strategies the CPD sessions recommended were not relevant or appropriate for their pupils without adaptation.

Reach

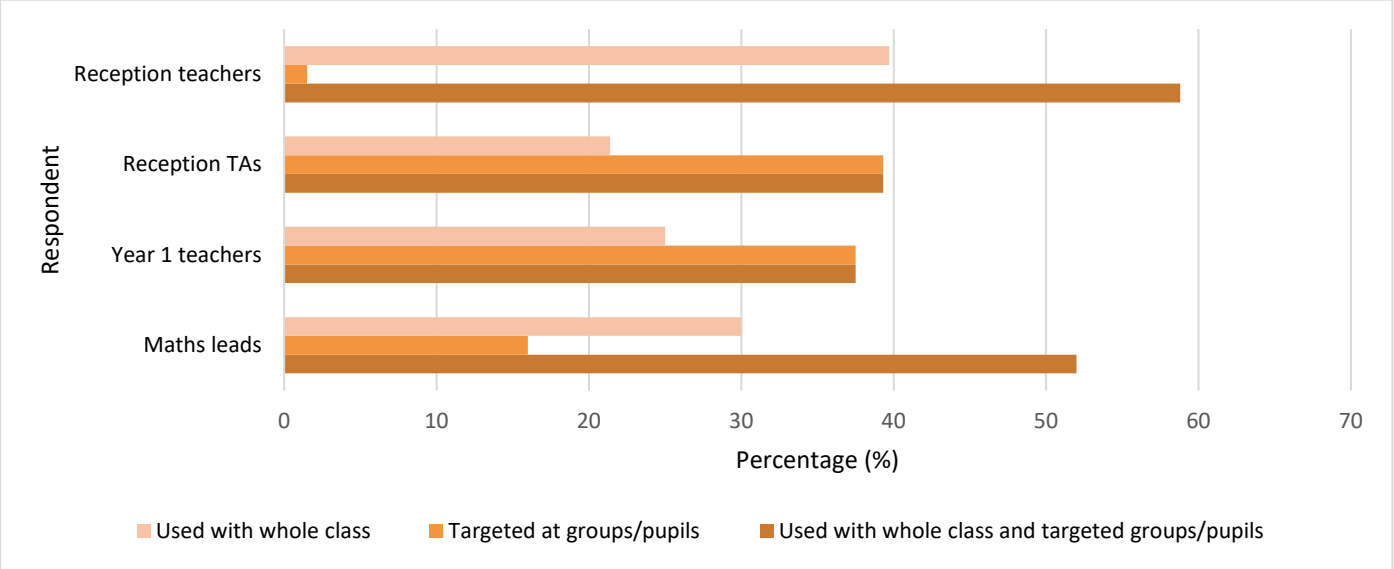
Do teachers deliver Reception Jigsaw to the whole class, or is it used as a targeted strategy?

The endpoint survey in Summer Term 2022 asked reception teachers, reception TAs, Year 1 teachers, and mathematics leads what their approach was to delivering Reception Jigsaw with their pupils. The survey asked respondents if: they only used the strategies with their whole class; they only targeted the Reception Jigsaw at specific groups/pupils; or if they used the Reception Jigsaw with their whole class and with targeted groups/pupils. Most respondents used the Reception Jigsaw with their whole class. Over half of reception teachers and mathematics leads reported also targeting specific groups/pupils, compared to just nearly two-fifths of reception TAs and Year 1 teachers. Compared to reception teachers and mathematics leads, larger proportions of reception TAs and Year 1 teachers reported only using the Reception Jigsaw with specific groups/pupils. These findings likely reflect the common working practices of TAs to work

[28] Areas in the classroom which allow children to learn through exploration which concrete resources. This is a key feature of reception classrooms.

with small groups of pupils who need additional support. As shown in Figure 14, Year 1 teachers were more likely to target pupils working below age-related expectations or with SEND. The findings can be seen in Figure 13 below.

Figure 13: Teachers', TAs', and mathematics leads' approach to delivering Reception Jigsaw

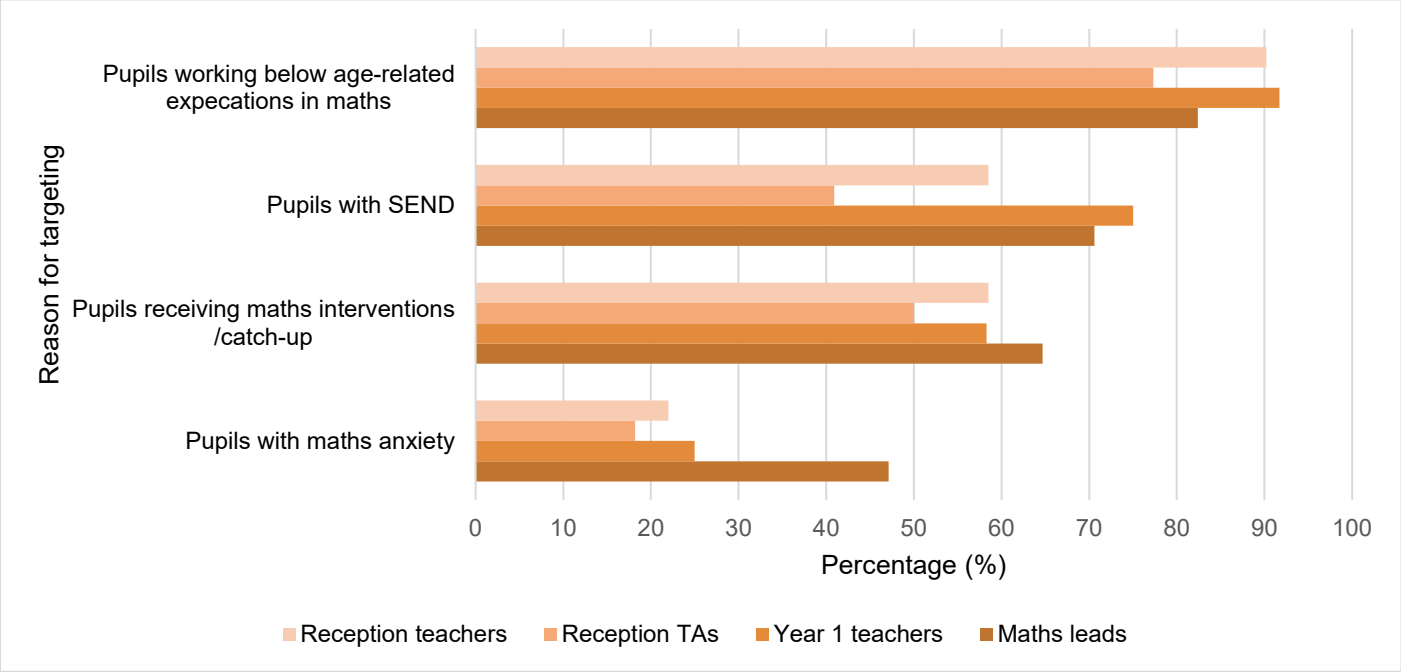


Notes: Data from the endpoint survey: What was your approach to delivering the Reception Jigsaw with pupils in your class? (Reception teacher N=68; reception TA N=28; Year 1 teacher N=16; mathematics lead N=25). Percentages have been rounded so many not sum to 100.
TA=teaching assistant.

Which groups/pupils are targeted for Reception Jigsaw?

The survey asked the respondents who said they targeted the Reception Jigsaw to identify, which specific groups/pupils they targeted for activities. As can be seen in Figure 14 below, across all respondent groups, the most common reason for targeting Reception Jigsaw was to support pupils working below age-related expectations. Around three-quarters of Year 1 teachers and mathematics leads targeted Reception Jigsaw at pupils with SEND, compared to around half of reception teachers and reception TAs. Between one-half and two-thirds of respondents reported targeting Reception Jigsaw at pupils receiving mathematics interventions or catch-up support. Across all respondent groups, mathematics leads were most likely to target Reception Jigsaw at pupils with mathematics anxiety (almost half, compared to a quarter or less of reception teachers, reception TAs, and Year 1 teachers).

Figure 14: Reasons for targeting Reception Jigsaw at specific groups/pupils



Notes: Data from the endpoint survey: Which pupils did you use the targeted strategies with? (Reception teacher N=41; reception TA N=22; Year 1 teacher N=15; mathematics lead N=17). Percentages have been rounded so many not sum to 100.

SEND=special educational needs and disability; TA=teaching assistant.

The interviews with teachers in the longitudinal and best practice case study schools provide further insight into teachers' approach to delivering the Reception Jigsaw. The evaluation team asked reception teachers and Year 1 teachers to comment upon whether more children had received Reception Jigsaw than others and if so, which children these were. These findings echo those from the survey. Reception teachers in both longitudinal and best practice case study schools took a mixed approach to delivering Reception Jigsaw in their classrooms. They reported that for the most part, all children received equal exposure to the Reception Jigsaw; however, there were instances of targeting activities, for example, at children who would not independently access mathematics activities or at pupils identified as struggling with a mathematics concept.

All four Year 1 teachers involved in the interviews in Summer Term 2022 reported that they had targeted Reception Jigsaw activities at pupils who had gaps in their early years mathematics knowledge and skills. They used the activities to secure children's understanding and ensure these skills were strong before moving onto Year 1 level tasks. When possible, some also worked Reception Jigsaw activities into their planning and delivered to their whole class during mathematics lessons and (where applicable) in their provision.

Did other year groups in intervention schools use learning from the Reception Jigsaw?

Through the endpoint survey, the evaluation team asked mathematics leads if other classes had implemented learning from the Reception Jigsaw. Over three-quarters of mathematics leads (77%) reported that learning had also been implemented in Key Stage 1 classes, but only 16% reported that learning had been implemented in Key Stage 2 classes. It can be expected that a high proportion of Key Stage 1 classes implemented the learning because Year 1 teachers were expected to attend the CPD sessions. Year 2 teachers were also invited to attend although their attendance was optional and the CPD sessions were not aimed at Key Stage 2 teachers. The interviews with mathematics leads and senior leaders in longitudinal and best practice case study schools provide further insight into these findings. Mathematics leads reported that naturally there is sharing of information and learning from CPD sessions within key stages, so Year 1 teachers had disseminated, either informally through conversations or more formally through key stage meetings, the activities and strategies suggested in the CPD sessions to their Year 2 colleagues.

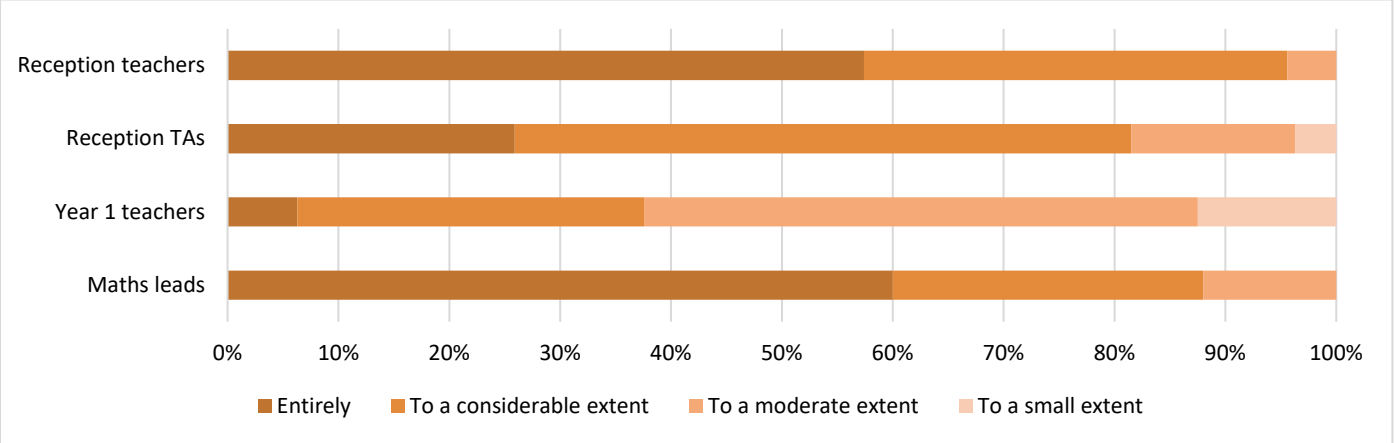
Responsiveness

How do school staff respond to the Reception Jigsaw programme?

Data from the endpoint survey, trainer reflection logs, and interviews with teachers and the developer team provided evidence that teachers' response to the Reception Jigsaw programme was positive.

The endpoint survey asked reception teachers, reception TAs, Year 1 teachers, and mathematics leads to comment on the extent to which they enjoyed delivering Reception Jigsaw in their mathematics teaching. Across all groups of respondents, data suggests that reception teachers enjoyed delivering the Reception Jigsaw most and Year 1 teachers enjoyed delivering the programme the least (Figure 15).

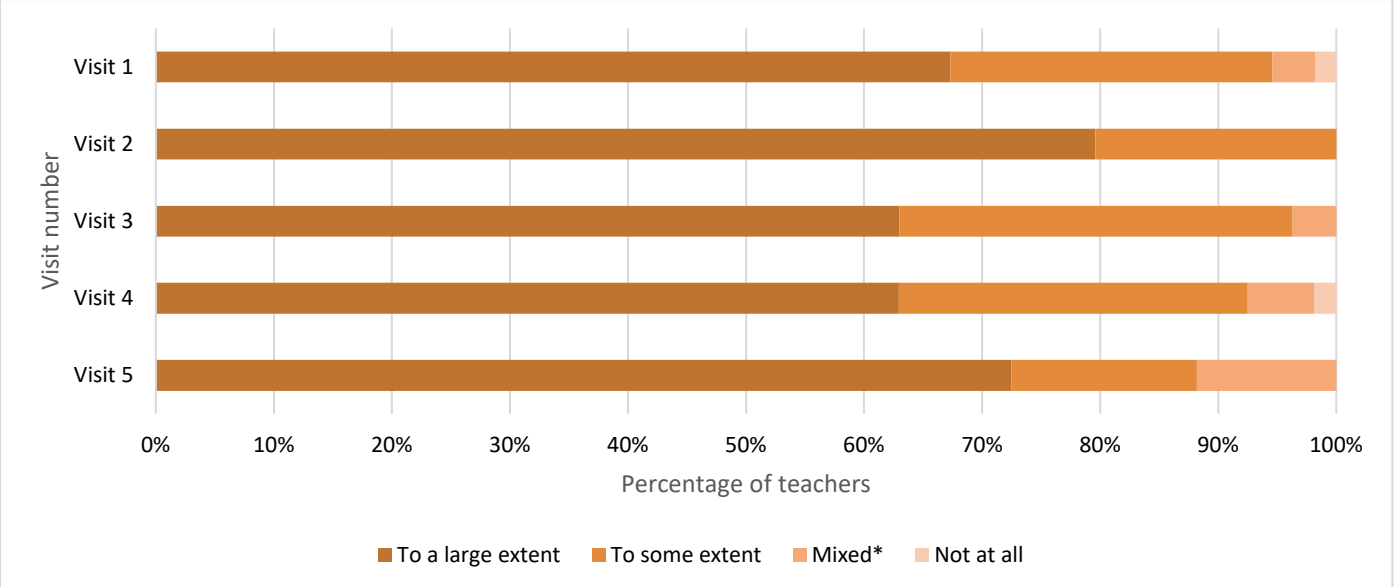
Figure 15: Extent to which teachers enjoyed delivering Reception Jigsaw



Notes: Data from the endpoint survey: To what extent have you enjoyed delivering Reception Jigsaw in your mathematics teaching? (Reception teacher N=68; reception TA N=27; Year 1 teacher N=16; mathematics lead N=25). Percentages have been rounded so many not sum to 100. TA=teaching assistant.

After each half-day coaching visit, Reception Jigsaw trainers indicated, through the reflection log, the extent to which reception teachers were engaged in the programme. They also commented on the extent to which reception teachers were engaged with the content of the most recent CPD session. Analysis of this data shows that the majority of reception teachers were highly engaged with both the programme and the CPD content. These findings are illustrated in Figures 16 and 17 below.

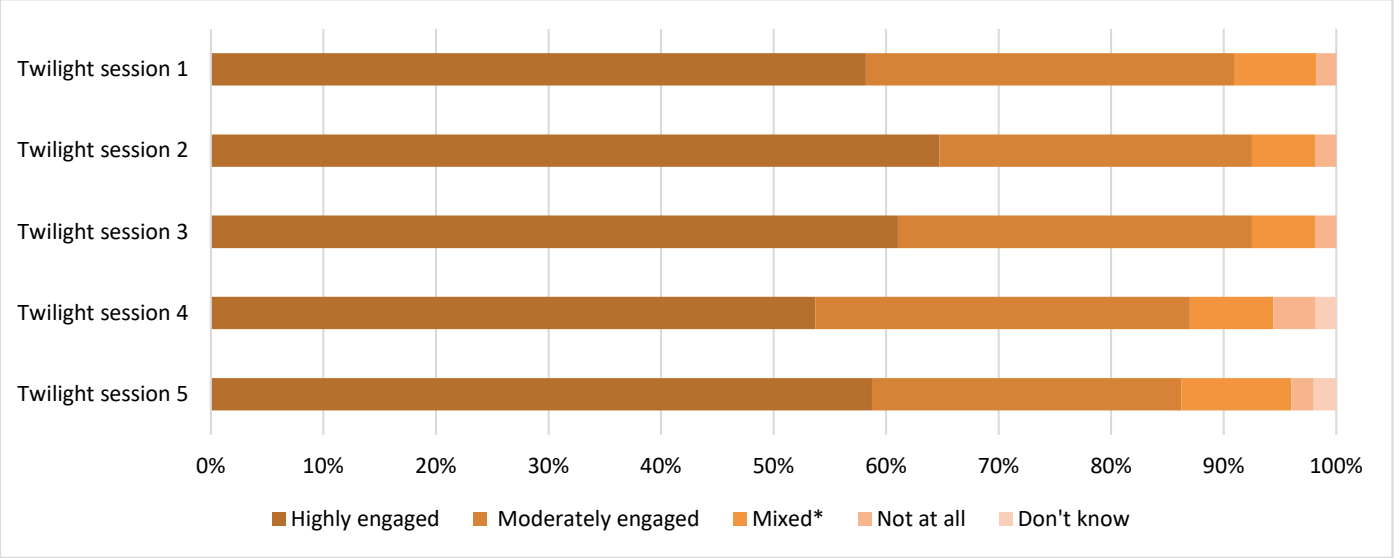
Figure 16: Reception teachers' engagement in the Reception Jigsaw programme over the year



Notes: Data from the trainer reflection logs: To what extent is/are the reception teacher(s) engaged in the Reception Jigsaw programme? (School visit 1 N=55; school visit 2, 3, 4 N=54; school visit 5 N=51).

** In schools where more than one reception teacher participated in the programme and there was variation in their engagement/involvement, trainers selected this response option.*

Figure 17: Reception teachers' engagement with CPD session content



Notes: Data from the trainer reflection logs: How engaged is/are the reception teacher(s) with the content of CPD session…? (CPD session 1 N=55; CPD session 2, 3, 4 N=54; CPD session 5 N=51).

** In schools where more than one reception teacher participated in the programme and there was variation in their engagement/involvement, trainers selected this response option.*

CPD=Continuing Professional Development.

The interviews with teachers in case study schools offered further insight into these findings.

Elements of the Reception Jigsaw that teachers and pupils responded well to

Teachers, mathematics leads, and senior leaders in case study schools responded positively to the Reception Jigsaw programme. The elements of the Reception Jigsaw that worked well, and therefore teachers responded well to, can be found below in the 'Quality' section of this report.

Elements of the Reception Jigsaw that teachers and pupils responded less well to

The majority of teachers had not experienced anything that worked less well with the programme, and most did not report anything that they or their pupils responded less well to. Discussion of the elements of the Reception Jigsaw that teachers report worked less well, which teachers also responded less well to, can be found below in the 'Quality' section of this report.

What challenges had teachers faced with implementing the Reception Jigsaw?

Very few teachers in both longitudinal and best practice case study schools reported challenges with implementing the Reception Jigsaw. Challenges were primarily related to implementing the Reception Jigsaw in Year 1. Two Year 1 teachers in the longitudinal case study schools reported not receiving the same level of support as their reception colleagues, for example they did not receive support from the trainer during the half-day coaching visits. They also did not receive the same resources as reception teachers, which made implementation of some of the activities modelled in the CPD sessions challenging. However, as acknowledged by the developer team, the Reception Jigsaw programme was targeted at reception teachers and pupils, and it was not expected that Year 1 teachers would engage with the coaching visits.

Year 1 teachers in best practice case study schools highlighted different challenges with implementation. One teacher reported challenges with the logistics of teaching in a more practical way while also ensuring there is a record and evidence of what pupils have achieved. However, this challenge had been overcome because their senior leaders saw the benefit for pupils of moving away from exercise books and replacing these with floor books, so had been supported with making this change. The Year 1 teacher in the other best practice school reported that the main challenge with implementation of the Reception Jigsaw had been the timing of the training. They reported that they would have benefited from many of the ideas earlier on in the academic year when they revisited topics with pupils, but they reported that the activities had still been helpful and they would benefit even more from these next year.

To what extent did teachers apply learning from the Reception Jigsaw?

Details on the extent to which teachers applied learning from the Reception Jigsaw in their mathematics teaching can be found above in the 'Dosage' section of this report.

Did teachers complete the gap tasks?

At the end of each CPD session, trainers set participants a gap task (an activity related to or exampled in the session content) to complete. Completion of the gap task was an essential element of the training for reception teachers and optional for all other participants.

The developer team believed that the gap tasks were an essential part of the Reception Jigsaw model and considered a mediator in the logic model. They reported that, because they link to the CPD session content, they provide an opportunity for teachers to put into practice something they have learned during the session. They reported that the gap tasks provided a useful first focus for the half-day coaching visits, regardless of the extent to which teachers had completed this. For example, in cases where teachers had not had time to put this in place, the trainer supported them to do this. The developer team reported:

> *The visits helped with the engagement in a lot of schools because of the time frame. If they hadn't had chance to do the gap task, the visit became an opportunity to do it with them, some needed more support. If there wasn't this support in place, in some schools, the gap task would never have been done so the impact of the training wouldn't have been as successful.* (Developer team)

The trainer reflection log asked Reception Jigsaw trainers to record if reception teachers had completed the gap task. Across all five visits, around half of trainers reported that teachers had completed the gap task and around a third reported they had 'partially' completed it. Smaller proportions of trainers reported that completion was 'mixed' (i.e. where there was more than one reception teacher in the school and there was variation in their completion) or that teachers had not completed the gap task at all.

All reception teachers in all six case study schools (longitudinal and best practice) reported completing the gap tasks. Reception teachers reported that they were a useful element of the training because they encouraged them to think about an area of their provision they could improve through implementing something new from the training: *'they are key to making sure you continue to think about the training'* (Reception teacher, best practice case study school). They also appreciated the chance to reflect on the task, such as how pupils had responded and the impact it had, at the next half-day coaching visit with the trainer.

What additional support is required to engage with the Reception Jigsaw programme?

The Reception Jigsaw logic model states that participation of the mathematics lead and support from the school senior leadership team are critical to successful implementation of the programme. Data from the endpoint surveys and interviews with teachers in case study schools suggest that teachers mostly received the necessary support for their school's mathematics lead and senior leadership to implement the programme.

Support from mathematics leads

The endpoint survey asked reception teachers, reception TAs, and Year 1 teachers to indicate the extent to which they agreed they felt supported by the school's mathematics lead to apply learning from the Reception Jigsaw. Overall, around three-quarters strongly agreed or agreed. Around two-fifths of reception teachers (44.1%) and reception TAs (41%) agreed that they felt supported by the mathematics leads to apply learning from the Reception Jigsaw. Around a quarter of reception teachers (28%) and a third of reception TAs (35%) strongly agreed and around

a quarter neither agreed nor disagreed with the statement. Just less than three-quarters (71%) of Year 1 teachers agreed with this statement. The same proportion strongly agreed (6%) and strongly disagreed (6%).

Box 1: What support had mathematics leads provided in best practice case study schools?

The evaluation team asked interviewees in the best practice case study schools what support mathematics leads had provided to teachers to support implementation. Mathematics leads and senior leaders reported that it was important teachers felt the changes were being made 'with them' rather than 'to them'. Mathematics leads had attended all the CPD sessions and visits, so felt well placed to support teachers with implementation.

Mathematics leads had participated in reception and Key Stage 1 planning meetings to support teachers with how they would implement the Reception Jigsaw into their mathematics teaching. In reception, mathematics leads had supported with making changes to the timetable to increase the time pupils spent in provision, and in both year groups, mathematics leads reviewed schemes of work and lesson designs to incorporate the Reception Jigsaw. In the classroom, support had included observations and team teaching, which informed the practical guidance mathematics leads could provide teachers to strengthen their delivery.

Support from senior leaders

As is noted in the logic model, the developers reported that support from senior leads had been key to successful implementation (mediator). They said:

> *The schools who go[t] the most from it, were the most engaged and the most successful were those with SLT [senior leadership team] buy-in, where SLT came to the sessions and where staff were released [for the half-day coaching visits]. Most staff were released for every visit and this had a massive impact.* (Developer team)

In some schools, the Reception Jigsaw trainer had helped gain SLT buy-in to the programme through raising their awareness of the importance of early years mathematics and sharing with them the rationale for implementing changes in the provision and curriculum.

> *Having a professional who can help and support as an expert has really helped, someone who can go and speak to the headteacher.* (Developer team)

The endpoint survey asked reception teachers, reception TAs, and Year 1 teachers to indicate the extent to which they agreed they felt support by senior leaders in their school to apply learning from the Reception Jigsaw. Overall, around three-quarters strongly agreed or agreed. Just over half of reception teachers (54%) and just over two-fifths of reception TAs (45%) agreed that they felt supported by senior leaders. Around a quarter of reception teachers (27%) and a third of reception TAs (31%) strongly agreed. Just less than three-quarters (71%) of Year 1 teachers agreed with this statement. The same proportion (6%) strongly agreed and strongly disagreed.

Mathematics leads and senior leaders indicated the extent to which they agreed that senior leaders in their school were supportive of staff applying learning from the Reception Jigsaw. Around two-thirds (65%) of mathematics leads strongly agreed and one-third (32%) agreed. In comparison, 80% of senior leaders strongly agreed that they were supportive of staff applying learning from the Reception Jigsaw, with a small proportion (17%) agreeing.

The endpoint survey also asked mathematics leads and senior leaders to comment on the extent to which they agreed, with the statement: 'senior leaders in my school are committed to sustaining the changes to teaching mathematics resulting from engaging with the Reception Jigsaw'. Around two-thirds (61%) of mathematics leads strongly agreed and one-third (32%) agreed. In comparison, 80% of senior leaders strongly agreed that they were supportive of sustaining the changes brought about, with a small proportion (17%) agreeing.

The interviews with staff in longitudinal and best practice case study schools provide further insights into the support senior leaders had provided to implement the programme. Similar support was provided across all six case study schools (longitudinal and best practice).

As mentioned above, teachers reported that senior leaders were keen for the Reception Jigsaw to succeed in their schools. To ensure this, senior leaders themselves reported that they had made Reception Jigsaw a school priority;

provided funding to engage with the programme, purchased resources recommended by the training, and paid reception TAs to attend; and provided cover for reception teachers and mathematics leads to fully engage with the coaching visits and had given teachers the flexibility to implement the programme. Teachers felt senior leaders had trusted them with implementing the programme in the way that worked for their pupils and had regularly checked in on the impacts they were seeing.

> *We are quite innovative with everything we do. We want to try new things and get the best out of the children so they [senior leaders] are happy for us to change the curriculum in the way we need to.* (Reception teacher, best practice case study school)

Box 2: How did best practice schools' CPD culture, and participation in other CPD, support implementation of the Reception Jigsaw and the impacts that teachers reported?

In both best practice case study schools, CPD was highly valued, with senior leaders encouraging teachers to source and attend CPD. They supported this through providing them with the funding and release time to participate.

> *SLT are very supportive of staff training and school improvement generally. I joined here three years and I've had lots of CPD since I joined…Anytime you need any help, we are supported to go and get training and cover is provided. It's a real strength here.* (Year 1 teacher, best practice case study school)

Senior leaders gave teachers the freedom to implement new approaches they gained from training, which allowed teachers to implement the Reception Jigsaw in a way that worked for them and their pupils.

Both best practice case study schools participated in the Teaching for Mastery Teacher Research Groups (TRGs), run by the West Yorkshire Maths Hub. They had embedded the mastery approach in their teaching of Year 1 to Year 6 maths and wanted to participate in Reception Jigsaw to embed the mastery approach into reception. The mathematics lead in one school felt that having a strong, established mastery approach across the school had supported implementation of the Reception Jigsaw because reception teachers were already aware of the principles of the mastery approach. They also reported that this would mean teachers in other years would be able to implement Reception Jigsaw principles with ease. Commenting on the links between Mastering Number, Teaching for Mastery, and Reception Jigsaw, they said: *'They are all really strong, interlinked programmes which is making for strong maths teaching.'* (Mathematics lead, best practice case study school).

One school had replicated the TRGs internally to support their reception teachers to implement the Reception Jigsaw. Senior leaders provided the time for teachers to meet for the TRG, as well as the time and cover for teachers to be released to observe each other's practice. At each TRG, teachers discussed how to implement strategies from the CPD session into the practice then following observations, identified good practice, and areas for development. The Reception Jigsaw trainer had supported teachers to plan out their TRGs and link Reception Jigsaw objectives with other early years targets, namely language, reading, and comprehension. From each TRG, teachers produced an action plan, which set out what they would achieve ahead of the next CPD session.

Headteachers encouraged reception TAs to participate in CPD (the impacts of which are discussed below in the 'Quality' section of this report), which fed into their decision to fund reception TAs time for the CPD sessions:

> *We invest in TAs too because they are valuable to the school but if you're not careful, you don't get value for money so it's important to keep developing them.* (Headteacher, best practice case study school)

What are participants' views about the sustainability of the Reception Jigsaw? What factors will facilitate or hinder the sustainability of the intervention?

As part of the interviews in Summer Term 2022, the evaluation team asked the staff in longitudinal and best practice case study schools whether they would continue to implement Reception Jigsaw in the future. Having seen the benefits and impacts of the programme for reception pupils this year, all six case study schools (longitudinal and best practice)

planned to continue using Reception Jigsaw. Reception teachers felt the approach was now embedded into their mathematics curriculum.

> *It's been the most enjoyable, informative training I've done in a very long time because it's been so practical, so well delivered, instantly achievable to put into the classroom and into your practice without it costing too much time, too much planning. We've been able to seamlessly weave it in.* (Reception teacher, longitudinal case study school)

The evaluation team asked staff what factors might facilitate or hinder them to continue implementing the Reception Jigsaw programme. Teachers said having the budget to buy the books and additional resources recommended by the CPD training would support continued implementation but recognised the constraints of school funding, which meant this could also hinder future implementation.

Teachers reported that movement of Reception Jigsaw trained staff out of EYFS and Key Stage 1 could hinder continued implementation of the programme in those year groups, but would support the natural cascading of the programme to other areas of the school. The mathematics lead in one of the best practice case study schools reported that their mathematics team is made up of a group of teachers from across the phases, so this will help with consistent and continued delivery should staff leave as the knowledge of Reception Jigsaw will continue to be shared. Mathematics leads reported that incorporating Reception Jigsaw into their school's mathematics CPD programme would facilitate the continued implementation of the programme in both in reception classes and across the school because it would ensure that new staff, both to the school and to EYFS, are competent in delivering mathematics lessons in line with the programme.

> *I have everything in a file so when a new teacher comes in, I can share this with them. Hopefully it will be more settled in September and I can sit down with everyone and share the training the same way [the trainer] did, put expectations in around how it is used and that we want to see the strategies being used.* (Mathematics lead, longitudinal case study school)

Along with delivering Reception Jigsaw through CPD sessions to support new and Key Stage 2 staff, interviewees reported that they would continue to revisit Reception Jigsaw action plans and the content of the CPD sessions to keep the intervention at the forefront of teachers' and TAs' minds.

> *Refreshers will be key, whether that is through the maths lead or White Rose Maths. It does slip if you don't get the refreshers of the ideas. I'm hoping through the maths team we will be able to do that.* (Year 1 teacher, best practice case study school)

3. Was the quality of training, support, and intervention materials provided by the developer adequate? Was preparedness and confidence of staff delivering the intervention at the right level? If not, why?

Quality

This section includes the findings from intervention group teachers and the developers on their perceptions of the quality and effectiveness of the Reception Jigsaw programme. The impacts that teachers recognised from implementing the programme for themselves, their pupils, and their school are also considered.

Quality of the Reception Jigsaw trainers

Teachers were very satisfied with the quality of delivery from the Reception Jigsaw trainers. The endpoint survey asked reception teachers, reception TAs, Year 1 teachers, mathematics leads, and headteachers to indicate the extent to which they agreed ('strongly agree' or 'agree') with the statement: 'the delivery of the CPD training sessions was of a consistently high quality'. Almost all reception teachers (96%), reception TAs (97%), mathematics leads (100%), headteachers (95%), and Year 1 teachers (94%) strongly agreed or agreed with this statement.

Satisfaction with the Reception Jigsaw trainers and their delivery was very high in case study schools. Teachers described their trainers as *'dedicated', 'enthusiastic', 'inspirational',* and *'professional'*. They commented on trainers' knowledge and experience with teaching early years mathematics, which they valued because it allowed trainers to draw upon their own experiences and meant they could relate to teachers.

> *The success of it all is that [the trainer] knows the EYFS and has worked as an EYFS teacher—it's not someone who has come in with a new idea without implementing it. [The trainer] has the experience so can relate it to classroom practice—everyone has appreciated that.* (Mathematics lead, longitudinal case study school)

Teachers in case study schools also appreciated the level of support they received from their trainers at the CPD sessions and coaching visits and their responsiveness to emails between sessions when teachers required ad-hoc support. They reported that trainers were able to provide solutions and 'quick fixes' to enhance their provision and did so in a supportive manner, which did not make teachers feel under pressure or scrutiny.

Quality of the Reception Jigsaw content

All teachers in case study schools felt the content of the Reception Jigsaw training was high quality. They liked that it provided them with lots of useful practical ideas, which they could take away and implement straight away, and which all children could engage with. Teachers also appreciated the links to research and the EEF early maths guidance report (EEF, 2021). The quotes below, from participants in a range of roles, demonstrate teachers' high regard for the training content:

> *I thought it was fantastic training, it was really well planned, the content was good. I always like it when people put practical suggestions in of what you can go away and use.* (Reception teacher, longitudinal case study school)

> *I was really impressed. It was very professionally delivered, delivered to a really high standard, I liked that it was backed by all of the research, there were links to this. The resources are high-quality. The content was really meaningful and taught to a high standard...It was really engaging and useful and it's really going to help develop our mathematics teaching so I'm really pleased so far.* (Year 1 teacher, longitudinal case study school)

> *It has been on[e of] the best pieces of training for staff that they could have had.* (Deputy headteacher, longitudinal case study school)

Relevance of the Reception Jigsaw content

The endpoint survey asked reception teachers, reception TAs, Year 1 teachers, mathematics leads, and headteachers to indicate the extent to which they agreed ('strongly agree' or 'agree') with the statement: 'the content presented in the CPD training sessions was relevant to me in my role'. Agreement was highest among reception teachers (93%), reception TAs (90%), and mathematics leads (84%). Around three-quarters of Year 1 teachers (77%) and headteachers (71%) agreed that the content was relevant to them.

As reflected in the survey findings, despite recognising that the training was high quality, some Year 1 teachers in case study schools found the content less relevant to their teaching practice. For example, continuous provision did not feature in all Year 1 classrooms, which they felt limited the extent to which they could implement the activities demonstrated in the CPD sessions. The focus on reception children and the development of their early mathematics skills also led them to feel the training was not always relevant to them. This focus was however made clear at the start of the training and, as reported by the developer and observed by the evaluation team, trainers ensured the content was as relevant as possible to Year 1 teachers through providing examples of how to make activities more challenging and through suggesting how they may incorporate the Reception Jigsaw into their practice.

Quality of the Reception Jigsaw resources

Teachers were very satisfied with the quality of the resources they received with the Reception Jigsaw training (such as the dot plates and picture cards received at the first twilight session). Almost all reception teachers (99%), all reception TAs (100%), and mathematics leads (100%) agreed or strongly agreed that the materials provided in the CPD sessions were of high quality. Almost all Year 1 teachers (94%) and headteachers (90%) also reported that they strongly agreed or agreed the materials were of high quality.

Similarly, to the survey findings, all reception teachers in case study schools were very satisfied with the quality of resources they received. Teachers commented that it was a welcome change to have training, which came with the resources demonstrated in the training content. Teachers commented that the resources were particularly useful

because of their versatility, they could be used during direct teaching and pupils could use them independently during provision time. They also reported that resources were inclusive because they were suitable for pupils with SEND and for pupils with English as an additional language. Additional sets of resources were available to buy but there was a lack of awareness from some Year 1 teachers that this was the case.

Elements of the Reception Jigsaw CPD sessions that worked well and less well

The evaluation team asked teachers in case study schools to identify elements of the CPD sessions that had worked well. Teachers in longitudinal and best practice case study schools highlighted the same elements of the training that had worked well and ultimately supported them to implement the programme.

Most teachers commented on the practical nature of the sessions, which they saw as a key element to successful implementation. They liked being provided first with a range of practical ideas and concrete examples followed by the opportunity to 'have a go' at the activities themselves. They commented that being able to try out the activities themselves in the sessions built their confidence with how to implement the Reception Jigsaw in their classroom, for example it gave teachers the opportunity to think about the questions to ask pupils to extend their learning. It also gave teachers an indication of how their pupils would explore and respond to the activities, as demonstrated in the following quote from a reception teacher:

> *All of the ideas are great, but when you actually get to do it yourself, you can see where the children might have problems with stuff, or see if your children will like that game, because you know your class and what they like really well, so that's been really useful.'* (Reception teacher, longitudinal case study school)

Mathematics leads and senior leaders praised the inclusive and open nature of the training, which had allowed a range of staff working in different year groups and at different levels to attend. They commented that it was an opportunity for staff to work with colleagues outside of their year group or phase teams and had initiated professional discussions around mathematics teaching. One senior leader reported that the practical activities had been a catalyst for these discussions:

> *The practical activities have initiated discussions between staff, between reception, Year 1 and the maths team that you don't always get the opportunity for. They've discussed maths, shared ideas then been able to implement in the classroom.* (Deputy headteacher, longitudinal case study school)

Teachers welcomed the option to tailor activities based on their pupils and provision, and the support the trainer provided during the CPD sessions to do this. One mathematics lead commented on the flexibility of the trainer's approach:

> *Every bit of knowledge that is shared with us, there's always a chance to discuss it specifically in terms of how it would work for us as a school or how things have worked with the gap tasks. That's been really positive, she [the trainer] has a very open approach.* (Mathematics lead, longitudinal case study school)

Teachers liked the design and organisation of the training and the ongoing cycle of CPD sessions and coaching visits that occurred over the year. They compared this favourably to one-off CPD sessions. The longitudinal nature of the Reception Jigsaw had allowed teachers to build on knowledge from session to session and had given them time to implement in small steps, with support in between through the coaching visits. They said the order of delivery was key for ensuring pupils' knowledge was secure and built upon as the programme progressed. One reception teacher said:

> *Quality leads to quality. The premise of the Reception Jigsaw was proper research, it was well thought through, well designed, well-resourced and the support was tailored to our school.* (Reception teacher, longitudinal case study school)

What are teachers' suggestions for improving the CPD sessions?

Only a very small number of teachers in case study schools identified improvements to the CPD sessions. Two teachers and one mathematics lead commented on the length of CPD sessions (two hours) and felt that reducing this to an hour and a half would be preferable. One teacher also commented that there was a lot of content to cover during the sessions and suggested that a lot of the time spent trailing the activities and resources could be reduced. As acknowledged by

the trainers, teachers in one school reported that the CPD sessions being delivered via Zoom did not work as well as in person.

Quality of half-day coaching visits

The endpoint survey asked reception teachers and mathematics leads to comment on the quality of the half-day coaching visits. Most reception teachers (92%) and mathematics leads (87%) strongly agreed or agreed that the support provided during the visits was of a consistently high quality. Although high, compared to satisfaction with the CPD sessions, these findings suggest teachers and mathematics leads were slightly less satisfied with the coaching visits.

All reception teachers and mathematics leads in case study schools were very complementary of the support the trainer provided them during the half-day coaching visits. They reported that this tailored support had been helpful, leading to improvements in the provision, and the trainer was realistic about what was achievable in their settings, not making teachers feel overwhelmed or criticised. One mathematics lead commented:

> *[The trainer] understands that the provision in our school is already well established and strong in EYFS so [the trainer] listened to us and is going to help us build on other areas.* (Mathematics lead, longitudinal case study school)

In one case study school, the reception teacher was an Early Career Teacher (ECT) in a one-form entry school, and thus did not have the support of a partner teacher. She reported that the trainer had acted like a mentor, giving advice and guidance to implement new activities and strategies. The trainer had helped her with adapting the activities demonstrated in the CPD sessions for her pupils and context, and they helped her see how to increase the opportunities for mathematics learning in the provision, a skill which was still developing as an ECT.

Elements of the Reception Jigsaw half-day coaching visits that worked well and less well

The evaluation team asked teachers in case study schools to identify elements of the half-day coaching visits that had worked well. Reception teachers reported that the personalised, one to one support had been invaluable. Given this, the support teachers received and what worked well differed across schools, but key elements teachers picked out were: the practical advice on planning and sharing of resources; the 'quick fixes', which were easily achievable to improve the provision; and the opportunity for teachers to reflect on their practice with an early years mathematics specialist.

> *[The trainer] has always given me the opportunity to discuss the maths that I feel the children needed more support with or they didn't quite get and helped me think of solutions and ways to adapt what I'm doing next to make sure that gap isn't left.'* (Reception teacher, longitudinal case study school)

Teachers in both longitudinal and best practice case study schools, as well as the developer, reported that the support from the trainer during the half-day coaching visits to reorganise the reception classroom and to incorporate mathematics activities into all areas of provision, had been key to successful implementation of the Reception Jigsaw. Trainers had provided teachers with an expert, fresh pair of eyes, which helped them reflect on how they could enhance their provision to maximise mathematics learning opportunities for pupils. The developer team reported:

> *We've done a lot of supporting teachers with classroom provision—this has been a bigger part of the visits than we anticipated. So many schools have completely changed and reorganised their provision. Trainers have rolled their sleeves up and helped teachers get this sorted.* (Developer team)

What are teachers' suggestions for improving the half-day coaching visits?

The only things that worked less well in case study schools were individual school-based challenges, including teachers having to use their planning, preparation, and assessment (PPA) time for the visits and being unable to take on some suggestions due to school policies. Teachers and mathematics leads did not have any suggestions for improving the half-day coaching visits.

What training do trainers receive and how well does this prepare them to deliver the CPD sessions and half-day coaching visits to teachers?

The evaluation team asked the developers what training and support the Reception Jigsaw trainers received. This information complemented the evaluation team's observations of the five 'train the trainer' sessions. They reported that trainers had an intensive package of support. Trainers saw each CPD session twice before delivering to schools. The developer went through the session presentations slide by slide, modelling how trainers would be expected to deliver the content. Trainers had the opportunity to try out the activities presented in the sessions in groups/pairs in the same way teachers would in the sessions when delivered in schools. The lead trainer also modelled how trainers may provide suggestions for adapting the content for children in older year groups, higher ability children in reception, and pupils with SEND. These train the trainer sessions provided trainers with an example of 'best practice' delivery, as well as the opportunity to discuss and reflect on the content together.

Trainers also received support to deliver the half-day coaching visits. The developers reported that the range of experience across the team differed, as did the experience of the teachers they would be working with in schools. To ensure all staff were at the same level, this training consisted of: a broad review of early years content and pedagogy; the history of early years; and what good practice looks like. The developers compiled a list of questions (and barriers) trainers could expect, prepared them for the support and challenge aspect of the visits, and modelled how to pitch suggestions to teachers, mathematics leads, and headteachers to ensure the visits were a positive and supportive experience for schools.

The developer team reported that through the recruitment process and training sessions, they had achieved the level of skills and experience expected of trainers. This is a mediating factor in the logic model.

Following the evaluation team's observations of CPD session one in the four longitudinal case study schools, it was concluded that trainers were well prepared to deliver the sessions. Their delivery was professional and enthusiastic, demonstrating their clear knowledge of early years mathematics and drawing upon their own teaching experience, which teachers responded well to. They clearly explained the aims of the Reception Jigsaw, the research upon which it is based and were able to answer all questions and challenges teachers posed. Trainers built good relationships and established effective rapport with all participants in the sessions observed. For example, while teachers completed the activities, trainers went around the pairs/groups to speak with staff on a more personal level and ensure everyone remained engaged and included in the session.

4. To what extent do participants feel the intended outcomes of the programme are being achieved for children, practitioners, and the school? How are they being achieved?

Perceived impact

Why did schools get involved with the Reception Jigsaw? What did they hope to gain from participation?

To understand schools' reasons for participating in the Reception Jigsaw, the evaluation team asked mathematics leads and senior leaders in case study schools what had attracted them to the trial and how the Reception Jigsaw fitted with wider school strategies/plans for improvement. Interviewees also highlighted what they hoped to gain for teachers, pupils, and the wider school. Reasons for participation were very similar across longitudinal and best practice case study schools.

WRM were highly regarded by all of the case study schools ahead of participating in the Reception Jigsaw, which had fed into schools' decision to sign up for the trial. Mathematics leads reported that their schools used WRM schemes and/or resources to teach mathematics in other year groups, for which teachers had received training. They saw Reception Jigsaw as a CPD opportunity for reception staff to receive similar WRM training and support them with creating a consistent mastery approach to mathematics teaching across the school. This was seen as a priority for these schools.

Another priority for some of the case study schools was to develop or improve their early years teaching, provision, and outcomes.

Did teachers feel that the support prepared them to deliver the Reception Jigsaw in reality?

The endpoint survey asked reception teachers, reception TAs, Year 1 teachers, and mathematics leads to indicate the extent to which the CPD sessions prepared them to implement the Reception Jigsaw. Overall, agreement ('strongly

agree' or 'agree') was highest from reception teachers (97%) and reception TAs (94%), followed by mathematics leads (91%) and Year 1 teachers (89%).

Mathematics leads and senior leaders were asked in the endpoint survey to what extent they felt prepared to support staff in their school to apply learning from the Reception Jigsaw. The vast majority felt prepared where 97% of mathematics leads and 87% of senior leaders reported 'strongly agree' or 'agree'.

The evaluation team also asked reception teachers during the longitudinal case study interviews if they felt the CPD sessions and half-day coaching visits had prepared them to implement the Reception Jigsaw in reality. All reception teachers felt the programme had upskilled them and they were confident to implement within their classrooms. The only additional support they would have liked, noted by a small number of teachers, was guidance on supporting the transition from reception to Year 1.

What are the perceived impacts for teachers?

The Reception Jigsaw logic model states that the short-term outcomes at the teacher level will be that practitioners gain confidence and improved knowledge in teaching mathematics. The findings from the endpoint survey and interviews supported the intended short-term outcomes stated in the logic model, with data from both sources being largely positive. The impacts for reception teachers, reception TAs, Year 1 teachers, and mathematics leads are set out below.

Impacts for reception teachers

All of the 68 reception teachers who responded to the endpoint survey reported that the Reception Jigsaw had a positive impact on their mathematics teaching. Over four-fifths (87%) reported a large positive impact and a small proportion (13%) reported a slight positive impact. The evaluation team explored the impact Reception Jigsaw had on reception teachers more thoroughly through the longitudinal and best practice case study interviews. The findings are outlined below.

Reception teachers, mathematics leads, and senior leaders in both longitudinal and best practice case study schools reported similar impacts for reception teachers. They reported that participating in the Reception Jigsaw had led to improvements in the quality of reception mathematics teaching and increased teachers' confidence. Areas where teacher confidence had improved included: in the use of mathematical vocabulary; planning extension activities; ensuring prior learning is concrete before moving onto the next topic; creating mathematics learning opportunities; and in understanding pupils' progression.

> *We feel really secure in where our children are up to. Those children who aren't at expected level, we know where their gaps are and what we need to do to support those children which is really important. Before, we didn't have that intervention for children to close the gaps.'* (Mathematics lead, best practice case study school)

Teachers reported that the Reception Jigsaw had provided them with a breadth of ideas, games, books, and resources, which they could draw upon to support their teaching of all mathematics topics. They had used these to enhance both their direct teaching practice and continuous provision. They felt these ideas had enabled them to create exciting and varied mathematics curriculums.

Impacts for reception TAs

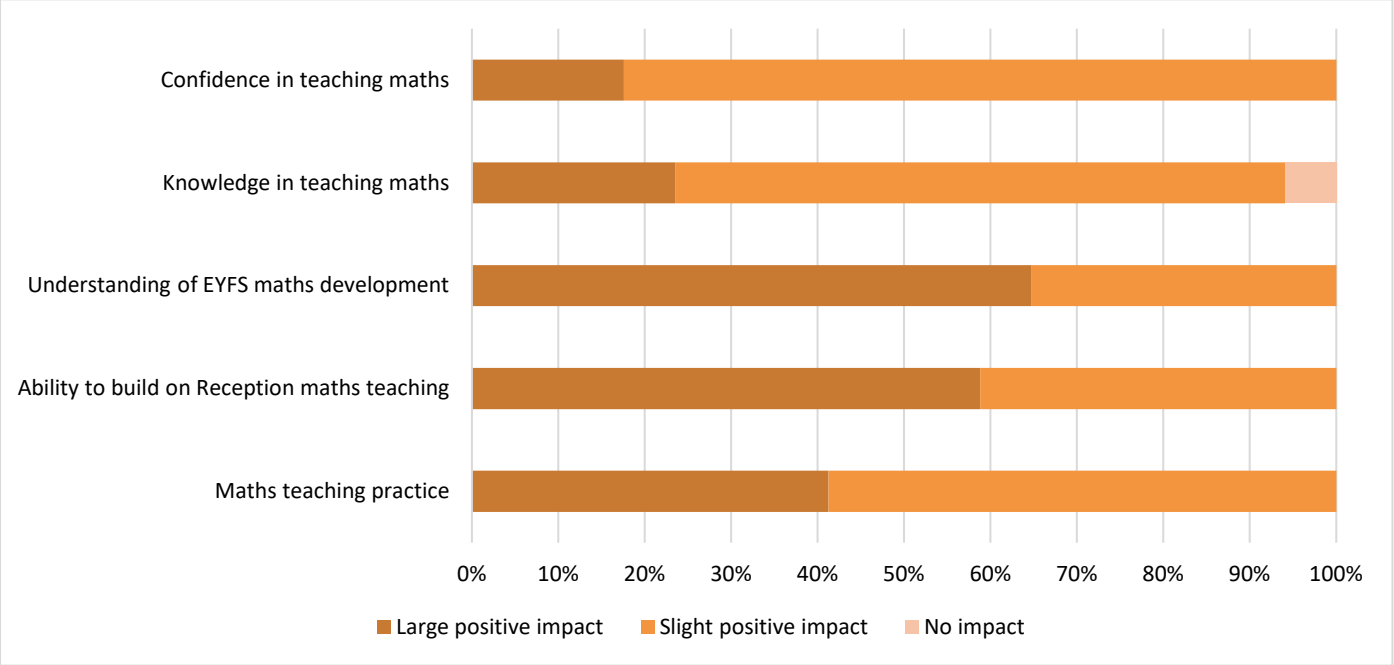
The majority of the reception TAs who responded to the endpoint survey reported that the Reception Jigsaw had a positive impact on their mathematics teaching. Around 69% of the 29 reception TAs who responded reported a large positive impact and 17% reported a slight positive impact. The remaining 14% were unsure of the impact. The evaluation team explored the impact Reception Jigsaw had on reception TAs more thoroughly through the longitudinal case study interviews with reception teachers, mathematics leads, and senior leaders. They reported that reception TAs were more confident supporting pupils during mathematics lessons and in provision, for example extending their learning through questioning and modelling mathematical vocabulary.

Interviewees reported that reception TAs do not often have the opportunity to engage in CPD alongside teachers, so the Reception Jigsaw had been a welcome opportunity. Reception teachers reported that the joint training had meant reception TAs were able to better support them with implementing and delivering mathematics activities in the classroom, because they had practiced the activities themselves and understood the rationale behind them.

Impacts for Year 1 teachers

Along with increased confidence and improved knowledge in teaching mathematics, the Reception Jigsaw logic model stated that practitioners in Key Stage 1 will have a better understanding of children's EYFS mathematics development and will build on reception mathematics teaching practice. Year 1 teachers in the intervention group did not receive the Confidence in Mathematics questionnaire, conducted for the impact analysis of this trial, so the evaluation team used the endpoint survey to understand if the intended outcomes of the logic model had been met for Year 1 teachers. The majority of Year 1 teachers reported a positive impact on their confidence and mathematics teaching practice, however the proportion reporting a large or slight positive impact differed across different dimensions. These survey findings are illustrated in Figure 18 below.

Figure 18: Impact of Reception Jigsaw on Year 1 teachers



Data from the endpoint survey: Overall, what has been the impact of the Reception Jigsaw on your… (N=17). Percentages have been rounded so may not sum to 100.

All Year 1 teachers in longitudinal and best practice case study schools reported that the training had increased their awareness of the mathematics content pupils cover in reception, which helped them to understand their starting point in Year 1 and their ability to build upon this. For example, one teacher commented:

> *For me, doing the Jigsaw was about understanding what the children had seen in EYFS so that I could use the ideas for those children in Year 1, not just for those who find maths difficult. It was also a good reminder about which skills I could actually be promoting through those sorts of activities. Like subitising, that has become a really important skill. Going into it I hoped it would supplement what we were doing and it definitely did that.* (Year 1 teacher, best practice case study school)

Impacts for mathematics leads

The endpoint survey asked mathematics leads to comment on the impact Reception Jigsaw had on their mathematics teaching practice. Most mathematics leads reported a positive impact—four-fifths (81%) of the 31 mathematics leads who responded to the survey reported a large positive impact and a small proportion (16%) reported a slight positive impact. Only a small proportion (3%) were unsure. The interviews with mathematics leads and senior leaders in case study schools provide further insights into the impact Reception Jigsaw had on mathematics leads.

The impacts reported were seen for mathematics leads in all six case study schools (longitudinal and best practice). Mathematics leads had improved in their role over the year, commenting that the support from the trainer during the half-day coaching visits had been pivotal to their development and had enabled them to better support teachers. One mathematics lead commented:

> *It's been great from my point of view as a curriculum lead to understand the changes to the curriculum within reception…actually having that training myself has been really useful…When the teachers come to me and say, 'they're not grasping numbers one to five'—I have those points [from the training] to go back to so I can help them. As a curriculum leader working in a different year it's often hard to do that but having those points and knowledge to reflect back on and the notebooks they provide at twilights is great.* (Mathematics lead, longitudinal case study school)

In addition to this, mathematics leads had gained a deeper understanding of all aspects of early years, including the curriculum, the provision, pupils' starting points, and what they need to achieve by the end of the year. Mathematics leads reported feeling more confident when speaking to senior leaders and Ofsted about reception mathematics.

How had schools monitored the impacts on staff?

The evaluation team asked mathematics leads and senior leaders in the two best practice case study schools how they had monitored and measured the impacts Reception Jigsaw training had upon staff. They reported doing this through observations, from which implementation of the Reception Jigsaw was clear, and staff voice surveys, which showed that teachers' confidence in teaching mathematics and using manipulatives in their lessons had improved. One mathematics lead reported:

> *I have discussed with [the trainer] the use of mathematical language and sustained shared thinking. It is clear from lesson observations that staff are using mathematical terminology when talking to children. The way they are probing questions has improved and it is becoming embedded in their practice.* (Mathematics lead, best practice case study school)

In the other school, the mathematics lead reported that their school improvement partner had identified similar impacts when observing reception teachers' mathematics lessons, along with the enthusiasm teachers display when teaching mathematics.

What is the impact of Reception Jigsaw on the reception classroom/learning environment and Key Stage 1 provision?

The Reception Jigsaw logic model stated that outputs of the programme would include changes to reception learning environment and changes to mathematics provision implemented across Key Stage 1. The findings from the endpoint survey are positive, suggesting this output was achieved in both reception and Year 1 classes.

The endpoint survey asked reception teachers, reception TAs, and mathematics leads to reflect upon the impact Reception Jigsaw had on the reception classroom/learning environment. Other than a small proportion of reception TAs who were unsure of the impact, all other respondents reported a positive impact. The majority reported a large positive impact (reception teachers [79%], reception TAs [83%], and mathematics leads [94%]) and smaller proportions reported a slight positive impact (reception teachers [21%], reception TAs [10%], and mathematics leads [7%]).

The endpoint survey asked Year 1 teachers to reflect upon the impact Reception Jigsaw had on changes to their school's mathematics provision in Key Stage 1. Most Year 1 teachers reported a positive impact—around half (53%) reported a slight positive impact and around a quarter (24%) reported a large positive impact. Just less than a fifth (18%) were unsure and a small proportion (6%) reported no impact on the Key Stage 1 mathematics provision.

These survey findings suggest that Reception Jigsaw had a greater impact in reception classrooms compared to Key Stage 1 classrooms. The case study interviews provide further insight into these findings.

All reception teachers reported that one of the greatest impacts of the half-day coaching visits had been the support with reorganising provision, as described in the quote below:

> *[Reception teachers] are implementing maths opportunities throughout the provision and they have really enhanced the continuous provision, rather than maths activities only being available in a designated maths area.* (Deputy headteacher, longitudinal case study school)

Case study interviewees reported that these changes had a positive impact on how pupils used and accessed mathematics activities in the provision. For example, one teacher reported that pupils engaged in mathematics activities more frequently and independently. Pupils were also combining resources in their play, thereby covering multiple areas of the curriculum at a time.

Year 1 teachers in case study schools did not report making any changes to their provision resulting from the Reception Jigsaw training.

What are the perceived impacts for reception pupils?

The Reception Jigsaw logic model states that in classrooms where reception teachers and reception TAs apply learning from the Reception Jigsaw, the short-term outcomes will be that pupils have a deeper understanding of mathematical concepts and improved mathematics attainment at the end of reception.

The findings from the endpoint survey and interviews support the intended short-term outcomes stated in the logic model, with data from both sources being largely positive. Nearly all reception teachers (98.5%) reported that the Reception Jigsaw had a perceived positive impact on pupils': engagement with mathematics teaching and learning; confidence in mathematics; and knowledge and understanding of mathematics concepts. Around four-fifths of the 68 reception teachers who responded reported a large positive impact and just less than one-fifth reported a slight positive impact. Only 1.5% of respondents reported no impact across the three dimensions. No teachers reported that the Reception Jigsaw had a negative impact upon pupils.

The evaluation team asked reception teachers in all six case study schools (longitudinal and best practice) to comment on the perceived impacts for reception pupils resulting from implementing the Reception Jigsaw. Teachers in both longitudinal and best practice case study schools reported similar perceived impacts for pupils. As reflected in the survey findings, teachers reported that pupils were demonstrating a deeper understanding of number and were showing strong number sense, which had allowed them to quickly grasp number bonds.

Reception teachers reported that pupils' confidence and interest exploring mathematics concepts was evident. For example, teachers reported that pupils were confident talking about their mathematics learning and articulating their reasoning to adults and frequently used mathematical vocabulary to do this:

> *They are so confident—they are using sentence stems and mathematical vocabulary. They can apply what they have been learning and you see it in provision when they are learning on their own.* (Reception teacher, best practice case study school)

Reception teachers felt that all pupils enjoyed mathematics lessons and engaging in mathematics activities in the provision, with one teacher reporting that not one of her pupils was frightened of mathematics as has been seen in previous cohorts, particularly among those pupils who find mathematics more challenging.

Reception teachers themselves were confident that they were sending their pupils into Year 1 with a strong mathematics foundation. One reception teacher said:

> *It just feels that this year, they have got such a solid base knowledge of numbers, shape, how the number system works which is really impactful because it means they grasp all the other skills, such as more than and less than, knowing how to use the number line easier.* (Reception teacher, longitudinal case study school)

How had schools monitored the impacts on pupils?

The evaluation team asked teachers, mathematics leads, and senior leaders in the two best practice case study schools how they had monitored and measured the impacts that implementation of the Reception Jigsaw training had upon pupils.

Reception teachers in both schools reported that the data on reception pupils' mathematics attainment and progress was positive, commenting that the percentage of 'on track' pupils had increased throughout the year. Reception teachers in one school reported using the EEF spiral of progression (shown in the first Reception Jigsaw CPD training) to monitor and record where pupils' progress throughout the year.

Reception teachers reported that observations of pupils in provision were particularly telling of the impact Reception Jigsaw had upon their mathematics abilities. Through observations, they picked up on pupils' confidence engaging in mathematics activities that, pupils accessed and explored mathematics more frequently in provision and the mathematical language they used when interacting with each other.

What are the wider school impacts?

The endpoint survey asked mathematics leads and headteachers to comment on the wider school impacts of the Reception Jigsaw, namely the impact upon staff attitudes to teaching and mathematics teaching practice. Findings from the survey are positive, with the majority of mathematics leads and headteachers reporting a positive impact to both these dimensions.

The majority of mathematics leads (77%) and headteachers (83%) reported that the Reception Jigsaw had a large positive impact on staff attitudes to teaching mathematics across the school. Smaller proportions (mathematics leads [16%] and headteachers [13%]) reported a slight positive impact. Small numbers of mathematics leads and headteachers reported no impacts or were unsure.

Around half of mathematics leads (55%) and headteachers (50%) reported that the Reception Jigsaw had a large positive impact on mathematics teaching practice across the school. Around a third (mathematics leads [36%] and headteachers [37%]) reported a slight positive impact. Small numbers of mathematics leads and headteachers reported no impacts or were unsure.

The evaluation team further explored wider school impacts through interviews with mathematics leads and senior leaders in the longitudinal case study schools. The findings from these interviews are outlined below.

During the final longitudinal case study interviews, the evaluation team asked mathematics leaders and headteachers to comment upon any wider school impacts they had seen from the Reception Jigsaw. At this point, they reported some wider school impacts, but this was limited. As reported above in the 'Reach' section of this report, mathematics leads and senior leaders reported that the Reception Jigsaw principles and activities had been implemented in other year groups to support pupils with SEND and pupils working below age-related expectations. Applying the learning from Reception Jigsaw with these pupils was helping to fill gaps in their early mathematics skills and knowledge. Mathematics leads and senior leaders reported that the Reception Jigsaw training would be shared formally with the rest of the school in the next academic year (2022/2023) through in-school mathematics CPD sessions so whole-school impacts would be more prevalent after this. Further details on this can be found above in the 'Responsiveness' section of this report.

The developers reported that being involved with the Reception Jigsaw trial had raised the importance of early years mathematics in intervention schools. Trainers raised reception teachers' and headteachers' awareness that reception pupils' mathematics results are equally important as phonics, often a competing priority, in contributing the pupils' good level of development scores.

5. What does business as usual consist of for the control group?

Usual practice

Usual practice pre-randomisation

To understand the business as usual practices in all schools participating in the trial, a baseline proforma was sent to each school's nominated contact[29] for the trial, to comment upon their approaches to teaching mathematics in EYFS and to detail any mathematics-related CPD reception teachers had received. This information was used to understand the practices in control and intervention schools pre-randomisation.

Almost all schools randomised to the intervention group (91% of 54 schools) and to the control group (84% of 80 schools) reported at baseline that reception teachers followed WRM schemes of work, suggesting similarities between current practice and the Reception Jigsaw.

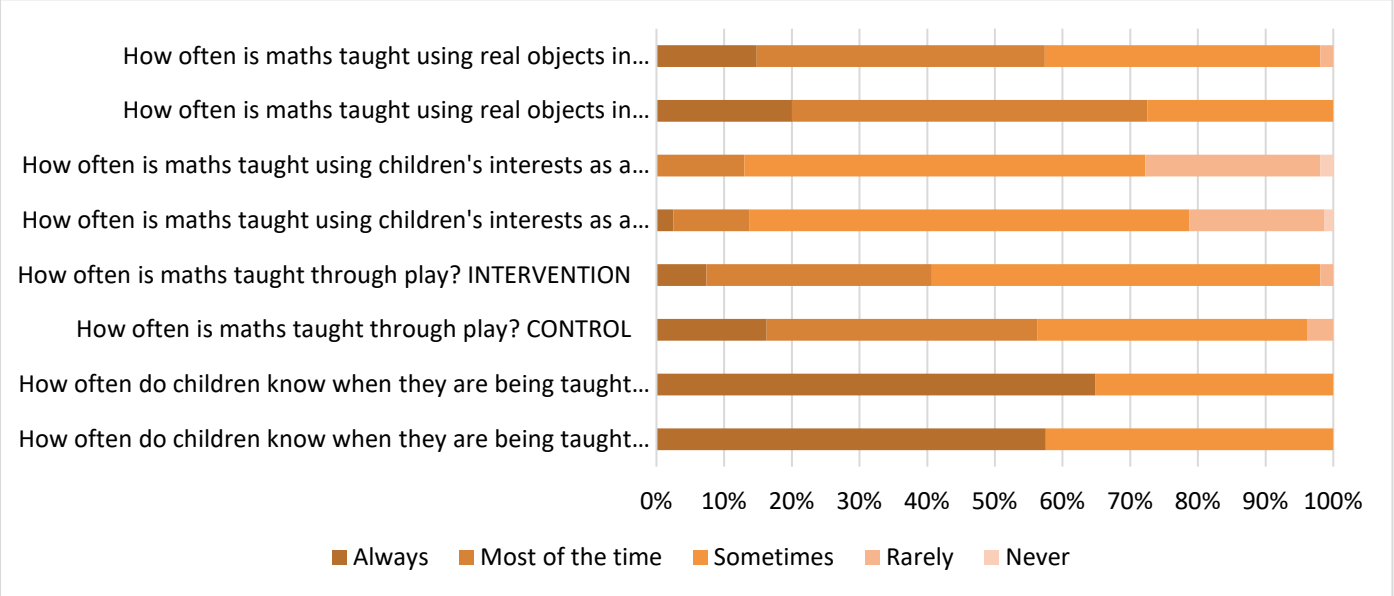
At baseline, many of the control schools (73%) taught mathematics five days a week in reception classes and around a quarter (24%) taught mathematics four days a week. Very few control schools taught mathematics only twice or three times a week. In comparison, around two-thirds of intervention schools (63%) taught mathematics five days a week in reception classes and around a quarter (24%) taught mathematics four days a week. The remaining intervention schools taught mathematics three days a week (7%) or one day a week (6%).

[29] It was suggested in the MoU that the main school contact be the school mathematics lead.

The baseline proforma asked mathematics leads in all schools to describe how mathematics was taught in relation to the rest of the subjects in reception. The proforma presented them with the following options: 'mathematics is taught as discrete sessions'; 'mathematics is taught as part of [other] topics'; or 'a combination' of the two approaches. Almost three-fifths of schools randomised to the control group taught reception mathematics through a combination of the two approaches and the remaining two-fifths taught mathematics as discrete sessions. Among schools randomised to the intervention group, just over half taught mathematics as discrete sessions and just less than half taught reception mathematics through a combination of the two approaches.

Figure 19 below provides mathematics leads' responses to the business as usual proforma, which further detail how mathematics was taught in schools randomised to the control and intervention groups at baseline.

Figure 19: Approach to teaching mathematics in schools randomised to the control and intervention groups at baseline



Notes: Data from the baseline business as usual proforma, control group (N=80), intervention group (N=54).
** For this question, mathematics leads could only respond with 'always', 'sometimes', or 'never'.*

The baseline proforma asked mathematics leads to comment on the balance of adult-led and child-initiated activities in mathematics. In schools randomised to the control group, around half (51%) reported that there was a 50/50 split of adult-led and child-initiated activities and the remaining half (49%) reported that mathematics was primarily adult-led. Analysis by region showed differences in approach between Essex and Yorkshire schools—whereas Essex schools were more likely to report that mathematics was primarily adult-led (73%), Yorkshire schools were more likely to report a 50/50 split (57%). These differences are unlikely to be due to chance, however given the small number of Essex schools in the sample, it is important to interpret the findings with caution. In schools randomised to the intervention group, just over two-fifths (44%) reported that there was a 50/50 split of adult-led and child-initiated activities and a small proportion (4%) reported that mathematics was primarily child-led.

Through the baseline proforma, mathematics leads in all schools provided details of the mathematics-related CPD their reception teachers had engaged with prior to taking part in the evaluation. The evaluation team coded responses into three categories:

- Internal CPD delivered by the school or MAT, such as staff meetings, CPD delivered by the mathematics lead or early years lead, and inset days.
- External CPD delivered by outside organisations, such as the NCETM, WRM, local authorities, the mathematics hub, or a mathematics consultant.
- In-school or school-to-school support, including learning walks and observations in teachers' own schools, or at other schools within their trust/local authority.

Similar proportions of mathematics leads in control schools reported that reception teachers received internal (35%) or external (33%) CPD and 18% reported that reception teachers received a mix of both internal and external CPD. A very small proportion of mathematics leads (1%) reported that their reception teachers engaged with in-school or school-to-school support. Few mathematics leads (13%) reported that their reception teachers had not recently received any mathematics-related CPD.

Findings were similar from schools randomised to the intervention group. Similar proportions reported that reception teachers received external (34%) or internal (32%) CPD and 17% reported that reception teachers received a mix of both internal and external CPD. Only a small proportion of mathematics leads (4%) reported that their reception teachers engaged with in-school or school-to-school support. At baseline, just less than a fifth of mathematics leads (17%) reported that their reception teachers had not recently received any mathematics-related CPD.

Business as usual in case study schools

The interviews with teachers and mathematics leads in the longitudinal case study schools in Autumn Term 2021 provide further insight into the business as usual mathematics practices prior to implementing the programme. Prior to engaging with the Reception Jigsaw, all four longitudinal case study schools reported using WRM schemes and/or resources to inform their teaching of reception mathematics.

Reception teachers and mathematics leads reported that in reception classes, mathematics was taught four to five times per week through direct teaching time (a short input of 15–20 minutes) and through continuous provision, where pupils accessed mathematics activities to build their understanding through practical learning and using concrete resources.

The evaluation team asked teachers and mathematics leads through the first case study interviews what mathematics-related CPD they had recently engaged in. Teachers and mathematics leads reported engaging in minimal CPD. Two teachers had virtually participated in WRM training for reception during the national lockdown and after receiving CPD session one of the Reception Jigsaw, could see similarities between the content of the two courses.

Business as usual in control schools

To understand how business as usual practices changed in schools randomised to the control group during the intervention year (academic year 2021/2022), an endpoint proforma asked mathematics leads in control schools to comment upon any changes made to their approach to teaching mathematics in EYFS and any mathematics-related CPD that teachers and TAs in either reception and/or Key Stage 1 received during the year. This information was used to understand how control schools' normal practices differed to the Reception Jigsaw.

At endpoint, just less than a third (30%) of the 57 mathematics leads who responded reported that there had been significant changes to the way mathematics was taught in reception in the 2021/2022 academic year compared to previous years. The most common change reported was the inclusion of new schemes and/or resources, which guided mathematics teaching, reported by 71% of these mathematics leads. Small proportions reported changes to the balance of direct teaching and time in provision (18%) and a few mathematics leads (12%) reported that mathematics was being taught more frequently this academic year.

Almost two-fifths of mathematics leads (39%) reported that there had been changes to their mathematics scheme or published resources used to guide mathematics teaching in reception. Just less than half of these schools (46%) reported they had started using schemes/resources from the NCETM Mastering Number programme this year, while around a third of these (32%) started using schemes/resources from WRM, suggesting similarity to the Reception Jigsaw. Just less than a fifth (18%) reported including another scheme into their mathematics teaching, such as Power Maths, or had developed their own schemes.

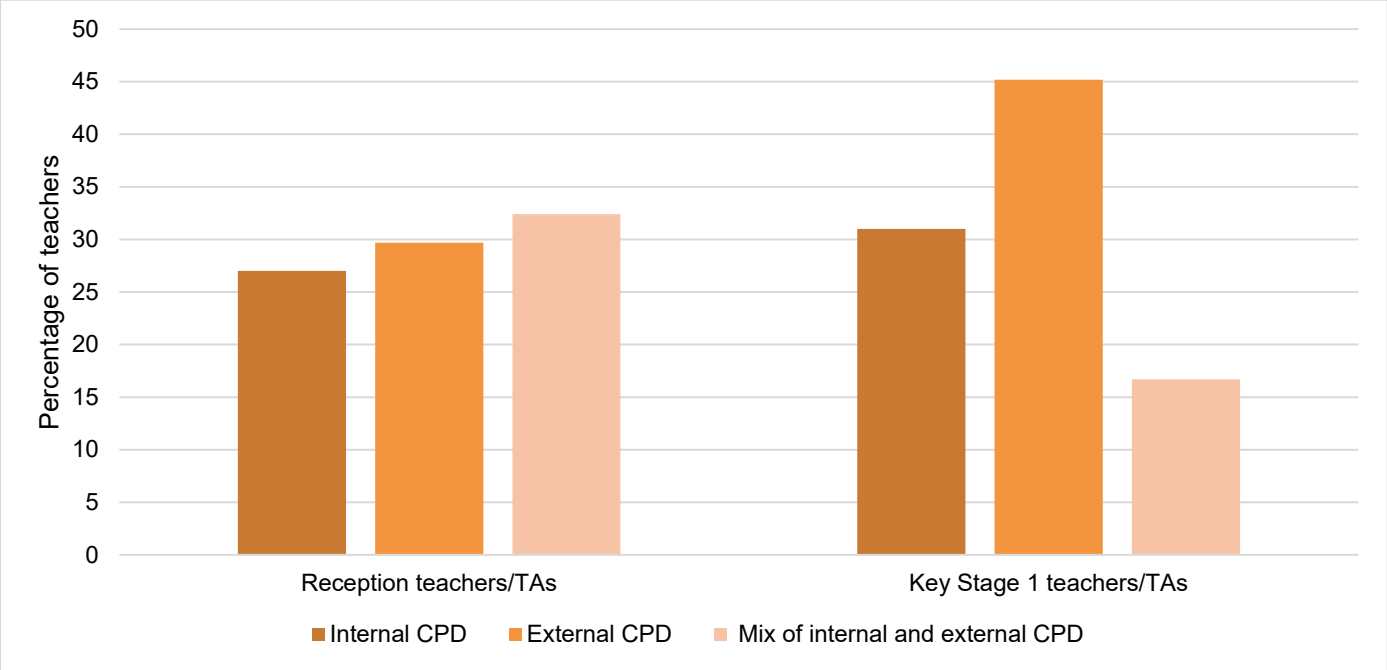
Analysis by region showed that Essex schools were more likely to report making changes to their mathematics scheme (55%) compared to Yorkshire schools (35%). Over half of schools in Yorkshire (56%) reported including Mastering Number into their mathematics teaching, compared to 17% of schools in Essex. In comparison, two-thirds of schools in Essex (67%) had included WRM schemes/resources into their teaching new for the 2021/2022 academic year, compared to 19% of schools in Yorkshire. Despite these differences, sample sizes, particularly in Essex, are too small to determine whether these differences are due to chance or not.

Around a third of all mathematics leads (35%) in control schools reported that they had participated in the Mastering Number programme in the 2021/2022 academic year.

The endpoint proforma asked mathematics leads in control schools to indicate if reception teachers and/or reception TAs and Key Stage 1 teachers and/or TAs had received any mathematics-related CPD during the academic year and if so, what this was. Around two-thirds (65%) reported that reception teachers and/or reception TAs had received mathematics-related CPD and three-quarters (75%) reported that Key Stage 1 teachers and/or Key Stage 1 TAs had received mathematics-related CPD.

The types of CPD reception teachers and reception TAs, and Key Stage 1 teachers and TAs received are shown in Figure 20 below.

Figure 20: Types of CPD reception teachers/reception TAs and Key Stage 1 teachers/Key Stage 1 TAs engaged with during 2021/2022 academic year



Notes: Data from the endpoint business as usual proforma: Please give brief details of the mathematics-related CPD provided to reception teachers and/or TAs this academic year (N=37) and Please give brief details of the mathematics-related CPD provided to Key Stage 1 teachers and/or TAs this academic year (N=42).CPD=Continuing Professional Development; TA=teaching assistant.

Programme differentiation

To what extent do teachers think that Reception Jigsaw is distinctive to their practices prior to randomisation?

The endpoint survey in Summer Term 2022 asked reception teachers, reception TAs, Year 1 teachers, and mathematics leads how distinct Reception Jigsaw was from their usual practice to teaching mathematics. Teachers responded along a three-point scale—'Not distinct', 'Partially distinct', and 'Totally distinct'. The majority of reception teachers (85%), Year 1 teachers (88%), and mathematics leads (87%) reported that the Reception Jigsaw was partially distinct from their usual teaching practice. Small proportions reported that the Reception Jigsaw was either totally distinct or not at all distinct (around 6–7%). In comparison, around half of reception TAs (55%) reported that the Reception Jigsaw was partially distinct and, compared to other respondents, a slightly larger proportion (14%) reported the Reception Jigsaw was totally distinct. However, around a quarter (24%) were not sure. The interviews in case study schools provide further insights into these survey findings.

The evaluation team asked reception teachers and Year 1 teachers at the second and third interviews to comment on the differences between the Reception Jigsaw and their practice before engaging in the trial. Teachers reported that Reception Jigsaw was very similar to their usual practice, but was enhancing this through providing new, practical ideas and further exploration of mathematics through real-life examples to develop children's understanding. The training was providing staff with a deeper understanding of mathematics practice and pedagogy. A Year 1 teacher said: *'It's similar*

to the approach we are already taking but it is deepening our understanding.' (Year 1 teacher, longitudinal case study school)

What, if any, practices and approaches does Reception Jigsaw replace in intervention schools?

Based on interviews with reception teachers, Year 1 teachers, and mathematics leads in longitudinal case study schools, Reception Jigsaw appeared to enhance rather than replace mathematics practices and approaches, as discussed above. Before the Reception Jigsaw, some teachers worked through WRM schemes in a linear way; whereas after the training, they had been able to improve the sequencing of their lessons, for example viewing the scheme as a whole and looking at how different topics can fit together and be delivered in the same week. They also reported delivering at a slower pace, not being afraid to pause and revisit a topic to ensure pupils' knowledge and skills were secure before moving on. They reported that Reception Jigsaw had helped them to deliver mathematics lessons with more structure and had led to lessons having a weekly focus or objective. Teachers reported that their classrooms were now 'language-rich environments' because teachers and TAs were confident in modelling mathematical language.

What (other) mathematics-related CPD activities and mathematics interventions have intervention group teachers taken part in during the trial?

Through the endpoint survey, case study interviews in longitudinal and best practice schools and an interview with the developer, the evaluation team aimed to ascertain what other mathematics-related CPD schools in the intervention group had participated in alongside Reception Jigsaw.

Over half of mathematics leads (59%) in intervention schools who responded to the endpoint survey reported participating in the Mastering Number programme.[30] Interviews with teachers in case study schools provide detail on how teachers delivered both Mastering Number and Reception Jigsaw.

Four of the six case study schools (two longitudinal and two best practice) participated in Mastering Number alongside the Reception Jigsaw, however use of the programme varied. In two schools (one longitudinal and one best practice case study), reception teachers chose not to follow the programme, instead dedicating their focus to Reception Jigsaw. In the other two schools (one longitudinal and one best practice case study), reception teachers used Mastering Number as an additional intervention or for starter tasks and used Reception Jigsaw to guide their main lesson.

Costs

Time

Data was gathered from the developers in terms of the time involved in undertaking the WRM training and coaching visits. Reception teachers, Year 1 teachers, and school mathematics leads are expected to attend the five two-hour CPD sessions. It is recommended the reception and Year 1 TAs also attend these sessions. At least one reception teacher is expected to attend the five half-day coaching sessions. The expectation set by WRM at the beginning of the trial was that school mathematics leads would also attend at least part of each coaching visit however on reflection WRM felt that going forward this would not be a requirement. Gap tasks were also completed by reception teachers as part of the intervention, but WRM reported that these could be completed as part of usual planning and preparation time and have therefore have not been included as an additional cost in this evaluation. Findings from case study interviews supported this.

Through the endpoint survey, data was collected on how much time was spent by teachers and TAs implementing Reception Jigsaw approaches in their schools. Respondents to the survey were asked to estimate time spent (beyond their normal mathematics planning and preparation time) on the activities listed in the Table 30 below. Reception TAs spend considerably less time than reception teachers and school leaders on delivery of Reception Jigsaw. Both reception teachers and school mathematics leads spent an average of more than four hours on trialling Reception Jigsaw techniques and planning Reception Jigsaw activities over the course of the academic year. Reception teachers also spent additional time (> 4 hours) focusing on resources and the learning environment. School mathematics leads also spent time (> 4 hours) adapting Reception Jigsaw techniques.

[30] Participating schools receive central training (online and face-to-face) and pupil-facing resources. There is an expectation that schools will provide a daily teaching session for all children of 10 to 15 minutes, in addition to their normal maths lesson. See for further information: https://www.ncetm.org.uk/maths-hubs-projects/mastering-number/.

Table 30: Total time devoted by personnel for preparation and delivery

			Year 1			Year 2			Year 3		
	Activity	**Staff member**	**Number of Teachers / TAs**	**Mean number of total hours per person (min, max) (reception teacher n=63, reception TA n=25, mathematics lead n=22)**		**Number of teachers**	**Mean number of total hours (min, max)**		**Number of teachers**	**Mean number of total hours (min, max)**	
Preparation	Attending CPD twilight	Reception teacher/Year 1 teacher/school mathematics lead	all	10	n/a	0	0	n/a	0	0	n/a
	Attending CPD twilight	Reception TA/Year 1 TA	all	10	n/a	0	0	n/a	0	0	n/a
	Attending coaching session	Reception teacher	1	15	n/a	0	0	n/a	0	0	n/a
	Attending coaching session	School mathematics lead	1	variable	n/a	0	0	n/a	0	0	n/a
	Teacher cover (while teachers attend coaching session)	Classroom teacher/TA	all	15	n/a	0	0	n/a	0	0	n/a
Delivery	Revisiting content from the RJ training sessions	Reception teacher	all	2.52	(0, 12)	0	0	n/a	0	0	n/a
	Revisiting content from the RJ training sessions	Reception TA	all	0.88	(0, 10)	0	0	n/a	0	0	n/a
	Revisiting content from the RJ training sessions	School mathematics lead	1	3.68	(0, 50)	0	0	n/a	0	0	n/a
	Trialling RJ techniques	Reception teacher	all	4.44	(0, 50)	0	0	n/a	0	0	n/a
	Trialling RJ techniques	Reception TA	all	0.88	(0, 10)	0	0	n/a	0	0	n/a
	Trialling RJ techniques	School mathematics lead	1	8.05	(0, 60)	0	0	n/a	0	0	n/a
	Adapting RJ techniques	Reception teacher	all	1.98	(0, 40)	0	0	n/a	0	0	n/a
	Adapting RJ techniques	Reception TA	all	0.64	(0, 3)	0	0	n/a	0	0	n/a
	Adapting RJ techniques	School mathematics lead	1	4.09	(0, 30)	0	0	n/a	0	0	n/a
	Planning RJ activities	Reception teacher	all	4.48	(0, 50)	0	0	n/a	0	0	n/a

	Planning RJ activities	Reception TA	all	0.64	(0, 10)	0	0	n/a	0	0	n/a
	Planning RJ activities	School mathematics lead	1	5.50	(0, 60)	0	0	n/a	0	0	n/a
	Organising/creating maths resources for the RJ	Reception teacher	all	4.52	(0, 50)	0	0	n/a	0	0	n/a
	Organising/creating maths resources for the RJ	Reception TA	all	0.4	(0, 2)	0	0	n/a	0	0	n/a
	Organising/creating maths resources for the RJ	School mathematics lead	1	3.55	(0, 20)	0	0	n/a	0	0	n/a
	Preparing the reception classroom/learning environment for the RJ	Reception teacher	all	5.05	(0, 50)	0	0	n/a	0	0	n/a
	Preparing the reception classroom/learning environment for the RJ	Reception TA	all	0.36	(0, 2)	0	0	n/a	0	0	n/a
	Preparing the reception classroom/learning environment for the RJ	School mathematics lead	1	2.64	(0, 20)	0	0	n/a	0	0	n/a
	Supporting other staff to implement RJ	Reception teacher	all	2.30	(0, 20)	0	0	n/a	0	0	n/a
	Supporting other staff to implement RJ	Reception TA	all	0.32	(0, 2)	0	0	n/a	0	0	n/a
	Supporting other staff to implement RJ	School mathematics lead	1	5.05	(0, 80)	0	0	n/a	0	0	n/a
	Incorporating RJ into maths policies	Reception teacher	all	1.02	(0, 30)	0	0	n/a	0	0	n/a
	Incorporating RJ into maths policies	Reception TA	all	0.2	(0, 2)	0	0	n/a	0	0	n/a
	Incorporating RJ into maths policies	School mathematics lead	1	2.59	(0, 40)	0	0	n/a	0	0	n/a
	Other RJ-related activities	Reception teacher	all	1.30	(0, 30)	0	0	n/a	0	0	n/a
	Other RJ-related activities	Reception TA	all	0.8	(0, 12)	0	0	n/a	0	0	n/a
	Other RJ-related activities	School mathematics lead	1	1.36	(0, 20)	0	0	n/a	0	0	n/a

CPD=Continuing Professional Development; n/a=not applicable; RJ=Reception Jigsaw; TA=teaching assistant.

Financial

Financial costs borne by schools associated with their participation in the Reception Jigsaw (as delivered in the trial) were collected directly from schools. The quantitative data presented below were collected through the two endpoint practitioner surveys. Data from case study interviews supported the findings presented. There were three groups of financial costs to the programme. The programme cost for the schools in the trial was a subsidised rate of £1,000 per school (plus valued added tax [VAT]). The other financial costs were those associated with staff attending the CPD twilight sessions and the coaching visits and the costs of resources and equipment to support the delivery of Reception Jigsaw.

The CPD sessions are held after school, so no teaching cover is required however, some schools reported that they paid TAs to attend these sessions. Around 20 of the 30 responding headteachers paid TAs for their attendance. A further three reported giving time off in lieu. At least one reception teacher in each school was expected to attend half-day coaching visits, which took place during the school day. Nine of the 30 responding headteachers paid some cover costs for reception teachers to be released from their classrooms. Around 90% of headteachers used internal cover by either other teachers or TAs/support staff to cover at least some of the coaching sessions. A smaller percentage reported that the coaching visits took place during non-teaching time (e.g. PPA time).

In the calculations of the per pupil costs reported in Table 31, all the costs described are considered as start-up costs and therefore are only considered in the first year of the programme. The per-pupils costs calculated over a three-year period is £15.77 and is therefore of very low cost as per the EEF cost thresholds (see Appendix A).

Table 31: Recurring costs of the implementation of the programme, per ingredient

			Nominal values			**Real values (deflate using Year 1 as base)**			**Present value**	**Cost in analysis year**
		Start-up or Recurring?	**£ Year 2022**	**£ Year 2023**	**£ Year 2024**	**£ Year 1 (2022 prices)**	**£ Year 2 (2022 prices)**	**£ Year 3 (2022 prices)**	**£ Present value (2022)**	**£ Present value (2022)**
Personnel for training	Cover (TA CPD training)	Start-up	371.69			371.69			371.69	371.69
	Cover (reception teacher coaching visit)	Start-up	262.87			262.87			262.87	262.87
	Cover (school mathematics lead coaching visit)	Start-up	46.83			46.83			46.83	46.83
Training and programme costs	Programme costs (subsidised)	Start-up	1,200			1,200			1,200	1,200
Facilities, equipment. and materials	Resources and equipment	Start-up	£131			£131			£131	£131
Total cost per school										**2,012**
Number of pupils per school year										43
Cost per pupil school year*										**15.77**

**Over a three-year period, three cohorts of pupils would have been exposed to the programme (3 x 43 pupils). This is a precise number (total cost per school and number of pupils has been rounded).*

CPD=Continuing Professional Development; TA=teaching assistant.

The programme cost of the intervention was subsidised for schools taking part in the trial. To get an understanding of the market costs to schools outside of the trial the full market costs must be included. Under the current delivery model, the costs to schools varies in terms of distance from WRM head office in Halifax as the time needed for trainers to travel increases as distance increases. The market cost to schools in addition to the £1,200 paid as part of the trial varies from £2,550 to £3,795. Table 25 shows the range of per pupil cost estimates at market prices. Schools are expected to cover the cost of travel and accommodation where this is needed, and this has not been included in these calculations. As can be seen from Table 32, even at market prices Reception Jigsaw is considered to be of very low cost under the EEF thresholds. It is also worth noting that costs could be reduced further if local trainers are recruited and trained as was originally planned for the Essex schools in this trial.

Table 32: Cost estimates at market prices

Estimate	Total cost per school	Cost per pupil school year
Lower estimate (schools close to Halifax)	£4,562	£35.75
Upper estimate (schools farthest from Halifax)	£5,807	£45.50

Conclusion

There was a positive impact of the Reception Jigsaw on teachers' confidence in teaching maths to reception-age pupils, with the analysis showing an effect size of 0.6. Triangulation of this impact finding with self-reported findings from the endpoint survey and the case study interviews strongly support this conclusion. It is important to note that there was a high level of attrition for the practitioner survey, although this was balanced across the control and intervention groups. Reception teachers, and to a lesser extent, reception TAs and Year 1 teachers, reported that they felt taking part in the Reception Jigsaw had a large impact on their mathematics teaching, their mathematics classroom, their confidence, and ultimately on their mathematical knowledge and understanding of their pupils.

As reported in Table 33, there was not enough evidence to conclude whether being taught by staff who were assigned to take part in the Reception Jigsaw had an effect on mathematics attainment at the end of reception. While the analysis of the primary outcome did show a small positive effect on pupil mathematics outcomes from schools participating in the Reception Jigsaw, it is not possible to conclude that this was not due to chance. This finding was reflected in all three pupil-level outcome measures looking at attainment. Despite there being no evidence of an impact in the impact evaluation, nearly all responding teachers reported that they perceived a positive impact on pupils': engagement with mathematics teaching and learning; confidence in mathematics; and knowledge and understanding of mathematics concepts.

The longitudinal analysis, based on assessments which pupils sat a year after the main trial when they were at the end of Year 1 (summer 2023), found similar results to the main analysis. Pupils who were in schools that had the opportunity to participate in Reception Jigsaw, achieved slightly higher in their mathematics assessments overall but there was not enough evidence to conclude that this finding was not due to chance.

Analysis on the impact of attending a high proportion of the Reception Jigsaw twilight and coaching sessions found that, where a pupil's reception teacher attended at least nine out of ten sessions, there was a small positive effect in terms of pupils' mathematics attainment. This is the equivalent to approximately two months of additional progress over that made by pupils who were taught by a similar teacher from schools, which did not receive the Reception Jigsaw training. The longitudinal follow-up analysis showed that this beneficial impact of being taught by a reception teacher who had attended at least nine Reception Jigsaw sessions, was maintained.

The subgroup analysis showed that there was no impact of the Reception Jigsaw on the outcomes of pupils eligible for FSM (Hedges' g = -0.02 [CI: -0.2, 0.2]) or for pupils in schools that signed up to take part in the Mastering Number programme (Hedges' g = 0.0 [CI: -0.3, 0.2]). Findings from the longitudinal follow-up subgroup analysis were in line with the subgroup analysis from the main trial period.

Table 33: Conclusions

Key conclusions	
1.	Children in Reception Jigsaw schools made the equivalent of one additional month of progress in maths, on average, compared to children in other schools. This result has a high-security rating. These results are not statistically significant. This means that the statistical evidence does not meet the threshold set by the evaluator to conclude that the true impact was non-zero
2.	Children eligible for free school meals made no additional progress in maths, on average, compared to children in other schools. This result has a lower security rating than the overall finding because of the smaller number of pupils
3.	There was evidence that the Reception Jigsaw programme led to reception teachers being more confident in teaching mathematics
4.	The programme was well received by schools. Reception teachers reported positively about the support they received throughout the Reception Jigsaw programme, finding it relevant to their teaching practice, high quality, and impactful in terms of the changes they made to their practice and environment
5.	Year 1 teachers also reported positively about the support, but less so compared to reception teachers. However, they could see that it would help them with supporting pupils to transition from reception to Year 1, and to build on knowledge pupils had gained in reception

6. Longitudinal analysis found that children in Reception Jigsaw schools in reception made an average of one month's additional progress at the end of Year 1 compared to children in other schools. However, the statistical evidence does not meet the standard set by the evaluator to conclude that the true impact was non-zero.

Impact evaluation and IPE integration

Evidence to support the logic model

The logic model presents increased practitioner confidence in teaching mathematics as the first short-term outcome of the hypothesised causal chain. This evaluation supports this claim with strong evidence from the impact evaluation that there was a statistically significant increase in confidence of teaching mathematics to reception-age pupils for those practitioners who had access to the Reception Jigsaw, compared to those who did not. Practitioners included in this part of the analysis were all the reception teachers, reception TAs, and school mathematics leads who were in the schools in the trial at the start of the academic year 2021/2022. The majority of reception teachers and reception TAs self-reported large impacts of the Reception Jigsaw on their teaching and on the learning environment. Year 1 teachers responses mirrored that of reception teachers although self-reported impacts were lower, seemingly due to the reception-focus of the programme. There was insufficient evidence to draw conclusions as to whether there is any impact on practitioner confidence in their own mathematics ability.

While there was not enough statistical evidence in this trial to conclude that there was a positive impact on pupils' mathematics attainment at the end of reception when analysed on an ITT basis, pupils whose reception teacher had had access to the Reception Jigsaw made more progress in all three mathematical attainment outcomes. Teacher-reported data from the endpoint surveys support this finding; nearly all of the reception teachers who responded to the survey reported a large (81%) or slight (18%) impact of the programme on their pupils' mathematics attainment.

The impact analysis supports the mediator in the logic model relating to completion of the training. The measure used in the statistical analysis to assess compliance was attendance at each of the ten training sessions (five twilights and five half-day coaching visits). It is important to note that this is a teacher-level measure of compliance and also that some teachers (and therefore their pupils) are considered non-complaint as they may have left the school part way through the academic year. Pupils taught by a reception teacher who had attended at least nine of the Reception Jigsaw sessions made on average two months more progress than pupils who were taught by similar teachers in schools, which were not receiving the Reception Jigsaw training. The longitudinal analysis provides some evidence that this additional progress is maintained to the end of Year 1.

It is also important to note that, expectations around the attendance to the sessions at school-level, as described by WRM, involve a slightly higher level of flexibility around reception teacher attendance than used in our teacher-level compliance measure. WRM expects attendance at the twilight CPD sessions for all reception teachers but only one reception teacher need attend the coaching visit. School-level compliance also requires the attendance of Year 1 teachers at the twilight training and mathematics leads at both the twilight training and the coaching visits.

Participation of the mathematics lead and support from senior leadership is highlighted as a mediator to successful implementation. An element of this is supporting collaboration and learning across year groups. Findings from the IPE suggest that professional collaboration is another outcome of the intervention and one that could be a mechanism for change. Professional collaboration could lead to joined up maths strategy throughout the school, culture of openness, and sharing learning, which then positively impacts all teachers' maths practice, and therefore maths progress and attainment throughout the school (linked to medium- and long-term impact).

Longitudinal follow-up assessments were sat by pupils in randomised schools in summer 2023 to assess the impact of the Reception Jigsaw on one of the hypothesised medium-term outcomes, namely improved attainment in mathematics at the end of Year 1. The additional month's progress made by pupils who were in Reception Jigsaw schools seen at the end of their reception year remained at the end of Year 1 although the confidence interval includes zero we cannot conclude this is a genuine effect.

Interpretation

Findings presented in this report shows strong evidence of the effectiveness of the Reception Jigsaw in terms of increasing practitioners' confidence in teaching mathematics to reception pupils. This proximal outcome is the crucial first step in raising mathematics attainment for pupils. Evidence suggests that early years teachers often do not have a

positive attitude to mathematics, including lacking confidence in their own mathematics ability (All Party Parliamentary Group for Maths and Numeracy, 2014; Boaler and Dweck, 2016). Negative attitudes to mathematics among educators can influence how they teach mathematics as well as how much time they spend teaching mathematics (Boaler and Dweck, 2016; Russo *et al.*, 2020; Norton, 2017; Christensen and Knezek, 2020). The EEF evidence review on improving mathematics in early years and Key Stage 1 suggests that mathematical knowledge for professional development is important alongside pedagogical content knowledge and knowledge of children's mathematical development (Hodgen *et al.*, 2020).

For impacts to be seen in pupil outcomes, the increased confidence of practitioners must translate into changes to pedagogy and/or the classroom environment. There is promising evidence that the programme has had a positive impact on the quality of teaching in reception and the learning environment, with the majority of teachers stating there had been a large impact on these areas. High proportions of teachers reported that they felt that the Reception Jigsaw was partially distinct from their usual approach to teaching mathematics so despite large proportions of schools using WRM resources prior to the trial, Reception Jigsaw did impact their approach. Findings from the IPE suggest that the impacts of the programme on Year 1 are weaker, which is understandable given the focus of the programme (WRM's Primary Jigsaw package is aimed at Key Stage 1 and Key Stage 2).

The finding from the CACE analysis in the impact evaluation suggests that high levels of attendance at Reception Jigsaw sessions are vital to achieving outcomes in attainment. Attendance at coaching visits is not compulsory for all reception teachers (where a school has more than one reception teacher), so where teachers have high levels of compliance this could signal teachers who are more centrally involved and invested in the programme. The majority of teachers attended the CPD sessions so differences in compliance likely reflect difference in attendance to coaching sessions. Teachers found the tailored support delivered through the coaching sessions very helpful. Completion of the gap tasks is highlighted as a mediator in the logic model, and this was found to be variable. This could also be a factor contributing to this impact finding.

Engagement in both the intervention and the trial was high, with no schools withdrawing from the intervention group or the intervention itself. This likely reflects findings from the endpoint survey and case studies, which show that the teachers were very positive about the quality of the training and coaching. While the statistical findings of the effect of the Reception Jigsaw on pupils' mathematics attainment only yielded a positive effect when compliance was taken into account, perceived impacts were reported by the teachers.

The longitudinal follow-up in Summer Term 2023 for the pupils at the end of Year 1 was important, as it provided an opportunity to understand whether there is an impact on attainment seen further down the causal chain. Although there was not enough evidence at the end of the main phase trial to draw positive conclusions around the effectiveness of the Reception Jigsaw on mathematics attainment, there were reasons to expect positive impacts at the end of Year 1.

First, pupils in intervention schools were likely to have had increased exposure to strategies and approaches advocated by Reception Jigsaw. Second, schools and teachers would have had longer to embed change initiated by the Reception Jigsaw in their schools. Third, benefits due to pupils' potentially deeper understanding of number may take a longer time to lead to improvements in mathematics assessments. Last, the use of a written assessment (such as the New PUMA) may have been more appropriate for pupils at the end of Year 1 rather than reception, as pupils are likely to have experienced more formal learning after a year of Key Stage 1. A mechanism in the logic model for improvement in attainment at the end of Year 1 is the Year 1 teachers and Year 1 TAs applying their learning from Reception Jigsaw in the classroom. At the end of the main phase of the trial Year 1 teachers reported using Reception Jigsaw less than reception teachers. However, this may not have reflected their practice during the academic year 2022/23 when their pupils would have already experienced Reception Jigsaw strategies in their reception year and practices may have become more embedded.

As the results from the longitudinal follow-up did not find a demonstrably larger effect than the main phase analysis, the hypothesised mechanisms for Reception Jigsaw increasing mathematics attainment in Year 1 are not supported by the evidence presented in this report. However, as the positive compliance result was maintained at the end of Year 1, this is indicative of Reception Jigsaw having longer-term benefits to pupils where their teachers attended at least nine of the ten sessions.

Limitations and lessons learned

Attrition was low at the school level, with only three out of 138 schools (2%) dropping out of the trial in the main stage of the trial. These were all control schools and were all picked up again at the longitudinal follow-up where there was no attrition at the school level. At pupil level, attrition was higher at 13% for the main trial. This was driven by absence on the day of the assessment (7%), with a further 2% of pupils having left the school, 2% having been withdrawn from testing by the teacher and the remaining 2% attending withdrawn schools. There were only small differences between intervention and control groups in terms of the reasons for pupil-level attrition. There is unlikely to be bias due to missing data.

The Mastering Number programme was rolled out by the DfE at the start of the academic year, meaning it coincided with the trial. Due to the overlap in subject and age group focus for this programme, data on Mastering Number participation was gathered from schools also participating in the trial. While involvement in Mastering Number, as reported at the beginning of the academic year, was similar across intervention and control groups, there was a difference between the intervention and control groups in terms of percentages of schools at endpoint, which stated that they had participated in the Mastering Number programme. Of the control group respondents, 49% of mathematics leads reported at endpoint that their school had participated in Mastering Number in the academic year 2021/2022 compared to 59% of intervention schools. This disparity in participation raises concerns around concurrent intervention bias as the effect being measured may be the impact of the Mastering Number programme (or the interaction between the two programmes). It is plausible that those schools allocated to the intervention group who had also participated in Mastering Number on some level, did not focus on Mastering Number in reception classes and instead prioritised Reception Jigsaw. Case study evidence suggests this may have been the case. Four (of the six) case study schools were signed up to participate in Mastering Number. Two of these did not continue with Mastering Number and two used Mastering Number as an additional resource but used Reception Jigsaw to guide their main lesson. However, it is unknown whether the case study findings fully reflect the broader intervention schools. Reassuringly, the subgroup analysis looking at schools who reported that they were participating in Mastering Number did not find any effect of the participation in Reception Jigsaw on pupils' mathematics outcomes. It should be noted though that the data on Mastering Number participation was collected prior to randomisation and there is evidence that some schools signed up to participate later in the school year (see 'Outcomes and analysis' section above).

Feedback from WRM and some schools involved in the study about their experience of taking part in the trial suggested that some aspects of the endpoint assessment (New PUMA) were not aligned with expectations for pupils of this age, or the teaching approaches most often used such as the use of manipulatives and a focus on talk. Some schools felt that the assessment was not appropriate for the age of the pupils and could have been more tailored to four- to five-year-olds. A few schools reported that they felt the assessment was at odds with the approaches followed in the Reception Jigsaw. Some of this misalignment related to the curriculum focus and is why the reduced PUMA was also analysed, although the result was similar.

The bespoke EN checklist created by NFER with input and advice from early years specialists at WRM was not a fully-validated measure. Evidence from the trial showed that it had excellent internal consistency (reliability) with a Cronbach's alpha index of 0.95. It also had a high pre-/post-test correlation with the New PUMA (0.59) contributing to quite a small MDES, which was the aim of its inclusion. A useful finding here is that well-informed assessments (or checklists) can be extremely useful in explaining outcome variance.

The trial was an efficacy trial and as such the developers recruited the schools and delivered the programme. Therefore, the results refer to this model—a small team of trainers delivering the Reception Jigsaw in schools—and the generalisability of the results to all schools should be made with some caution. As WRM recruited the schools to the trial, they were likely to be schools known to WRM and are therefore perhaps more likely to be open to the approach used by WRM, compared to the broader population of all schools. Indeed, over 80% of schools were using WRM resources before the start of the trial. Any scaling-up of the programme would necessarily involve broader recruitment of both schools and trainers. The pool of trainers in the trial was smaller and more experienced than planned, due to not being able to recruit new trainers in the Essex area. This also led to a deviation in the model for schools in Essex, whereby coaching sessions and twilight sessions happened on the same day. In the Essex case study school, this also meant the trainer frequently spent more time in school as they stayed there for longer than the half-day coaching visit to support staff in the school.

An extension to the IPE in the longitudinal follow-up phase of the trial would have enabled more understanding of the longer-term mechanisms such as senior teachers enabling more embedded practices across the school and Year 1 teachers building on Reception Jigsaw into Key Stage 1.

Future research and publications

If resource was not limited, it would be worth exploring other assessments for young pupils, which better reflect the style of teaching experienced at this age (and promoted by Reception Jigsaw). In mathematics, this is likely to involve manipulatives, as current early mathematics teaching emphasises the importance of concrete objects to facilitate a deep understanding of number.

References

Allen, R., Burgess, S. and Mayo, J. (2012) 'The teacher labour market, teacher turnover and disadvantaged schools: new evidence for England' CMPO Working Paper 12/294. CMPO.

All Party Parliamentary Group for Maths and Numeracy, (2014) Maths and Numeracy in the Early Years. Available at: https://www.nationalnumeracy.org.uk/sites/default/files/images/news-and-media/appg_paper_-_eys.pdf (accessed 26 January 2013).

Andrade, J., Styles, B. and Morton, C. (2022) 'Randomised Controlled Trial: Evaluation of the White Rose Maths Reception Jigsaw: Statistical Analysis Plan'. London: EEF. Available at: https://educationendowmentfoundation.org.uk/projects-and-evaluation/projects/reception-jigsaw (accessed 16 January 2023).

Bates, D., Maechler, M., Bolker, B. and Walker, S. (2015) 'Fitting Linear Mixed-Effects Models Using lme4'. *Journal of Statistical Software*, 67(1): 1–48. https://doi.org/10.18637/jss.v067.i01.

Boaler, J. and Dweck, C. (2016) 'Mathematical Mindsets: Unleashing Students' Potential Through Creative Math, Inspiring Messages, and Innovative Teaching'. San Francisco, CA: Jossey-Bass.

Chen, J.-Q., McCray, J., Adams, M. and Leow, C. (2014). 'A Survey Study of Early Childhood Teachers' Beliefs and Confidence about Teaching Early Math'. *Early Childhood Education Journal*, 42: 367–77. https://doi.org/10.1007/s10643-013-0619-0.

Christensen, R. and Knezek, G. (2020) 'Indicators of Middle School Students' Mathematics Enjoyment and Confidence'. *School Science and Mathematics*, 120(8): 491–503. https://doi.org/10.1111/ssm.12439.

Costa, H.M., Outhwaite, L.A. and Van Herwegen, J. (2021) 'Preschool Teachers' Training, Beliefs and Practices Concerning Mathematics in Pre-Schools in the UK: Implication for Education and Practice'. *PysArXiv,* https://doi.org/10.31234/osf.io/rdx6c.

Davis-Kean, P.E., Domina, T., Kuhfeld, M., Ellis, A. and Gershoff, E.T. (2021) 'It Matters How You Start: Early Numeracy Mastery Predicts High School Math Course-Taking and College Attendance'. *Infant and Child Development,* 31(2): e2281. https://doi.org/10.1002/icd.2281.

DfE (Department for Education). (2017) Primary Assessment in England: Government Consultation Response. Available at: https://www.gov.uk/government/consultations/primary-assessment-in-england (accessed 17 January 2023).

DfE. (2021a) Statutory Framework for the Early Years Foundation Stage: Setting the Standards for Learning, Development and Care for Children From Birth to Five. Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/974907/EYFS_framework_-_March_2021.pdf (accessed 17 January 2023).

DfE. (2021b) Development Matters: Non-Statutory Curriculum Guidance For the Early Years Foundation Stage. Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1007446/6.7534_DfE_Development_Matters_Report_and_illustrations_web__2_.pdf (accessed 17 January 2023).

DfE. (2022a) Opportunity For All: Strong Schools With Great Teachers For Your School. Available at: https://www.gov.uk/government/publications/opportunity-for-all-strong-schools-with-great-teachers-for-your-child (accessed 26 July 2022).

DfE. (2022b) 'Early Years Foundation Stage Profile: 2023 Handbook'. Available at: https://www.gov.uk/government/publications/early-years-foundation-stage-profile-handbook (accessed 17 January 2023).

EEF (Education Endowment Foundation). (n.d.) Pipeline of EEF Trials: Stages of the EEF Project Pipeline. Available at: https://educationendowmentfoundation.org.uk/projects-and-evaluation/evaluation/process-and-people/pipeline-of-eef-trials (accessed 23rd February 2023).

EEF. (n.d.) Early Years Toolkit. Available at: https://educationendowmentfoundation.org.uk/education-evidence/early-years-toolkit (accessed 24 August 2023).

EEF. (2018) Statistical analysis guidance for EEF evaluations: March 2018.

EEF. (2019) 'Longitudinal Analysis of EEF Trials'. London: EEF. Available at: https://educationendowmentfoundation.org.uk/public/files/Grantee_guide_and_EEF_policies/Evaluation/Writing_a_Protocol_or_SAP/longitudinal_guidance.pdf (accessed 25 January 2023).

EEF. (2021) 'Improving Mathematics in the Early Years and Key Stage 1: Guidance Report'. London: EEF. Available at: https://d2tic4wvo1iusb.cloudfront.net/eef-guidance-reports/early-maths/EEF_Maths_EY_KS1_Guidance_Report.pdf?v=1674116008 (accessed 17 January 2023).

Farquharson, C., McNally, S., Tahir, I., (2022), 'Education Inequalities', IFS Deaton Review of Inequalities, Available at: **https://ifs.org.uk/inequality/chapter/education-inequalities/** (accessed 24 January 2023).

Fox, J., Kleiber, C. and Zeileis, A. (2021) ivreg: Instrumental-Variables Regression by '2SLS', '2SM', or '2SMM', with Diagnostics. R Package Version 0.6-1. Available at: https://CRAN.R-project.org/package=ivreg (accessed 23 November 2021).

Hodgen, J., Barclay, N., Foster, C., Gilmore, C., Marks, R. and Simms, V. (2020) Early Years and Key Stage 1 Mathematics Teaching: Evidence Review. London: EEF. Available at: https://educationendowmentfoundation.org.uk/public/files/Early_Years_and_Key_Stage_1_Mathematics_Teaching_Evidence_Review.pdf (accessed 17 January 2023).

Husain, F., Chidley, S., Piggott, H., Averill, P., Basi, T., Gilbert, A., Comanaru, R., Fenton, C. and Corteen, E. (2019) Early Years Foundation Stage Profile (EYFSP) Reforms: Pilot Report. London: EEF. Available at: https://educationendowmentfoundation.org.uk/public/files/Projects/Evaluation_Reports/Early_Years_Foundation_Stage_Profile_(EYFSP)_Reforms.pdf (accessed 17 January 2023).

IES (Institute of Education Sciences). (2013) Teaching Math to Young Children, Educator's Practice Guide. Available at: https://ies.ed.gov/ncee/wwc/Docs/PracticeGuide/early_math_pg_111313.pdf (accessed 17 January 2023).

James, F. (2022) Q&A: The New EYFS Framework. Available at: https://www.yorkshireridingsmathshub.co.uk/news/qa-the-new-eyfs-framework/ (accessed 19 January 2023).

Jiang, Y. and Small, D. (2014) ivpack: Instrumental Variable Estimation. R Package Version 1.2. Available at: https://CRAN.R-project.org/package=ivpack (accessed 23 November 2021).

Knowles, C. (2017) 'Closing the Attainment Gap in Maths: A Study of Good Practice in Early Years and Primary Settings'. Brixton: Fair Education Alliance. Available at: https://static1.squarespace.com/static/543e665de4b0fbb2b140b291/t/58aaeac429687f223f0ff369/1487596235907/FEA+Numeracy+Report_FV.pdf (accessed 24 January 2023).

National Foundation for Educational Research (NFER). (2019) 'Reception Baseline Assessment: Privacy Notice'. Bonn: NFER. Available at: https://www.nfer.ac.uk/media/4554/rba_pupil_data_privacy_notice.pdf (accessed 24 January 2023).

National Statistics. (2022) Schools, Pupils and Their Characteristics. Available at: https://explore-education-statistics.service.gov.uk/find-statistics/school-pupils-and-their-characteristics (accessed 17 January 2023).

NCETM (National Centre for Excellence in the Teaching of Mathematics). (2021) MASTERING NUMBER: Supporting Pupils in Reception, Year 1 and Year 2 to Develop Good Number Sense. Available at: https://www.ncetm.org.uk/maths-hubs-projects/mastering-number/ (accessed 16 January 2023).

Norton, S.J. (2017) 'Primary Mathematics Trainee Teacher Confidence and its Relationship to Mathematical Knowledge'. *Australian Journal of Teacher Education*, 42(2): 47–61. http://dx.doi.org/10.14221/ajte.2017v42n2.4

Nunnally, C. and Bernstein, I.H. (1994) 'The Assessment of Reliability'. *Psychometric Theory*, 3: 248–92.

Outhwaite, L.A., Anders, J. and Van Herwegen, J. (2022a) 'Mathematics Attainment Falls Behind Reading in the Early Primary School Years'. CEPEO Working Paper No. 22–06. London: Centre for Education Policy and Equalising Opportunities, UCL. Available at: https://repec-cepeo.ucl.ac.uk/cepeow/cepeowp22-06.pdf (accessed 26 July 2022).

Outhwaite, L., Ranzato, E. and Van Herwegen, J. (2022b) Raising Mathematical Attainment From the Early Years. Impact, Issue 16. Available at: https://my.chartered.college/impact_article/raising-mathematical-attainment-from-the-early-years-2/ (accessed 24 January 2023).

Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D. and R Core Team. (2021) nlme: Linear and Nonlinear Mixed Effects Models. R Package Version 3.1–153. Available at: https://CRAN.R-project.org/package=nlme (accessed 10 December 2021).

Poet, H., Styles, B., Miani, G., Andrade, J. and Ghosh, A. (2021). Randomised Controlled Trial Evaluation of the White Rose Maths Reception Jigsaw: Evaluation Protocol. London: EEF. Available at: https://educationendowmentfoundation.org.uk/projects-and-evaluation/projects/reception-jigsaw (accessed 16 January 2023).

Ritchie, S. J., and Bates, T. C. (2013) 'Enduring Links From Childhood Mathematics and Reading Achievement to Adult Socioeconomic Status'. *Psychological Science*, 24, 1301– 1308.

Robinson-Smith, L., Fairhurst, C., Stone, G., Bell, K., Elliott, L., Gascoine, L., Hallett, S., Hewitt, C., Hugill, J., Torgerson, C., Torgerson, D., Menzies, V. and Ainsworth, H. (2018). 'Maths Champions: Additional Appendices'. London: EEF Available at: https://educationendowmentfoundation.org.uk/public/files/Maths_champions-appends.pdf (accessed 27 July 2021).

Russo, J., Bobis, J., Sullivan, P., Downton, A., Livy, S., McCormick, M. and Hughes, S. (2020) 'Exploring the Relationship Between Teacher Enjoyment of Mathematics, Their Attitudes Towards Student Struggle and Instructional Time Amongst Early Years Primary Teachers'. *Teaching and Teacher Education*, 88: 102983. https://doi.org/10.1016/j.tate.2019.102983

Schwendel, G. (2023). Randomised Controlled Trial: Evaluation of the White Rose Maths Reception Jigsaw Statistical Analysis Plan – longitudinal follow-up. London: EEF. Available at: **https://educationendowmentfoundation.org.uk/projects-and-evaluation/projects/reception-jigsaw** (accessed 16 September 2023).

Taggart, B., Sylva, K., Melhuish, E., Sammons, P. and Siraj, I. (2015) Effective pre-school, primary and secondary education project (EPPSE 3-16+): How pre-school influences children and young people's attainment and developmental outcomes over time. Research Brief June 2015. Available at: **https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/455670/RB455_Effective_pre-school_primary_and_secondary_education_project.pdf.pdf** (accessed 23 January 2023).

WRM (White Rose Maths). (n.d.) Timeline. Available at: https://whiterosemaths.com/who-we-are/our-story#more (accessed 17 January 2023).

Appendix A: The EEF cost rating

Costs

Appendix Table 1: Cost rating

Cost rating	Description
£ £ £ £ £	*Very low:* less than £80 per pupil per year.
£ £ £ £ £	*Low:* up to about £200 per pupil per year.
£ £ £ £ £	*Moderate:* up to about £700 per pupil per year.
£ £ £ £ £	*High:* up to £1,200 per pupil per year.
£ £ £ £ £	*Very high:* over £1,200 per pupil per year.

Appendix B: Security classification of trial findings

Rating	Criteria for rating			Initial score		Adjust		Final score
	Design	**MDES**	**Attrition**					
5	Randomised design	<= 0.2	0-10%					
4	Design for comparison that considers some type of selection on unobservable characteristics (e.g. RDD, Diff-in-Diffs, Matched Diff-in-Diffs)	0.21 - 0.29	11-20%	4				4
3	Design for comparison that considers selection on all relevant observable confounders (e.g. Matching or Regression Analysis with variables descriptive of the selection mechanism)	0.30 - 0.39	21-30%			Adjustment for threats to internal validity **[X]**		
2	Design for comparison that considers selection only on some relevant confounders	0.40 - 0.49	31-40%					
1	Design for comparison that does not consider selection on any relevant confounders	0.50 - 0.59	41-50%					
0	No comparator	>=0.6	>50%					

Threats to validity	Risk rating	Comments
Threat 1: Confounding	Moderate	Randomisation was appropriate and conducted by an independent statistician. But there is an important imbalance among pupil characteristics. Imbalance was low for the EN baseline measure but there was imbalance of FSM with higher proportions of FSM-eligible pupils in the control group. (Note: the % of FSM was lower than anticipated so a sampling approach was used to boost this).
Threat 2: Concurrent Interventions	Moderate	This disparity in participation could raise concerns around concurrent intervention bias, but it is more likely to underestimate the impact instead of overestimating the impact that the intervention had.
Threat 3: Experimental effects	Moderate	There was some contamination as almost two-fifths of mathematics leads (39%) reported that there had been changes to their mathematics scheme or published resources used to guide mathematics teaching in reception. Among those, a third of the schools from the control group (32%) started using schemes/resources from WRM, suggesting similarity to the Reception Jigsaw.
Threat 4: Implementation fidelity	Low	There were some adaptations to delivery resulting from the Covid-19 situation. However, any adaptations did not affect the fidelity with which the programme was delivered.
Threat 5: Missing data	Moderate	Missing data was moderate at 14% but this was the only variable associated with missing outcomes. Sensitivity analysis was not required as the EN baseline is included in the substantive model as a covariate.
Threat 6: Measurement of outcomes	Low	Well known measures.
Threat 7: Selective reporting	Low	The trial was registered, and all analyses were conducted as specified in the protocol and SAP.

- **Initial padlock score:** 4
- **Reason for adjustment for threats to validity:** [-0] Padlocks- There is no evidence of serious threats to the internal validity of the study
- **Final padlock score:** Initial score adjusted for threats to validity = **[4] Padlocks** due to attrition being 13.9%

Appendix C: Effect size estimation

Appendix Table 2: Effect size estimation

			Intervention group		Control group			
Outcome	**Unadjusted differences in means**	**Adjusted differences in means**	**n (missing)**	**Variance of outcome**	**n (missing)**	**Variance of outcome**	**Pooled variance**	**Population variance (if applicable)**
PUMA Total Score (primary measure)	0.67	0.45	952 (124)	33.73	1,365 (248)	29.47	31.22	N/A
Confidence in teaching mathematics (secondary measure II)	3.82	3.37	135 (99)	39.05	182 (160)	24.73	30.82	N/A
Confidence in mathematics ability (secondary measure III)	0.34	0.12	135 (99)	9.88	181 (161)	10.23	10.08	N/A
PUMA 20 score (secondary measure IV)	0.43	0.27	952 (124)	16.05	1,365 (248)	14.56	15.17	N/A
PUMA (Year 1) Total Score (secondary measure 5a)	0.73	0.51	905 (171)	31.00	1360 (253)	27.93	29.16	N/A
PUMA (Year 1) Total Score (secondary measure 5b)	0.73	0.30	905 (171)	31.00	1360 (253)	27.93	29.16	N/A

Further Appendices

Please see accompanying document '*Further appendices*'.

This document is available for download at https://educationendowmentfoundation.org.uk

The Education Endowment Foundation
5th Floor, Millbank Tower
21–24 Millbank
London
SW1P 4QP

https://educationendowmentfoundation.org.uk
@EducEndowFoundn
Facebook.com/EducEndowFoundn