SPAtial Cognition to Enhance mathematical learning (SPACE)

Pilot Report

March 2025

Amy Hall, Paula Verdugo, Katherine Young, Ann Dowker, Jane Lewis

The Education Endowment Foundation (EEF) 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 to 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

Contents

About the evaluator

SPAtial Cognition to Enhance mathematical learning (SPACE) was independently evaluated by a team from the Centre for Evidence and Implementation (CEI): Jane Lewis; Amy Hall; Dr Katherine Young; and Dr Paula Verdugo. The evaluation team was also advised by Dr Ann Dowker, University Research Lecturer at the University of Oxford, Department of Experimental Psychology.

CEI is a global, not-for-profit evidence intermediary dedicated to using the best evidence in practice and policy to improve the lives of children, families, and communities facing adversity. Established in Australia in late 2015, CEI is a multidisciplinary team across five offices in London, Melbourne, Sydney, Singapore, and Oslo. CEI works with clients, including policymakers, governments, practitioners, programme providers, organisation leaders, philanthropists, and funders in three key areas of work:

- understand the evidence base;
- develop methods and processes to put the evidence into practice; and
- trial, test, and evaluate policies and programmes to drive more effective decisions and deliver better outcomes.

For more information about CEI visit our website at www.ceiglobal.org

The lead evaluator was Jane Lewis.

Contact details:
Jane Lewis
Centre for Evidence and Implementation
Albany House,
London, SW1H 9EA

Email: jane.lewis@ceiglobal.org

Glossary

ANCOVA	analysis of covariance
ANOVA	analysis of variance
BLOCS	Block Construction Skills for Mathematics
CEI	Centre for Evidence and Implementation
CoGDev	Cognitive, Genes and Developmental Variability
DF	degrees of freedom
EAL	English as an Additional Language
EEF	Education Endowment Foundation
EYFS	Early Years Foundation Stage
FSM	free school meals
GDPR	General Data Protection Regulation
ID	identification
IDEA	Intervention, Delivery, and Evaluation Analysis
LEA	Local Education Authority
m-ANOVA	mixed-analysis of variance
MoU	Memorandum of Understanding
Ofsted	Office for Standards in Education, Children’s Services and Skills
PiM	Progress in Mathematics
SAT	Standardised Assessment Test
SD	standard deviation
SE	standard error
SEND	special educational needs and disabilities
SLT	senior leadership team
SPACE	SPAtial Cognition to Enhance mathematical learning
STEM	science, technology, engineering, and maths

Executive summary

The project

The SPAtial Cognition to Enhance mathematical learning (SPACE) programme was developed by the University of Surrey to improve the spatial thinking and maths outcomes of six to seven-year-old (Year 2) pupils. Through structured LEGO® sessions within maths lessons, the programme aims to improve pupils' spatial language, spatial thinking skills, and ultimately their maths skills. This is achieved through teaching training and support that aims to increase their confidence, understanding, and perceived importance of spatial reasoning for maths.

This pilot evaluation involved 15 schools, with one Year 2 class teacher and teaching assistant (TA) taking part in each school, and a total of 415 pupils. Teachers were trained via a half-day in-person training session and provided with three 30-minute support sessions, in addition to online and email support throughout the programme. Training materials included prompt cards, training videos, and a delivery manual.

SPACE is designed to be delivered as twelve sessions of 30 minutes each across six weeks, as a whole-class intervention. In each session, every pupil was given a box of LEGO® bricks and a model booklet, which contained six diagrams of exploded models (showing the component pieces separated and indicating how they fit together) for pupils to build. Teaching staff were asked to use spoken language to support pupils to complete the models, rather than physically manipulating the pieces and to encourage children to work independently.

This pilot evaluation was designed to assess the feasibility of implementation, evidence of promise, and readiness for an efficacy trial of the SPACE programme. It involved observations of training, support sessions, intervention delivery, and pupil assessments; interviews with teachers, TAs and senior leadership; and focus groups with the delivery team and school staff surveys. This pilot was undertaken during the 2023/2024 academic year with delivery of the programme between September 2023 and December 2023 and data collection between November 2023 and April 2024. The evaluation was carried out by the Centre for Evidence and Implementation (CEI). The project was funded by the Department for Education's Accelerator Fund, as part of the Education Endowment Foundation (EEF)-led initiative to build the evidence about programmes that show promise for increasing pupil attainment.

Key findings

Table 1: Summary of pilot findings

Research question	Findings
Is the SPACE programme feasible for implementation?	The SPACE programme was well received by teaching staff, who found it was easy and enjoyable to deliver. The programme was delivered as intended, with sufficient training, resources, and support provided. Teachers and TAs suggested pupils engaged well with the programme, including those from disadvantaged backgrounds, maintaining focus and motivation throughout the sessions. However, concerns were expressed by school staff that SPACE took time away from their usual maths teaching. While they saw value in learning using physical manipulatives, teachers were unsure how the programme linked with the wider maths curriculum.
Is there evidence of promise?	Teaching staff reported that the SPACE programme had improved their confidence and understanding of teaching spatial reasoning and use of spatial language, in line with the Theory of Change. Teaching staff perceived positive improvements in pupils' spatial language and spatial skills, which applied to pupils from all backgrounds. They also highlighted the perceived positive effects on pupils considered to have fewer opportunities to engage with physical manipulatives in the home. There were mixed views from teaching staff on whether perceived improvements in pupils' spatial language and reasoning leads to improved maths attainment.
Is the SPACE programme ready to be evaluated in a trial?	Development work will be required to adapt the delivery model for delivery at scale. Two possible pathways to scale have been identified through this evaluation, depending on the long-term goal for the programme. First, adaptations are recommended to the intervention to better integrate the programme with existing teaching. Second, some more extensive developments to the delivery model (such as a train-the-trainer process) are recommended to facilitate scale up. The pilot showed that the whole-class assessment measures used in this study are not appropriate for use in their current form. Although the results indicate that they can detect change over the intervention period, the tests took significantly longer than intended and pupils were unfamiliar with the content, making them unacceptable to teachers and pupils.

Additional findings

Schools were positive about the self-contained nature of the SPACE programme and training, and the resources and support were well received. Staff felt pupils engaged well with the programme, including those from disadvantaged backgrounds.

The programme was feasible for delivery overall, but the evaluators concluded that an area of challenge was integrating it into normal maths lessons so that it is seen by teaching staff as a way of teaching maths rather than being viewed as ‘additional to’ the planned curriculum. The evaluators identified some mitigations for this issue: earlier recruitment of schools would allow for further support to build SPACE into planned teaching in advance of the start of the school year; more direct linking of spatial thinking and the school’s maths curriculum in the training and session content could increase the comfort of delivery for teaching staff; delivery in the Spring Term when most schools teach ‘Space and Shape’ could help integrate the topics. Additionally, the storage of large LEGO® boxes was challenging in some schools due to limited space.

Teacher- and pupil-level outcomes show promising evidence and support for several elements of the Theory of Change. Notably, there is evidence that the programme improved teaching staff’s perceived understanding of spatial reasoning and confidence in using spatial reasoning teaching strategies in the classroom. There were also perceived improvements for pupils’ spatial language and spatial reasoning skills. However, there were mixed views from teaching staff on how spatial language and spatial reasoning are linked to maths skills.

In addition to the key findings, teachers reported that the programme was seen to have a positive effect on fine motor skills, resilience, perseverance, and confidence—areas not described in the Theory of Change. They also noted that the programme supported the transition from continuous provision into the more structured learning environment in Year 2.

A variety of potential causal mechanisms for change in teacher and pupil outcomes were identified through this pilot evaluation. Teaching staff’s increased awareness of the importance of spatial skills and the effect of their language on children’s learning was thought to increase their use of spatial prompts in the SPACE sessions. This appears to have led to more sustained changes in teaching strategies to incorporate spatial thinking more generally in the classroom. For pupils, the repetition of the model-building task was thought to increase confidence and engagement in maths and their exposure to spatial language led to them using it with teachers and peers.

The evaluators make two recommendations, which would help prepare the programme ready for scale up:

1. The first is to test an initial set of modifications, which would be deliverable at efficacy trial scale but not beyond. This would involve adaptations to address the concerns regarding how the programme is integrated into maths teaching and streamlining the support and monitoring processes as well as ongoing work by the delivery team to promote spatial reasoning with policymakers.

2. The second recommendation is to test a fully scalable model in an efficacy trial. This would involve these adaptations plus more significant investment in developing the programme for sustainable scale, including developing a less logistically challenging delivery modality or securing partnerships that provide the required logistical support and developing a train-the-trainer model and network of ‘champions’ to support delivery.

Introduction

Background evidence

Many adults experience difficulties with maths, and low numeracy is consistently found to have negative social and economic consequences for both the individual and society (Gross *et al.*, 2009; Pro Bono Economics, 2014; Ritchie and Bates, 2013). The development of early maths knowledge is critical to later mathematical skills (Davies-Kean *et al.,* 2022; Chu *et al.*, 2016; Dowker, 2005) and children from disadvantaged families start to fall behind from an early age (Ayala *et al.,* 2024), indicating the importance of early interventions to prevent or ameliorate such difficulties (Dowker, 2017; Janus and Duku, 2007; Rojo *et al.,* 2024; Svane *et al.,* 2023).

One promising area of intervention to address this gap in early maths skills is training in spatial reasoning. Spatial ability involves being aware of the location and dimension of objects and their relationships to one another and is an important part of children's mathematical development (Farran *et al.,* 2024). Spatial ability has a number of key components including:

- navigational ability;
- shapes and shape properties;
- mental transformation;
- visuo-spatial working memory;
- spatial numeric linear relationships; and
- spatial relationships in two-dimensional and three-dimensional space.

While the evidence from studies in the UK and internationally form different conclusions regarding, which aspects of spatial ability are most predictive of, which mathematical abilities at different ages, they converge in indicating that spatial ability is important to maths. Recent evidence from the UK, Europe, and USA suggests a relationship (medium to large effect size) between spatial ability (including spatial working memory, mental rotation, and other spatial transformations), and performance in arithmetic (Allen and Dowker, 2022; Cornu *et al.,* 2018; Mix *et al.,* 2021). Longitudinal studies suggest that mental rotation and other spatial skills at or before school entry may predict later mathematical skills (Frick, 2019; Gilligan *et al.,* 2019; Mix *et al.,* 2016; Zhang and Lin, 2015).

Hawes and Ansari (2020) discussed four possible reasons why spatial and mathematical ability might be related:

- numbers and numerical relationships are often mentally represented spatially, from the approximate representation of quantities on an internal number line to the more complex mental representation of mathematical problems;
- spatial and mathematical thinking use overlapping brain areas;
- education often involves the use of spatial models, such as number lines and number frames; and
- visual-spatial working memory is very important for mathematical thinking.

The authors conclude that these explanations are not mutually exclusive alternatives, but operate together, interact with each other, and reinforce each other.

One aspect of spatial ability that has received some attention recently is the use of spatial language. Several studies suggest a strong relationship between spatial language and other spatial abilities (Brainin *et al.,* 2022; Pruden *et al.,* 2011), and there is increasing evidence for a relationship between young children's spatial language and their mathematical abilities (Lindner *et al.,* 2022). In a study of preschool children, Georges *et al.* (2023) found that children's spatial language, in particular the use of locative prepositions, predicted ability in forward and backward counting, symbolic magnitude judgements and ordinal judgements even after controlling for other spatial abilities, phonological awareness, age, socio-economic status, and home language. In a study of primary school-age children, Gilligan-Lee *et al.* (2021) found that children's spatial language predicted several components of spatial skills, including scaling and perspective-taking. Furthermore, spatial language accounted for variance in the standardised maths outcomes for Progress in Mathematics (PiM) test series by the National Foundation for Educational Research, even when spatial skills and receptive vocabulary skills were controlled for. Exposure to spatial language in the classroom through teachers' use of it when describing spatial properties and relationships is therefore, likely to be a crucial component of any intervention, which aims to improve spatial skills.

One crucial question is whether spatial training can improve mathematical skills. The fact that spatial and mathematical skills are correlated, and even that spatial skills predict mathematical skills longitudinally, does not automatically imply that spatial training improves maths. In order to use such training to achieve relevant change, more evidence is needed both that spatial skills can be taught and that such teaching improves maths.

In relation to whether spatial skills can be taught, Uttal *et al.* (2013) carried out a meta-analysis of over 240 studies of spatial skill training, which used a variety of techniques with children, adolescents, and adults. The results showed that such training had significant effects on spatial skills and that effectiveness was not dependent on age, gender, or precise form of training. Yang *et al.* (2020) conducted a meta-analysis specifically of studies of spatial interventions in young children from birth to eight years and found the effects of such training on spatial skills to be significant.

There have been relatively few studies conducted internationally on the impact of spatial training on children's long-term maths outcomes. However, those undertaken, usually using physical manipulatives, have generated promising results (Cornu *et al.,* 2017; Hawes *et al.,* 2022; Gilligan-Lee, Hawes *et al.,* 2023; Lowrie *et al.*, 2017). This appears to be particularly the case for children from disadvantaged families (Bower, Zimmerman *et al.,* 2020; Schmitt *et al.,* 2018; Verdine *et al.,* 2014), perhaps because such children are less likely to have already been exposed to materials and activities (such as LEGO® play) that may help to develop spatial awareness and therefore, have more scope for development. These findings highlight the potential importance of interventions to promote the development of spatial abilities, and particularly the potential role of these approaches in narrowing the maths attainment gap.

Despite previous research demonstrating the potential importance of spatial skills to children's understanding of the world around them, spatial reasoning is not a current focus of policymakers in the curriculum (Moss *et al.,* 2016). Indeed, in 2018, shape, space, and measures were removed as learning goals from the Early Years Foundation Stage (EYFS) statutory framework for children in England. Although these goals have not been reintroduced, their importance to children's mathematical development was acknowledged by the Department for Education in 2020 (Gibbons, 2020) by updating the EYFS statutory framework to include the importance of opportunities for children to learn about shape, space, and measures. Furthermore, spatial reasoning skills and the use of manipulatives are described as important to children's development in all areas of maths in the current iteration of the EYFS statutory framework (Department for Education, 2024). Furthermore, while spatial thinking is not explicitly included in the Key Stage 1 Ready-to-Progress criteria produced by the Department for Education, the manipulation of shapes in terms of size and orientation is (Department for Education, 2020). These recent changes imply a shift in policy towards recognising the value of spatial reasoning for children's development, though the evidence suggests that the benefits require training in spatial skills beyond the typical geometry curriculum. Early years and maths specialists argue that embedding spatial reasoning in the early years and primary curriculum more explicitly would benefit all children, and particularly those from lower socio-economic status families (Bower, Zimmerman *et al.,* 2020; Gifford, 2019; Gilligan-Lee *et al.,* 2022).

The Block Construction Skills for Mathematics (BLOCS) study in the UK explored the impact of a previous iteration of the 'SPAtial Cognition to Enhance mathematical learning' (SPACE) intervention. In the BLOCS study, 198 children in Year 3 and Year 4 (aged seven to nine) completed a slightly more advanced version of the SPACE intervention (based on building LEGO® models) in small groups as a lunchtime club between January 2022 to December 2022, delivered by the class teacher or a postgraduate student trained by a researcher. This study focused on the impact of building the models (both physically and digitally) from visual instructions on children's spatial skills, mathematical skills, and LEGO construction ability. They identified an impact on LEGO construction ability and some evidence for transfer to arithmetic skills, as measured by a pre- and post-intervention comparison of the numeracy subtest of the Wechsler Individual Achievement Test III (Wechsler, 2001). The experience and results from the BLOCS study have informed the development of the SPACE intervention and its pilot evaluation, described here in the next section 'Intervention' below.

Intervention

SPACE is a whole-class structured approach to LEGO® play, which aims to improve six to seven-year-old (Year 2) children's spatial thinking and maths outcomes. It was developed by a team at the University of Surrey, following their experience of implementing BLOCS. The Theory of Change diagram is illustrated in Figure 1 below.

The aim of the programme is to improve teaching staff's confidence, understanding and perceived importance of spatial reasoning, and increase their use of spatial prompts with their Year 2 pupils. For children, their participation in the SPACE LEGO® sessions was anticipated to increase their exposure to spatial language (through teachers' use of spatial prompts) and draw on their spatial thinking skills, such as visualisation, to build the models. These processes, in turn, are expected to improve children's comprehension and production of spatial language and their spatial reasoning ability, including mental rotation, visuo-spatial working memory, embedding, and spatial-numerical relationships. In the longer term, these outcomes are anticipated to lead to sustained improvements in children's spatial language and spatial reasoning and, subsequently, their mathematical ability.

In this pilot evaluation, teachers and teaching assistants in 15 schools were trained to deliver SPACE, with teachers leading and, where possible, teaching assistants supporting delivery in each school. Teaching staff training was delivered by the University of Surrey delivery team and consisted of a half-day in-person training session. The delivery team provided each participating class with a set of support materials and fortnightly support with a mixture of compulsory and optional sessions. Teachers were offered three compulsory 30-minute support sessions across the six-week programme, in which they provided monitoring information, discussed their experiences of delivery, and received delivery support from the delivery team. In the intervening three weeks, teachers were also offered optional online support sessions, or they could opt for email support. Teachers were also given resources to help them encourage children's spatial thinking and spatial language during SPACE sessions. These included spatial thinking prompt cards, online training videos, and a comprehensive delivery manual. Delivery materials also consisted of weekly themed two-minute videos for teachers to use to introduce the 'theme' (e.g. scientists, explorers) of the sessions each week.

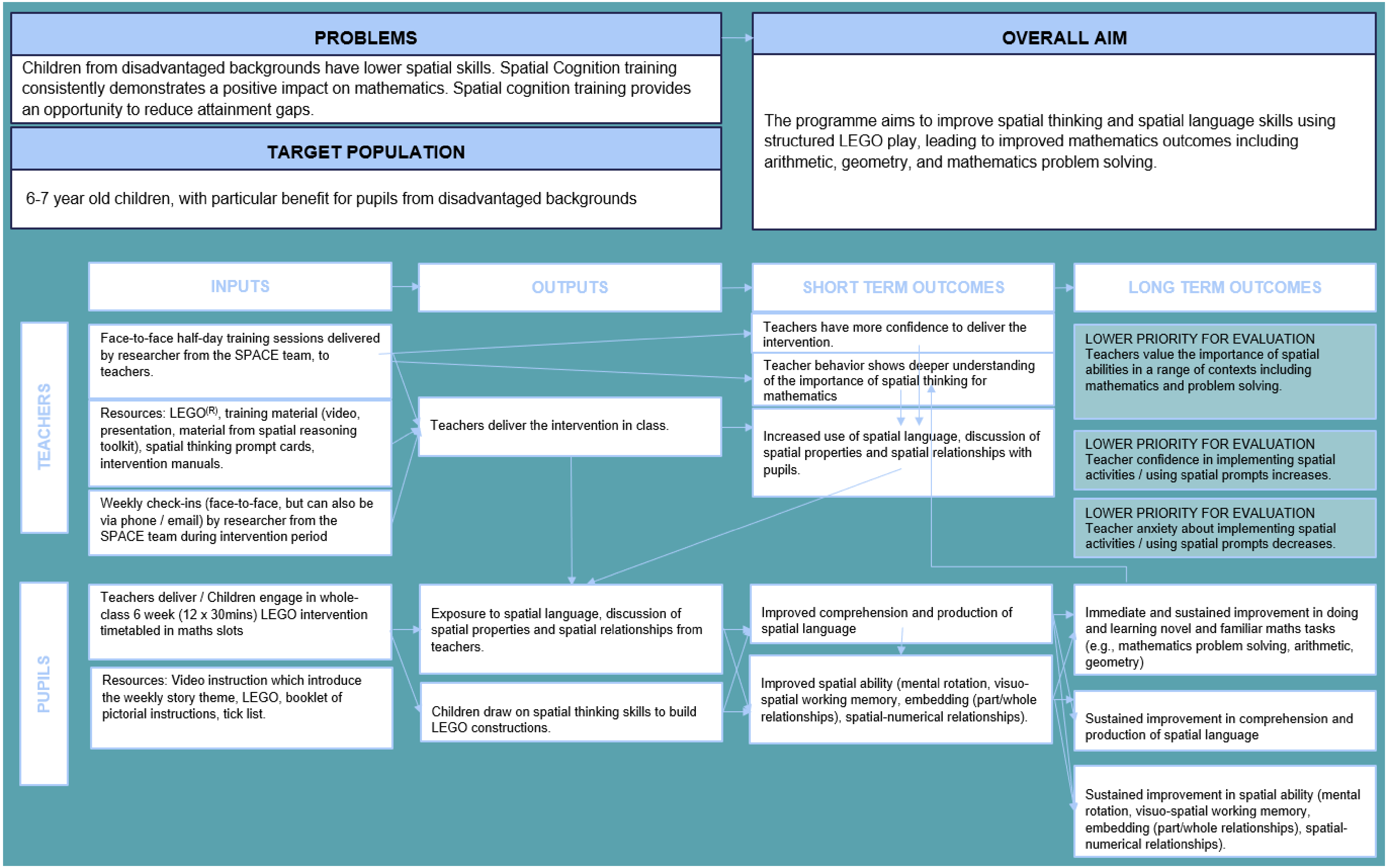
Each class taking part in the evaluation was provided with LEGO® packs for each child, and booklets containing the visual instructions.

The SPACE programme was delivered during maths lessons by teachers and teaching assistants in two weekly 30-minute sessions per week for six weeks (12 sessions in total). SPACE sessions follow a common format. Each session begins with the teacher or teaching assistant presenting the 'theme' for the session, and presentation of the two-minute video. For the main activity, each child was given a model booklet to follow, containing a series of exploded diagrams of LEGO® models of familiar objects (such as a crown or a tree), which the children then built out of LEGO bricks. The models show the relative position of the pieces by separating them, with arrows indicating how they fit together. Children were asked to build up to six LEGO models per session, working through the models in sequence, independently and at their own pace. Teachers and teaching assistants were asked to use their judgement on when to provide support and to do so via spoken prompts rather than physical support to encourage children to use spatial thinking to problem-solve and to expose children to spatial language.

The 'model booklet' given to each child consisted of the pictorial exploded models to follow as well as a tick box page for children to tick the models they completed per session. The tick boxes were intended to act both as motivation and a way of monitoring task completion, though children were asked to focus on accuracy rather than speed and were not expected to finish all models in each session.

Sessions took place in normal school classrooms, within the regular maths timetable. SPACE was designed to complement and support the maths curriculum in the following areas: Geometry—properties of shapes; Geometry—position and direction; Measurement; Fractions; Place value; Multiplication; Addition; and Subtraction.

Figure 1: Theory of Change



Research questions

The pilot evaluation focused on three areas: feasibility of implementation; evidence of promise; and readiness for an efficacy trial. This evaluation explores ten research questions under these three areas.

Feasibility of implementation:

1. Is SPACE feasible for an implementation in maths lessons?
2. Are the training and support provided sufficient for teachers and teaching assistants to implement SPACE?
3. Was SPACE implemented as intended?
4. Is SPACE acceptable to teaching staff?
5. Do children, particularly those from disadvantaged backgrounds, engage well with SPACE?

Evidence of promise:

6. Do the training and delivery of SPACE sessions lead to changes in teaching staff's understanding of the importance of spatial skills, use of spatial language, and confidence in teaching spatial skills?
7. What impacts do teaching staff perceive for children, particularly those from disadvantaged backgrounds (e.g. spatial language, spatial skills, and maths skills and attainment)?
8. What are the perceived mechanisms and possible causal chains?

Readiness for an efficacy trial:

9. How could SPACE be delivered on a larger scale?
10. Are the proposed whole-class assessment measures appropriate for use in an efficacy trial?

Ethical review

Ethical review was conducted separately for the delivery of the SPACE programme in schools and for the evaluation of the programme.

Ethical approval for the delivery of the SPACE programme in schools was requested by the delivery team and granted by the University of Surrey Ethics Committee in February 2023.

The ethical approval for the evaluation of the SPACE programme was submitted by the evaluation team and positively appraised by the Social Research Association Ethics Appraisal Service in March 2023.

In brief (and described further below), the pilot evaluation involved: a survey of teaching staff at three timepoints; interviews with teaching staff and senior leadership at selected schools; focus groups with the delivery team; observations of teacher training, support sessions, delivery of the intervention and the child assessments; and administrative monitoring data. In addition, the delivery team asked schools to administer child assessments at two timepoints. The evaluation involved assessing their feasibility, including observing their administration.

The headteacher of each school gave consent for the school to be involved in SPACE. Parents of participating classes were provided with information about SPACE (in an information sheet) by the evaluation team via the school and had the right to withdraw their child's assessment data from the study. School staff participants were provided with information sheets by the evaluation team and consented to participate in surveys or interviews, see Appendices B, C, and D. They were provided with information sheets, which explained the research in plain language, and where participants were given the opportunity to ask clarifying questions before providing their data. Participants were provided with clear steps to follow if they wished to end their participation or withdraw their data from the research.
The evaluation team followed the procedures outlined here in the 'Data protection' section below, to protect the anonymity of all research participants and no identifiable information is included in public outputs or reports.

Data protection

Data were processed and stored in line with the General Data Protection Regulation (GDPR) and the UK Data Protection Act 2018. All data procedures were overseen by the Centre for Evidence and Implementation (CEI) independent data protection officer. CEI's legal basis for processing personal data is legitimate interest under UK GDPR Article 6.1 (f).[1] The University of Surrey's legal basis for processing personal data is public task.

The child assessments were completed in paper form, and their name was written on each. These documents were transported by secure courier delivery from schools to CEI, where a unique identifier was assigned to each child. The assessments were marked by CEI and scored, then digitised by a data entry provider before the paper copies were securely destroyed. CEI holds the key linking children's names and their unique identifiers in order to link the child school administered assessments over the two timepoints. This key is password-protected and stored securely and separately from the assessment data.

Pseudonymised survey and assessment data were shared by CEI with the University of Surrey delivery team. Interview data were not shared between the institutions. Personal data will be held by CEI for two years after which point it will be securely destroyed. Personal data held by the University of Surrey will be destroyed after six years.

A data privacy impact assessment was undertaken by CEI. All information was securely stored on a dedicated drive, accessible only to the CEI research team.

Information sheets included a digital link to a Data Privacy Notice as a link and/or attachment, which set out the rights of participants, including their right to withdraw data from the evaluation.

Project team

Evaluation team

CEI is a global, for-purpose evidence intermediary and advisory organisation specialising in the use of evidence-based methods in practice and policy to improve the lives of people facing adversity. The evaluation team's affiliation and responsibilities were as follows:

- **Jane Lewis, CEI, Managing Director.** Project director, involved in all stages of the pilot evaluation, accountable for the overall project.
- **Amy Hall, CEI, Advisor.** Project manager, involved in all stages of the pilot evaluation and responsible for the day-to-day project management, liaison with the delivery team and funder, fieldwork, and reporting.
- **Dr Paula Verdugo, CEI, Advisor.** Project researcher, responsible for the coordination of qualitative data collection and child assessments, analysis of qualitative data, and reporting.
- **Dr Katherine Young, CEI, Principal Advisor.** Project researcher, responsible for the analysis of survey and child assessment data and reporting.
- **Professor Ann Dowker, University of Oxford, University Research Lecturer.** Advisor on this project, providing subject area expertise throughout the set-up, analysis, and reporting stages.

Delivery team

[1] 'Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (United Kingdom General Data Protection Regulation) (Text with EEA relevance)' (2016) *Official Journal* L 679. Available at: https://www.legislation.gov.uk/eur/2016/679/article/6# (accessed 1st August 2024).

The majority of the delivery team are based within the Cognition, Genes and Developmental Variability (CoGDev) Lab at the University of Surrey, led by Professor Emily Farran. The team specialises in spatial thinking in relation to science, technology, engineering, and maths (STEM) in primary school-age children.

- **Professor Emily Farran, University of Surrey, Principal Investigator, Delivery Team Lead.** Designed the intervention and delivered the training.
- **Professor Camilla Gilmore, Loughborough University, Co-Investigator, Delivery Team.** Provided subject area expertise throughout the project.
- **Dr Katie Gilligan-Lee, University College Dublin, Co-Investigator, Delivery Team.** Provided subject area expertise throughout the project.
- **Dr Marija Živković, University of Surrey, Postdoctoral Researcher.** Delivered the training and provided delivery support to half of the participating schools.
- **Dr Anna Korzeniowska, University of Surrey, Research Assistant.** Provided delivery support to half of the participating schools.
- **Rachel Baxter, University of Surrey, Project Officer.** Led the recruitment and onboarding of schools and coordination of delivery materials.
- **Professor Tim Jay, Loughborough University, Co-Investigator,** Delivery Team. Provided subject area expertise throughout the project.
- **Professor Denis Mareschal, Birkbeck, University of London,** Co-Investigator, Delivery Team. Provided subject area expertise throughout the project.
- **Professor Derek Bell, Learnus, Co-Investigator, Delivery Team.** Supported recruitment and provided subject area expertise throughout the project.

Methods

Recruitment and consent

Recruitment for the SPACE intervention programme and the pilot evaluation were undertaken in tandem and are therefore, described as a single process below.

The University of Surrey led the recruitment of schools, seeking to enlist 15 schools to participate and three schools to a waiting list in case any schools withdrew their participation early. Schools were recruited from two geographical areas, namely, Portsmouth and Surrey that were selected based on: i) proximity to the delivery team; and ii) ensuring that at least half of the participating schools were recruited from educational investment areas. A back-up area was selected in case of recruitment challenges, but it was not required.

To be eligible for the trial, schools had to be mainstream state-funded primary schools with at least one Year 2 class. Schools with mixed-year group classes were not eligible to participate. Additionally, primary schools could not be delivering any other maths interventions (beyond usual teaching) in their Year 2 class.

The participating schools were recruited between March 2023 and June 2023. A variety of strategies were used to contact schools: advertisement of the pilot evaluation on the Education Endowment Foundation (EEF) website; outreach to local maths hubs; targeted emails and phone calls by the delivery team or their partners to their existing networks of schools and contacts; and to schools identified through their websites. Schools were asked to complete an expression of interest form to indicate their interest in participating. Table 2 details the number of schools approached through these strategies and the proportion of schools recruited through each strategy.

Table 2: School recruitment strategies

Advertisement channel	# Schools contacted via route	# Expressions of interest via route	Recruited via route[2]	% Schools recruited via route
Northeast Hants and Surrey Maths Hub	372	13	6	29%
The EEF website	409	5	4	19%
Solent Maths Hub	37	4	4	19%
Portsmouth Primary Heads Conference	37	12	4	19%
Individual email from the delivery team to local schools	87	2	2	9%
Personal contacts of the delivery team	4	2	1	5%
Individual emails to schools involved in the intervention development phase	8	1	0	0%
Learnus[3] (partner of the University of Surrey) email to Portsmouth and Surrey primary schools	409	0	0	0%
Portsmouth Education Partnership bulletin	37	0	0	0%
Total	1,400	39	21	100%

The reasons not to participate provided by contacted schools included schools having other priorities, schools embedding new staff, and schools focusing on doing fewer interventions with higher quality. A total of 39 schools

[2] Schools that signed a Memorandum of Understanding (MoU).
[3] Learnus (partner of the Surrey University) is a community that brings together educators and those specialising in the study of the brain, mind, and behaviour to use the evidence from high-quality research to improve and enrich learning for all. Its members include neuroscientists, cognitive scientists, educationalists, psychologists, teachers, policymakers, and commentators.

expressed interest in taking part. Once schools expressed interest in participating, the delivery team selected 15 schools to participate, based on ensuring a spread of characteristics (school type, school size, Office for Standards in Education, Children's Services and Skills [Ofsted] rating, and percentage of pupils eligible for free school meals [FSM]) and feasible distance for the delivery team research assistants to travel. The headteachers from these 15 schools were asked to e-sign an MoU detailing what the intervention and pilot evaluation were about, what the participation would involve, the responsibilities of schools, the delivery team and the evaluation team, and a data sharing and data protection policy. Five schools were also informed they were on a waiting list and would be contacted in the event that one of the initial schools dropped out.

The recruited schools were invited to a 45-minute webinar in June 2023. The session included the background on spatial reasoning, an introduction to the intervention, and an overview of the evaluation activities they would be asked to engage with.

One school withdrew from the trial[4] after one staff member attended the half-day training prior to starting delivery and was replaced by one of the five schools on the waiting list. Table 3 describes the final sample of schools that delivered SPACE as part of this pilot evaluation.

Table 3: Characteristics of schools participating in the pilot evaluation of SPACE

Local authority	School type	Ofsted rating	% of pupils eligible for FSM	School size
Location 1	Local Education Authority (LEA)-maintained	Good	<18%	Small
	LEA-maintained	Good	>25%	Small
	Academy	Good	Between 18% and 25%	Medium
	Academy	Good	>25%	Large
	Academy	Requires improvement	Between 18% and 25%	Small
	Academy	Requires improvement	>25%	Medium
	Academy	Good	Between 18% and 25%	Small
	Academy	Good	>25%	Medium
Location 2	Academy	Good	<18%	Large
	Other type of school	Good	<18%	Large
	Academy	Good	Between 18% and 25%	Medium
	LEA-maintained	Outstanding	Between 18% and 25%	Small
	Academy	Good	>25%	Large
	LEA-maintained	Outstanding	<18%	Small
	LEA-maintained	Good	<18%	Large

There were three forms of informed consent for participation in this evaluation.

[4] According to communications between this school and the delivery team, the reason for withdrawal involved concerns from the school to fit the SPACE sessions within teaching time and meeting their targets, and their decision to prioritise providing support to pupils impacted by Covid-19.

- Approval for the delivery of the intervention and classroom evaluation activities (observations, administering the child assessments) was sought from each school's headteacher and provided by signing the MoU (see Appendix E). Headteacher approval for these activities was considered appropriate for this evaluation because SPACE occurred during regular maths lessons and the content fell within the range of usual curriculum activities.
- All parents/guardians of children in participating classes were informed about the programme by means of an information sheet (Appendices F and G) and given the option to withdraw their child's assessment data from the evaluation (through their right to withdraw) in July 2023 and again in September 2023, in case new children had joined the class in the meantime. To withdraw permission for their child's data being shared, parents could return the form to the class teacher who would remove their child's assessments from those delivered to the evaluation team. As this was a whole-class intervention considered typical for classroom delivery, parents did not have the option to opt-out of SPACE sessions.
- Teaching staff who took part in interviews (senior leadership team [SLT] representatives, teachers, and teaching assistants) received an information sheet (Appendix B) and consent form (Appendix C) and consented either by returning the signed form or verbally (and digitally recorded) prior to their interviews.

Once headteachers had signed the MoU, they were asked to release at least one but preferably two members of the participating teaching staff (including the class teacher, teaching assistants) and invite a member of the SLT to attend a half-day in-person training session. SLT staff attended where possible, based on their availability and interest in the programme, and we identified no pattern in the characteristics of schools who sent an SLT member. Schools were also asked to timetable SPACE lessons to replace normal maths lessons twice a week for six weeks during the 2023 Autumn Term.

The half-day training[5] was delivered to a total of 34 teaching staff (6 SLT members, 16 teachers, and 12 teaching assistants, one of whom was from the school that withdrew prior to delivery) over three dates in September 2023, two in Surrey and one in Portsmouth. Additionally, the teacher from the school that replaced the withdrawn school received individual face-to-face training from one member of the delivery team.

Retention of schools

Table 4: Teaching staff and students per school involved in the pilot study of SPACE[6]

School identification (ID)	Teachers	Teaching assistants	Students
A	1	1	29
B	1	1	29
C	1	1	28
D	1	1	29
E	1	1	29
F	1	1	21
G	1	1	22
H	1	1	27
I	1	1	30
J	1	0	24
K	1	1	30
L	1	1	30
M	1	0	29
N	1	1	30
O	1	1	28
Total	15	13	415

All 15 schools in the participating sample remained engaged in the pilot and delivered the SPACE programme throughout the 2023 Autumn Term as intended. There was no further attrition of schools and no evidence of attrition of students other than due to students joining or leaving classes. Schools were not asked to provide updated enrolment lists and all new students to the classrooms were invited to participate in the intervention and post-intervention assessment. Table

[5] The training manual is available on the following website: https://www.surrey.ac.uk/sites/default/files/2023-08/SPACE-training-manual-final.pdf.
[6] The numbers in this table indicate the initial number of people involved in each classroom in the pilot study. The number of teaching assistants and students initially reported by the teachers may have changed during the intervention period.

4 summarises the number of schools, teaching staff, and students participating in the pilot study by school. Additional children in the participating classes received the intervention but did not take part in evaluation activities.

Data collection

The evaluation followed a mixed-methods research design, collecting qualitative and quantitative data to address questions about the feasibility of implementation, evidence of promise, and readiness for an efficacy trial. The data collection methods used were:

- surveys and individual interviews with teaching staff;
- individual and group interviews with the delivery team;
- observations of training sessions, support sessions, delivery sessions, and child assessments;
- administrative monitoring information; and
- baseline and endline child assessments.

A detailed description of the methods, sampling, and analysis is presented in the following subsections.

Set-up and Theory of Change

In November 2022, teams from the University of Surrey, CEI, and the EEF met for a face-to-face Intervention, Delivery, and Evaluation Analysis (IDEA) workshop at the EEF offices. Two further set-up meetings were held between November 2022 and January 2023 to ensure all teams had a shared understanding of the recruitment, intervention, and evaluation activities.

The Theory of Change (Appendix A) was developed prior to the evaluation by the delivery team with support from the EEF. The Theory of Change was reviewed in the IDEA sessions, but no substantive changes were made. The Theory of Change was designed to include all outcomes that would be explored in an efficacy trial and as such describes the full range of anticipated causal pathways and outcomes, including the direct impact of the SPACE programme on children's mathematical performance. However, as a pilot trial with no counterfactual, these pathways are explored throughout this report through qualitative findings rather than quantitative evidence. As such, child-level outcomes are explored through teaching staff's perceptions and unanticipated outcomes are also described. The IDEA workshop was also used to identify fidelity criteria and success criteria for the pilot trial.

Teaching staff surveys

The purpose of the teaching staff surveys was to assess the feasibility and acceptability of SPACE, teachers' knowledge of, confidence in, and use of teaching spatial skills, and perceptions of SPACE's impacts on children. Surveys were administered at three timepoints: immediately post-training (Timepoint 1); post-intervention delivery (Timepoint 2); and at three-months follow-up (Timepoint 3).

Survey items were piloted with an early year's teaching assistant to ensure clear and appropriate language, ease of completion, estimate the timing, and to check for ceiling effects at baseline. Estimated teaching staff survey timings were also calculated using metrics provided by the Qualtrics survey platform and deemed to be appropriate (7 to 15 minutes). All surveys were reviewed by the delivery team and the EEF before being administered.

- **Timepoint 1.** The baseline survey of teaching staff was completed immediately after the training sessions[7] in September 2023. This survey consisted of 17 items and gathered information about teachers' roles in delivery and their confidence and experience with teaching spatial reasoning prior to the intervention. The survey was designed to take less than seven minutes to complete.
- **Timepoint 2.** A post-delivery survey was distributed immediately after the end of delivery and completion of child assessments in November 2023 and December 2023. This survey consisted of 34 items, covering participants' confidence and practices in teaching spatial reasoning applying adapted items described below, the experience of the SPACE programme implementation,

[7] The survey at Timepoint 1 was initially planned to be delivered at the start of each of the teacher training sessions, to capture teachers' prior knowledge and perceptions. However, due to an administrative error, this survey was instead administered by email immediately after the training.

acceptability and feasibility from standardised implementation outcome measures, bespoke items on their views on the impacts of delivery for them and their Year 2 class, and their experience of administering the child assessments. The survey was designed to take less than 15 minutes to complete.

- **Timepoint 3.** A follow-up survey was distributed towards the end of the Spring Term in March 2024 to April 2024. This survey consisted of 20 items and was designed to understand participants' experience of teaching and supporting spatial skills in their classroom after the intervention and their views about the sustained impacts of the SPACE programme on children's attainment. The survey was devised to take less than ten minutes to complete.

All teaching staff involved in the evaluation were invited to complete each survey by email and sent a maximum of three reminders if they did not respond. The surveys were hosted online in the Qualtrics platform and could be accessed using any standard laptop, tablet, or smartphone. The information gathered from the survey responses was analysed with descriptive statistics and with repeated measures analysis of variance (ANOVA) to establish change over time and presented accordingly in the 'Findings' section below. Due to the sample size limitation, results were interpreted as indicative. Responses to open questions were thematically coded. Table 5 shows the number of participants at each timepoint.

Table 5: Achieved sample per survey

Timepoints	No. of responses	Completion rate of all participating staff N=28
Timepoint 1: Baseline survey (September 2023)	22	79%
Timepoint 2: Post-delivery survey (November 2023–December 2023)	18	64%
Timepoint 3: Follow-up survey (March 2024–April 2024)	16	57%

Teaching staff confidence in delivery and understanding of spatial reasoning

Confidence and the perceived importance of spatial skills were assessed using items adapted from items previously used by the delivery team in previous theoretical research with Reception and Key Stage 1 teachers (Bates *et al.,* 2023). The adapted items included a question on the level of confidence of teaching staff to explain spatial reasoning to someone else and a scale used to measure the perceived importance of spatial reasoning. Both items were adjusted to avoid heavy weighting on the extremes of the scales and therefore, allow for granular responses. Adapted items also included a question on how often teachers and teaching assistants incorporated eight spatial activities into their classroom practice, selected from a longer list from Bates *et al.* (2023) and Gilligan-Lee, Bradbury *et al.* (2023).

Perceived impact on children's numeracy and spatial skills

The evaluation team developed survey items to explore teachers' and teaching assistants' perceptions of the impact of SPACE on children, informed by the delivery team as subject area experts. These items considered the perceived level of improvement in children's comprehension of spatial language, use of spatial language, spatial reasoning, problem-solving, geometry, overall maths skills, engagement in maths learning, and attention and concentration. As previously noted, the pilot trial was not designed to address all elements of the Theory of Change—notably children's performance in numeracy and spatial skills is measured based on teaching staff's perceptions rather than based on direct evidence from children. Indications of change in these areas are therefore, indirect and subjective and are interpreted proportionately in the 'Findings' section below. Some elements explored in the teaching staff surveys were beyond the initial Theory of Change, namely, engagement in maths learning and attention and concentration. These measures are nonetheless useful as indicators of change beyond those anticipated by the delivery team.

Measure of implementation

Implementation effectiveness (based on the dimensions of feasibility and acceptability) was assessed using a set of scales with substantive and discriminant content validity as well as test-retest reliability (ranging from 0.73 to 0.88) (Weiner *et al.,* 2017). Additional questions about specific elements of the implementation strategy were developed by the evaluation team.

Interviews

To gather more in-depth information about the implementation and impact of the intervention, semi-structured interviews were conducted with teachers, teaching assistants, and SLT representatives, as well as group interviews with the delivery team.

Interviews with school staff

Semi-structured interviews took place with school staff to assess the feasibility of implementing the SPACE intervention, barriers and facilitators to implementation, the sufficiency of support and materials provided by the delivery team, and perceptions of the impacts of SPACE on teaching strategies and children's abilities. See Appendix H for interview topic guides. The research team intended to interview a teacher or teaching assistant from each of the 15 participating schools and an interview with an SLT representative in eight selected schools. The allocation of schools to each interview group was purposive, to ensure diversity in school type, size, and location. In schools where a teaching assistant was selected but had not been closely involved with the delivery of the intervention, the class teacher was invited to take part in an interview instead. Table 6 presents the achieved sample of school staff interviews by role. A teacher or teaching assistant was interviewed in all but one of the participating schools.

Table 6: Achieved sample of school staff interviews

School role	No. of interviews
SLT representative	6*
Teacher	11
Teaching assistant	3
Total	20

* *One SLT representative responded to a set of key questions by email rather than in a verbal interview.*

Interviews took between 30 to 60 minutes to complete. Interviews were scheduled by members of the evaluation team with the selected interviewees via email at a time convenient to the participant. Interviews were held online through Microsoft Teams towards the end of delivery (between November 2023 and January 2024).

Delivery team interviews

Group and individual interviews with the core SPACE development team were chosen to explore the shared understandings of the different roles within the delivery team regarding implementation issues. These interviews explored their experiences of recruitment, training and support, implementation challenges, and issues pertinent to the readiness of the programme for a larger efficacy trial. See Appendix J for the interview topic guide. These group interviews had an approximate duration of 60 minutes and took place at three different timepoints. The first interviews were held with the core SPACE development team shortly after the initial teacher training sessions in October 2023. The second timepoint was towards the end of delivery (December 2023) when an interview was held with the delivery team research assistant and postdoctoral researcher to explore implementation issues raised in the fortnightly support sessions. The third timepoint was post-delivery (January 2024), when an interview was carried out with the principal investigator and the delivery team research assistant to discuss their vision and capacity for scaling the intervention. At the first timepoint, two interviews were held in order to focus on specific topics during the set-up and initiation phase. One interview was held with the project officer, focusing on recruitment and onboarding processes and the second interview was held with the principal investigator and postgraduate researcher to explore the training and initial implementation of the intervention. All these interviews were scheduled through email and were carried out online by two members of the evaluation team through Microsoft Teams.

Interview analysis

All interviews were audio recorded with the consent of the interviewees and transcribed verbatim. Once transcribed, thematic content analysis was completed for interview transcripts using the framework approach (Gale *et al.,* 2013; Goldsmith, 2021; Spencer *et al.,* 2013), primarily using deductive analysis (based on the Theory of Change, implementation dimensions, and research questions) and to a lesser extent inductive analysis (based on emerging issues and themes). Data from each interview transcript were systematically extracted and summarised within the themes identified. The summarised and categorised data were then reviewed and synthesised.

Observations

The evaluation team engaged in overt non-participant observations. Four different types of activities were observed: training sessions; delivery support sessions; SPACE classroom sessions; and the delivery of endpoint child assessments. Evaluation team members took handwritten notes in situ to reduce the disruption of the sessions, and observation guides can be found in Appendix K.

Schools were purposively sampled for each type of observation to ensure a representative range of school type, size, Ofsted rating, and location, although recognising that diversity was restricted by the overall sample size of the study sample. In some cases, the initial sample of schools for observation was adapted due to availability constraints.

Training sessions

In September 2023, one member of the evaluation team (AH) attended and observed two in-person training sessions, one in Surrey and one in Portsmouth. These observations sought to provide the evaluation team with a more in-depth understanding of the SPACE programme and serve as a supplement to interview and survey data when assessing the sufficiency of the training provided for teaching staff and experiences of observation. During the observations, the researcher kept a detailed record of the training provided using an observation guideline.

Delivery support sessions

Two members of the evaluation team (AH and PV) observed a total of four in-person support sessions (one evaluator at each session) to understand the sufficiency of support for delivery and identify arising implementation issues.

SPACE sessions

Two members of the evaluation team (AH and PV) observed SPACE delivery sessions (one evaluator at each session). These observations were intended to provide an understanding of how SPACE was delivered in the classroom, including engagement from children and adaptations by teachers, and to assess whether the sessions were implemented as intended. Evaluation team members particularly noted evidence of core fidelity features, including the duration of the sessions, the support strategies used by teachers with the students, the use of spatial prompts, and whether children were working independently.

Endline assessment sessions

In November 2023, two members of the evaluation team (AH and PV) individually observed post-intervention assessment sessions in five schools. In these sessions, evaluation team members observed all three assessment booklets delivered across one, two, and three sessions. These observations aimed to provide an understanding of the feasibility of delivering the child assessments with Year 2 classes within the intended time. Members of the evaluation team particularly noted teachers' use of the assessment scripts, children's engagement with the tests, the sequence of application, and the time spent on each assessment.

Table 7 provides a summary of the activities and number of observations undertaken:

Table 7: Summary of observations

Activity	No. of observations
Training sessions	2
Delivery support sessions	4
SPACE sessions	4
Endline assessment sessions	5
Total	15

All observations were attended by one researcher to avoid overcrowding the spaces. Additionally, keeping to the principle of minimising the burden on schools and participants, none of these observations involved visual or audio recording as this would have been intrusive and required specific parental consent. Observation notes were digitised by the observer for analysis. These were then thematically coded using a deductive approach and analysed descriptively to assess fidelity, implementation issues, and children's overall engagement as well as inter-site variation on these dimensions.

Baseline and endline child maths assessment measures

The evaluation assessed the feasibility of administering the three child outcome measures and involved preliminary validation of the measures to assess their viability for future evaluation.

The assessments were designed and piloted by the delivery team based on pre-existing validated assessment tools, with input from the EEF and the maths expert from the evaluation team. The assessments measured children's maths, spatial language, and mental rotation skills before and after the intervention and were piloted in non-participating schools prior to the evaluation with three Year 2 classes in two schools that had not implemented the intervention by the delivery team postdoctoral researcher.

- **Maths.** The maths assessment involved items adapted from existing standardised assessments and test items commonly used in Year 2 (White Rose, Standardised Assessment Tests [SATs], Key Stage 1 papers, and Test base), including geometry and problem-solving. These items were revised for clarity and appropriateness with input from Year 2 teachers prior to the piloting and the number of items was reduced to ensure they could be administered within 25 minutes. The assessment used in the pilot consists of two practice items, followed by 12 test items with a total of 18 marks available—9 marks for geometry and 9 marks for problem-solving. The problem-solving and geometry items were presented in a fixed random order. The maths assessment was administered in two versions, Form A and Form B, each presenting the same items but in a different order. The order of items was counterbalanced in Form A and Form B to account for ordering effects potential fatigue when answering the latter items due to the length of the assessment. This counterbalancing was considered appropriate by the delivery team who designed the assessment based on their expertise in this area. Roughly half of the schools (8 of 15) received Form A at baseline, and the other half Form B, and each school received the alternate version at endline.
- **Spatial language**. This assessment covered language production and language comprehension and incorporated items from previous studies (Gilligan-Lee *et al.*, 2021; Gilligan-Lee *et al.*, 2019; Farran and Atkinson, 2016), which were adapted for this age group and whole-class administration. The language production section was administered first, to ensure that children were producing the answers spontaneously. The language production items consisted of a total of two practice items and six test items (one per page) and were presented first. On each page, students were asked to fill in a gap in a sentence so that it matched what they saw in a picture (see Figure 2). These items were followed by two practice items and 17 test items on language comprehension. In this case, students were presented with four pictures, and needed to circle the one they thought matched a sentence read out loud by their teacher (see Figure 3).

Figure 2: Example of language production item

'The ball is_____ the table'

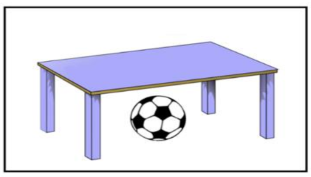
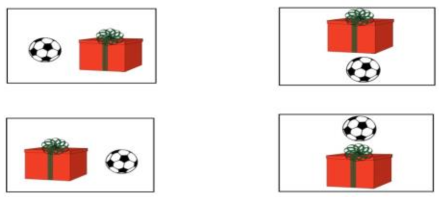


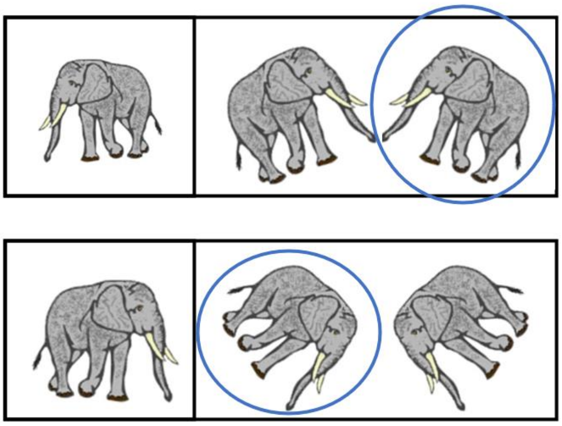
Figure 3: Example of language comprehension item

'The ball is above the present'



- **Mental rotation.** To try and prevent children from copying each other, the mental rotation task was designed with two versions to be distributed within classrooms in an alternated fashion. For each version, the assessment consisted of two practice items and 20 test items with four trials per page, asking students to identify, which out of the animals in the right column matched a target animal at the left of a column (see Figure 4). Items included rotations of an animal model at 45°, 90°, 135°, 150°, and 180°. The difference between the two versions is the way in which the animal needs to be rotated (clockwise or anticlockwise), while the degrees of rotation remain the same (e.g. Form A, the first item is a horse at 150°, clockwise; Form B, the first item is a horse at 150°, anticlockwise).

Figure 4: Examples of mental rotation items



Both the spatial language and mental rotation assessments were based on assessments previously used by the delivery team and administered by them with individual children in Year 3 (Gilligan-Lee *et al.,* 2021; Gilligan-Lee *et al.,* 2019; Farran and Atkinson, 2016). They were adapted to be used with Year 2 pupils and administrated by the class teacher to the whole class.

All the child assessments were paper-based and completed by each child individually during class time and administered by the class teacher. School staff were trained on the assessment measures during the half-day of SPACE training and were supported to administer them by the SPACE delivery team. They were also provided with a comprehensive manual, including step-by-step instructions[8] and a script, and a video. All children involved in the evaluation were asked to complete the assessment measures, with all children present on the day of completion expected to complete it. The evaluation team did not collect data on children whose parents formally withdrew them from the study, so withdrawal numbers are not presented here. Schools were asked to complete the assessments with children who were absent in a separate session, if feasible within their available staffing. At baseline, there was available data for 400–408 out of an anticipated total of 415 pupils (96%–98% complete). At endline, there was available data for 397–402 pupils (96%–97% complete). The variance is because not all children completed all three assessments. Of the pupils with parental consent, 7–18 pupils were absent or refused to complete the assessments at each timepoint, a total of 2%–4% of the possible evaluation sample.

In combination, the three measures were designed to take a maximum of 45 minutes to administer so as not to overburden schools (25 minutes for maths, 15 minutes for spatial language, and 5 minutes for mental rotation). Schools were asked to administer them either in a single session or across two sessions within the same week at the teacher's discretion. The measures were also intended to be administered with maths first, but the order for the spatial language and mental rotation measures was not fixed.

One member of the evaluation team (PV) scored the paper assessments (against a set of answers provided by the delivery team), and a second team member (AH) scored a random subsample of 3% per timepoint to check accuracy with 99.6% agreement. Scores were entered by an external company on an Excel spreadsheet with one row per student and one column per item. Data was double entered by the data entry company for 10% of the total sample to ensure consistency with 100% agreement.

Changes from pre- to post-intervention were analysed using an analysis of covariance (ANCOVA) to account for any differences by school, in which the dependent variable was the post-intervention score, the pre-intervention score was the covariate, and the school was entered as an independent variable. The range and distribution of results are reported here, including mean and standard deviation (SD), and floor and ceiling effects per assessment (maths, mental rotation, and spatial language) at pre- and post-assessment.

Administrative programme data

[8] The step-by-step instructions for each of the child assessments are available at the following link: https://www.surrey.ac.uk/spatial-cognition-enhance-mathematical-learning-space/space-programme-resources.

Administrative data monitoring programme delivery were collected by the delivery team and shared with the evaluation team. For each school, the information gathered consisted of:

- attendance at training sessions (including date of attendance, role of school representatives in attendance, and reasons for absence);
- participation in weekly support sessions (including date, attendees, format—face-to-face or online, notes on SPACE delivery);
- report on baseline assessment delivery (including number of assessment sessions and duration of each assessment); and
- SPACE lesson delivery (in support sessions, the delivery team verbally completed a set of pre-determined monitoring questions with the class teacher and recorded their responses on paper, i.e. dates of session delivery, duration of the sessions, presence of teacher and teaching assistant, number of children in the sessions, and number of models completed per child per session [digitised from the model booklets by the delivery team]).

These data were analysed using descriptive statistics and as part of the fidelity composite scoring described under the ‘Data synthesis’ subsection below.

Data regarding SPACE lesson delivery and assessment delivery was missing in several cases. This may suggest that the monitoring data collection processes were burdensome and could be improved for an efficacy trial—this is explored in greater detail in the ‘Readiness for an efficacy trial’ subsection under the ‘Findings’ section below.

Table 8 summarises how much of this information was gathered per school for analysis.

Table 8: Final achieved sample of monitoring data collected per school

School code	*Attendance to training*	*Participation in support sessions*	*Baseline assessment delivery*	*SPACE lesson delivery*
A	✓	✓	Missing all task duration data	✓
B	✓	✓	✓	✓
C	✓	✓	✓	✓
D	✓	✓	✓	✓
E	✓	✓	Missing mental rotation task duration data	Session duration missing for eight lessons
F	✓	✓	Missing all task duration data	Session duration missing for two lessons
G	✓	✓	Missing all task duration data	Session duration missing for two lessons
H	✓	✓	✓	✓
I	✓	✓	✓	Session duration missing for two lessons
J	✓	✓	✓	✓
K	✓	✓	✓	Session duration missing for six lessons
L	✓	✓	✓	Session duration missing for four lessons
M	✓	✓	✓	Session duration missing for ten lessons
N	✓	✓	✓	✓
O	✓	✓	✓	Session duration missing for two lessons

Data synthesis

Data from each data collection method were analysed as outlined above, then triangulated and integrated, identifying areas of difference and reinforcement and using different data sources to support and explain findings. Table 9 shows, which sources of information were used to answer each research question (see 'Research questions' subsection under the 'Introduction' section above).

Table 9: Evaluation data matrix

Data collection method		Research question no.									
		1	2	3	4	5	6	7	8	9	10
Monitoring data		x		x							
Teaching staff surveys		x	x		x	x	x	x			
Observations	Training sessions		x	x							
	Support sessions		x	x							
	SPACE sessions		x	x							
	Assessment sessions										x
Interviews	Teaching staff	x	x		x	x		x	x		
	SLT	x	x	x	x		x				
	Delivery team			x						x	x
Baseline–endline child outcome measures											x

Assessing fidelity

The evaluation team used a combination of data to explore whether SPACE was delivered as intended at each school site. Fidelity criteria were developed by the evaluation team based on the delivery team's description of delivery as intended. The criteria were established prior to delivery to ensure a non-biased rating of fidelity with ambitious but realistic expectations.

The fidelity was assessed using five criteria across three key dimensions: attendance; dosage; and quality of delivery. The criteria were assessed using information from the administrative data, interviews, and teaching staff surveys. The dimensions for measuring fidelity were:

- **Attendance was measured across two criteria teaching staff attend.** Training (criteria 1); and implementation support sessions (criteria 2).
- **Dosage.** Delivery of 12 sessions delivered twice weekly for six weeks lasting 30 minutes each (with flexibility to allow the 12 sessions to be delivered over eight weeks with a maximum of three sessions per week) (criteria 3).
- **Adherence to SPACE strategies/delivery quality was measured by combining two criteria.** Teaching staff use of spatial prompts (criteria 4); and children working autonomously on LEGO® models (criteria 5).

Attendance at training (criteria 1) was rated as follows:

- **Low.** The training was not attended by any staff.
- **Medium.** The training was attended by the delivering class teacher.
- **High.** The training was attended by more than one staff member including the delivering class teacher.

Attendance at support sessions (criteria 2) was rated as follows:

- **Low.** School staff attended one or fewer in-person implementation support sessions.
- **Medium.** School staff attended two in-person implementation support sessions.
- **High.** School staff attended all three in-person implementation support sessions.

Dosage (criteria 3) was assessed to establish whether children had received the intended dose of the intervention. This was assessed using the total number of sessions delivered per school. Dosage was rated as follows:

- **Low.** The school delivered <80% of the intended dosage.
- **Medium.** The school delivered 80%–90% of the intended dosage.
- **High.** The school delivered 90%–100% of the intended dosage.

The quality of delivery dimension was assessed based on interviews with school staff and observations of delivery and support sessions. These delivery features are indicative of the quality with which the teaching staff were delivering the intervention. We rated each school based on evidence for the use of spatial prompts to support children's problem-solving (criteria 4), and whether children worked autonomously (criteria 5). As non-mandatory delivery elements, these features were scored separately then half-weighted for the overall rating. They were scored as follows:

- **Low.** There is no evidence that the delivery staff adhered to the SPACE strategies.
- **Medium.** There is some evidence that the delivery staff adhered to the SPACE strategies.
- **High.** There is strong evidence that the delivery staff adhered to the SPACE strategies.

Composite ratings were assigned as follows: schools rated as high for three or more of the four criteria were rated as high overall; schools rated as high for two criteria were rated as medium overall; schools rated as high for fewer than two criteria were rated as low overall.

Timeline

Table 10 provides an outline of the activities related to the delivery of the intervention and the evaluation.

Table 10: Project timeline

Date	Activity
November 2021 – February 2022	Project set-up including IDEA workshop, ethical approval, evaluation plan, data sharing agreements
March 2023 – June 2023	Recruitment period
June 2023 – September 2023	School onboarding and evaluation material development
September 2023	Training of teaching staff
September 2023	First teaching staff survey
September 2023	Pre-intervention child assessments
September 2023 – November 2023	Start of SPACE intervention delivery
October 2023 – November 2023	Observations of SPACE lesson delivery and support sessions
November 2023	End of SPACE intervention delivery
November 2023 – January 2023	Interviews with school staff
November 2023 – December 2023	Post-intervention child assessments and observations
December 2023	Second teaching staff survey
March 2024 – April 2024	Third school staff survey
January 2024 – June 2024	Evaluation data analysis
June 2024	Presentation of preliminary results to the EEF
June 2024 – August 2024	Reporting

Findings

Participants

A total of 15 schools participated in the SPACE pilot evaluation between September 2023 and April 2024. Table 11 details the characteristics of these schools. A total of 415 children took part in the evaluation of SPACE, 46% of whom were female. The average number of children participating in the evaluation per class was 27.7. All children were in a Year 2 class and the youngest child was 6.0 years old, and the oldest child was 7.7 years old, with an average age of 6.5 years.

One classroom teacher per school was involved in the delivery of the programme, as reported in Table 4 above (see subsection 'Recruitment and consent' in the 'Methods' section).

Table 11: Summary of characteristics of participating schools in the pilot evaluation

Local authority	N
Portsmouth	8
Surrey	7
Total	15
School type	N
LEA-maintained	5
Academy	9
Other school type	1
Total	15
School size (n of pupils)	N
Large (over 400)	5
Medium (200–399)	4
Small (less than 199)	6
Total	15
% of pupils eligible for FSM	N
High (over 25%)	5
Medium (between 18% and 25%)	5
Low (less than 18%)	5
Total	15
Ofsted rating	N
Outstanding	2
Good	11
Requires improvement	2
Total	15
Number of Year 2 classes	N
1	3
2	12
Total	15

Feasibility of implementation

The feasibility of implementing an intervention depends on various factors at each layer of delivery, the recipients of the programme, the delivery staff, the delivery context, and the wider policy context. In this section, the necessary conditions and facilitators at the individual, class, and school level to deliver SPACE are explored.

The research questions have been reordered in this section. We start by presenting information about whether the programme was delivered as intended and the sufficiency of the training and delivery materials, then provide information on acceptability and engagement and end with a section on the feasibility of delivering SPACE in maths lessons. The overall feasibility of the programme is informed by various elements across these questions.

Research question 3: Was SPACE implemented as intended?

A key starting point in understanding feasibility is whether SPACE was delivered as intended, with fidelity. An evaluation of a programme that was not delivered as intended could lead to spurious answers to key questions regarding the feasibility and promise of a programme.

Composite fidelity score

Delivery in each school was assessed across the three fidelity domains using a composite fidelity score, as described in the Assessing fidelity subsection under the 'Methods' section above: attendance, dosage, and quality of delivery.

A composite fidelity rating was assigned to each school composed of five criteria across the three domains. The quality criteria were half-weighted as they consisted of non-mandatory features of delivery. Schools rated as high for three or more of the criteria were rated as high overall; schools rated as high for two criteria were rated as medium overall; schools rated as high for fewer than two criteria were rated as low overall.

The fidelity scores by school are shown in Table 12. Overall, the data show that the SPACE intervention was delivered as intended, with a medium or high rating for every school in every one of the fidelity criteria, apart from where data were not available. In the following sections, we present some further insights into the variations in delivery within the three dimensions.

A total of 12 schools delivered SPACE with high overall fidelity, and two delivered it with medium overall fidelity. One school could not be rated due to missing data for the quality criteria, as no observations or interviews were conducted with that school but it was rated high across all other criteria.

Table 12: Fidelity scores per school per measure

School ID	Attendance training	Attendance support sessions	Dosage	Quality of delivery: use of spatial prompts	Quality of delivery: children working independently	Overall
A	High	High	High	Med	High	High
B	High	Med	High	High	Med	Med
C	High	High	High	Med	High	High
D	High	High	High	High	High	High
E	High	High	High	Med	High	High
F	High	High	High	Med	High	High
G	High	High	High	No data	No data	Insufficient data
H	High	High	High	High	High	High
I	High	Med	High	High	High	High
J	Med	High	High	High	High	High
K	High	High	High	High	High	High
L	High	High	High	High	High	High
M	Med	High	High	Med	Med	Med
N	High	High	High	High	Med	High
O	High	High	High	High	High	High
Total no. high (of 15)	13	13	15	9	11	12

Attendance

Attendance to both training and the in-person support sessions are key elements of delivery to ensure that the principles behind the programme were transferred as designed to the practitioners leading the SPACE sessions, to enable them to deliver the sessions as intended.

Overall, schools showed high fidelity, with 13 schools out of 15 rated high for attendance at training and 13 rated high for attendance at the support sessions.

Most schools sent two members of staff to the half-day training session as requested, though two schools only sent the relevant class teacher. Three schools did not send the relevant teaching assistant but one of them sent a member of the SLT to make up the two staff overall. In the interviews, some schools shared that it was challenging to release two members of staff, especially in the first week of term after the summer holidays. This data does not include the school that withdrew prior to delivery commencing.

Only the number of in-person support sessions attended was monitored, as the online support sessions were optional. Two teachers missed one support session each. However, these sessions were missed due to travel issues or illness and teachers were willing and able to replace the sessions with online sessions to cover the relevant content and were not perceived to diminish the quality of delivery, demonstrating that it is feasible to substitute in-person sessions with online sessions when required. Some online support sessions were missed due to teachers not having time due to children's behaviour. Interviewees reflected that the online support sessions were not necessary, the required information could be gathered by email, and the fortnightly in-person sessions were sufficient.

We can, therefore, conclude that teaching staff felt they received the training and support required to deliver the programme as intended.

Dosage

The delivery of the intended number of sessions was core to the programme having the anticipated outcomes for teaching staff and children, by allowing for sufficient exposure, repetition, and reinforcement.

All schools delivered the required number of sessions, with some agreed deviation in the elapsed period for delivery and acceptable variation in the session length.

The intervention was designed to be delivered as two sessions per week for six weeks. However, some flexibility was built in to allow for school holidays and unanticipated circumstances—full dosage was considered achieved if a school delivered the 12 sessions within eight weeks, with a maximum of three sessions in one week. The sessions were delivered across an average of 7.3 weeks, which included the one-week Autumn Half-Term break and, in some schools, formal assessment weeks. One school's delivery was delayed due to a teacher's illness. No schools delivered more than two sessions in a week; the maximum number of sessions delivered in one week was two, and the minimum was zero.

Sessions were designed to take 30 minutes in total—20 minutes of delivery plus ten minutes of set-up and break down time. The average duration of the sessions across schools was 30 minutes, as intended. Though there was some variation in session length across schools, nine of the 15 schools' average session length was 28–32 minutes. Of the remaining five schools, three delivered the sessions in a shorter period (on average between 20 and 25 minutes per session), and three spent longer (on average between 33 and 40 minutes per session). Session length decreased across the duration of the intervention, with the first one or two sessions tending to be the longest before stabilising, as is reflected in the following quote:

> *Occasionally at the beginning, it would've taken us a bit longer in terms of the setting up and setting down, but half an hour's fine*. (Teacher)

Schools were asked to deliver SPACE sessions on different days during their scheduled maths time. One school initially held the first two SPACE sessions outside their timetabled maths lessons. However, this was identified in a support session and corrected from the second week of delivery onwards. A different school delivered two SPACE sessions back-to-back in each of the first two weeks of the programme before being reminded by the delivery team to deliver the sessions on separate days, as this intended intensity was designed to encourage repetition and reinforcement.

Therefore, although some variation in programme duration and session length was observed, it was within the parameters of the planned programme. Minor deviations were addressed early through the delivery team's established support mechanism and should not affect the programme's outcomes.

Quality of delivery

The core activity of the programme was children independently working on the models, by using visualisation and problem-solving. The quality of delivery was also anticipated to be enhanced through teaching staff's use of spatial prompts.

The quality of delivery was the criterion delivered with the lowest fidelity—nine schools were rated high for the use of spatial prompts and 11 schools were rated high for children working independently. The lower (but still good) fidelity ratings for quality compared to the other fidelity criteria can be explained by the fact that these two criteria on which quality of delivery was assessed were not mandatory elements of the programme, though they were strongly advised. Additionally, these criteria were rated largely based on qualitative interview data, using observations in a subsample of schools and semi-structured interviews. This means these criteria were not systematically assessed across schools, and the ratings may underestimate the prevalence of these delivery elements in the classroom.

Teaching staff were asked to use spatial prompts in SPACE sessions and were given bespoke prompt cards with suggestions for spatial prompting (including spatial language vocabulary) for this purpose. Teaching staff widely reported using spatial prompts to support students. They reported modelling spatial language by 'shoehorning' language from the prompt cards into their support at 'any and all opportunities'. Although they tended not to directly use the prompt cards during sessions, teaching staff reported that they reviewed them prior to delivery and found the ideas for vocabulary to use and questions to ask helpful.

Though a second adult was not required for delivery, teachers also commented that they found it much easier to use spatial prompts to support individual children when there was another adult in the classroom. A teaching assistant also commented that even with two teaching staff in the room, they could not effectively support all 30 pupils and had to prioritise whom they supported based on their knowledge of ability.

Teaching staff were asked to encourage children to work independently rather than in groups and not to provide manual help with the models. Overall, children were observed to work independently and any support from teachers or their peers was verbal rather than manual—teaching staff rarely handled the LEGO® apart from taking models apart. The teaching staff used various techniques to facilitate children working independently. To check that their models were correct, children were asked to show them to a teacher, or a peer, or check the diagram themselves. Teachers found that checking all models took away time from supporting children with spatial prompts, so encouraged children to check each other's models after the first few sessions. Pairing higher and lower-ability children together was a strategy used to encourage this and to keep the focus on accuracy rather than on speed.

These insights indicate that teaching staff were willing and able to incorporate delivery strategies to facilitate high-quality implementation, but that it is delivered more consistently where there are more members of staff available in the classroom. Future iterations of SPACE should consider staffing requirements to facilitate quality of delivery.

Research question 2: Are the training and support provided sufficient for teachers and teaching assistants to implement SPACE?

Schools' perceptions of the half-day training session and resources provided for the SPACE programme were generally positive. Teaching staff reported that they had sufficient training and support to deliver the programme; however, some room for improvement was reported for the training, videos, prompt cards, and booklets that would facilitate higher quality delivery and better integration of the content with their usual classroom teaching.

The SPACE delivery resources were:

- a comprehensive delivery manual for teaching staff;
- prompt cards for teaching staff, which provided suggested vocabulary and questions to use in the SPACE sessions and more widely in classroom teaching;

- short online training videos for teaching staff to refer to for 'top up' training tips and guidance on delivery;
- LEGO® boxes (one per child plus spares) with all the materials required to build the models;
- model booklets, which contain the pictorial exploded models for children to build and indicate, which models they completed; and
- weekly themed two-minute videos to introduce the story of each week's theme to children.

Teaching staff found that the half-day training, manual, and other materials provided them with the information they needed to deliver the intervention confidently and that the physical resources were well-prepared.

In the Timepoint 2 survey, teachers, teaching assistants, and SLT representatives were asked how satisfied they were with the half-day training session. All participants (14/14) indicated that they were quite (5) or very (9) satisfied with the training. Furthermore, most participants indicated that they felt ready to implement the programme in their classroom after the training. Of the 16 respondents, six indicated that they felt very ready, nine that they were somewhat ready, and one that they were only a little ready.

In interviews, teaching staff commented that they enjoyed the training and valued learning about the scientific background of the programme and the 'why' of what they were being asked to do, as well as detailed information about how to run the sessions. Training attendees particularly enjoyed the modelling of delivery by the delivery team and the chance to try out the LEGO® boxes themselves. Teaching staff felt that the duration and level of the training were appropriate, though some felt that they would not know how it would work in practice until they implemented it with their class.

> *We [teaching staff] got to know the subject matter but also like the nitty gritty of actually, what am I supposed to do in front of the class, so that was really good.* (Teacher)

Not all reflections on the training were positive. Some interviewees found that the scheduling of the training at the start of the academic year made it difficult to attend and took too much time out of the classroom at an important stage of settling into their new class. There were also indications that teaching staff needed more advance notice and information about how the content of SPACE links with the wider maths curriculum (see the feasibility within maths lessons subsection below) to enable them to embed it in their planned maths teaching. At Timepoint 2, survey participants were also asked how satisfied they were with the weekly support sessions provided by the delivery team. All participants (12/12) indicated that they were 'quite satisfied' (6) or 'very satisfied' (6) with the support sessions. In the interviews, teachers reported that they felt very supported by the delivery team and appreciated how available and responsive they were, both in the frequent support sessions and by email. However, there was a consensus that the in-person meetings did not need to be so regular and that most of the monitoring information collected in the support sessions each week by the delivery team could have been sent by email.

> *If we continued it [the programme] for longer I'd maybe say [support sessions] every other week or at the end of every block. Because she's [delivery team research assistant] so good at replying to her emails, if I do ever have a question, she'll reply to it, or I can call her if I need to.* (Teacher)

The resources (teacher training videos, training manual, prompt cards, weekly theme videos, model booklets, and LEGO® boxes) were well received by survey participants (Timepoint 2) and interviewees (Table 13), though some issues were identified, which could facilitate more consistently high-quality delivery. Storage of the large LEGO boxes was challenging in some schools due to limited space. Teachers from most schools reported that the weekly theme videos were initially useful but quickly became repetitive, and children became disengaged with them. While the language and ideas included on the prompt cards were generally considered useful, both for supporting children during the sessions and for extending the spatial thinking techniques to other teaching, teachers found the format cumbersome for use during the SPACE sessions. This limited how frequently they were used—teaching staff tended to quickly review them before delivering a session instead of during the sessions. In interviews and support sessions, teaching staff made suggestions for making the content more accessible, for example, flash cards to hang on their lanyard or a poster on the wall.

Other specific suggestions to improve the format and usability of the physical resources, such as making the two-minute videos more varied and engaging, with explicit reference to the session's theme and links to the curriculum and ensuring that the booklet models had good contrast between LEGO® pieces, were proposed by several teachers and teaching assistants.

Overall, the training and support provided were sufficient for delivery, but further improvements to the timing and content could facilitate better embedding into maths teaching and more consistent high-quality delivery across all schools.

Table 13: Satisfaction with the intervention resources from Timepoint 2 survey (N=14)

Intervention resource	Very dissatisfied	Quite dissatisfied	Neither satisfied not dissatisfied	Quite satisfied	Very satisfied
Training videos	0	0	0	4	10
Training manual	0	0	1	5	8
Prompt cards	0	0	6	5	3
Weekly short videos	0	4	4	4	3
Model booklets	0	2	1	6	5
Assessment booklets	1	0	6	4	3

Research question 4: Is SPACE acceptable to teaching staff?

The acceptability of a programme refers to the extent to which the teaching staff perceive the programme to be agreeable and includes satisfaction with the programme's content, complexity, comfort, delivery, and credibility (Proctor *et al.,* 2011). A high degree of acceptability is considered important to successful implementation, as those delivering the programme are less likely to deliver it with fidelity or sustain it if they do not like it.

Overall, the teaching staff interviewed found the SPACE programme acceptable in terms of its content and complexity. However, as discussed in research question 1, there were indications that more direct signposting regarding how the content relates to their maths curriculum would increase the comfort of delivery. Relating to credibility, some teachers also reported in interviews that they wanted more information regarding the efficacy of the programme before making it standard practice in their classroom.

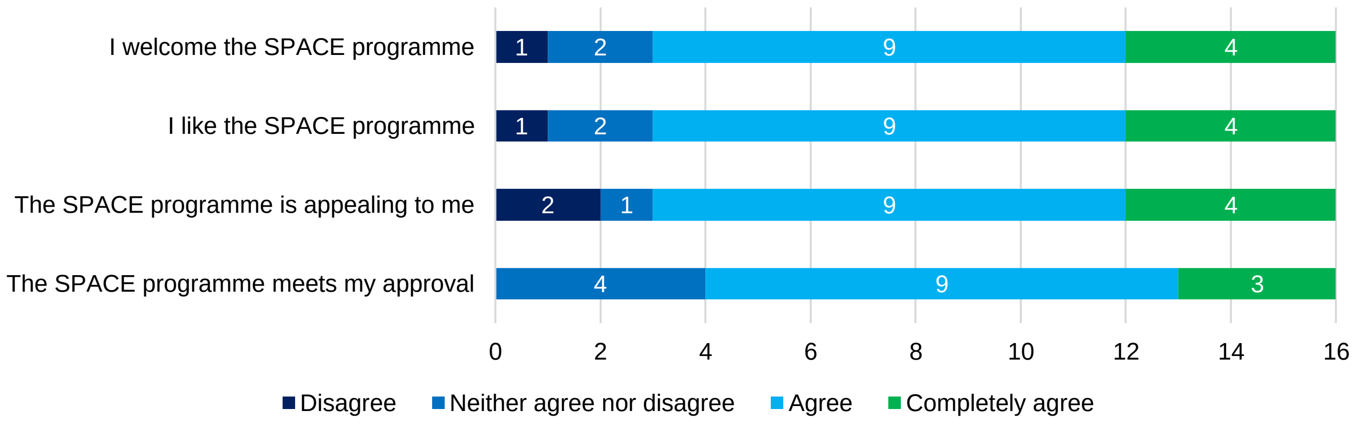
> *If it was ever something that was intended to be used as a whole scheme, I'd need to see more evidence of how that's going to get them to be able to do their timetables, do adding, subtracting, fractions, that kind of thing.* (Teacher)

The Acceptability of Implementation Measure (Weiner *et al.,* 2017) was administered to teaching staff in the Timepoint 2 staff survey. Participants were asked to indicate whether or not they agree with four statements:

- The SPACE programme meets my approval.
- The SPACE programme is appealing to me.
- I like the SPACE programme.
- I welcome the SPACE programme.

Figure 5 presents the findings from this scale. Overall, 80% of responses to these items were 'agree' or 'completely agree', with an average acceptability score of 4.0/5, demonstrating a high degree of acceptability.

Figure 5: Acceptability of the SPACE programme (N=16)



Interviews and observations also suggest that the programme was considered highly acceptable by teaching staff as well as SLT representatives. Teachers and teaching assistants described how much they enjoyed delivering the sessions and found it an interesting and positive experience.

> *I look forward to delivering the sessions.* (Teacher)

We also heard from a teaching assistant that they had anticipated it being 'chaotic' and difficult to deliver due to the large amounts of equipment involved and were surprised at how smoothly it had gone.

> *I honestly thought it would be a disastrous car crash, and what are we doing, if I'm honest? It was a lot more enjoyable than I thought it was going to be.* (Teaching assistant)

Teaching staff commented that they like the content and purpose of the programme and could see its value.

> *It's a very clever idea and I can see why and it's really important I think, it's really good […] I can see it's a lovely programme.* (Teaching assistant)

Teaching staff identified some unexpected positive aspects of the programme. Teachers valued having two sessions per week, which required no lesson planning or marking, as it gave them more time to focus on planning other lessons. They also appreciated the reminder that children in Year 2 still take a lot of value from learning through play. Teachers and teaching assistants also liked seeing children work independently and with focus and mentioned that it helped with the transition from continuous provision in Year 1 to structured learning in Year 2, which is sometimes difficult to achieve with the standard curriculum early in the academic year.

> *They're very used to getting up and doing things, and the time of year that the programme was happening was sort of a nice transition year from that doing to the sitting down, having a real proper focused lesson. It was quite a nice introduction to Year 2 that way.* (Teacher)

Through the interviews and qualitative survey responses at Timepoint 2 and Timepoint 3, we heard lots of interest from schools in continuing SPACE, in some capacity, which suggests that teachers liked and recognised the value of SPACE. This included rerunning the full programme, continuing to use strategies and language from the prompt cards, using more manipulatives in the classroom, and interest in other LEGO®-based interventions. Many interviewees shared that they had already changed how they approach modelling actions and instructions for children, shifting to using spatial language rather than manually manipulating or guiding.

> *We really enjoyed it, and it was something that actually we were like, if it was rolled out to be like a programme that you could buy into, we would probably buy into it and do it at the start of every year to increase their spatial awareness.* (Teacher)

Interviewees felt that if the content provided more direct links to the curriculum, for example, through weekly learning objectives or vocabulary, this would make them more comfortable with delivering it again, even if it meant postponing some other curriculum content. This theme is discussed further in the subsection 'Overall feasibility' (under research question 1) below.

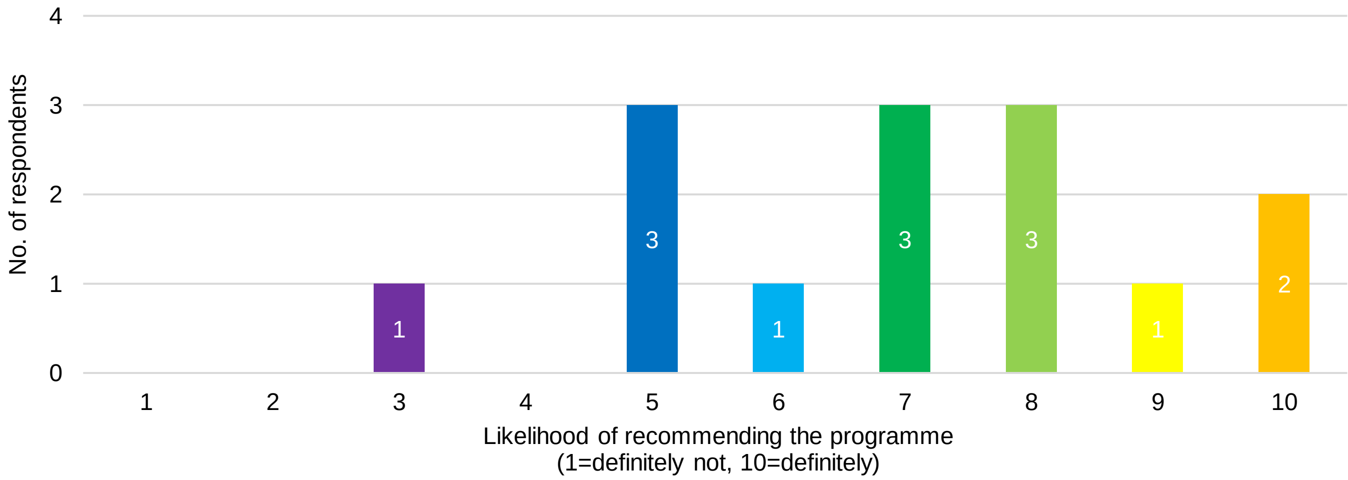
> *I see how it's beneficial for them and that they really enjoy it, but in terms of things like addition and subtraction, in terms of doing the actual sums and seeing the actual numbers, that's the only bit where I'm like, I don't see a correlation.* (Teacher)

Another indication of the acceptability of a programme is whether delivery staff are likely to recommend it to other settings. At Timepoint 2, survey participants were asked how likely they were (on a scale of 1, 'definitely not' to 10, 'definitely') to recommend SPACE to other schools.

The results in Figure 6 show that while most participants (64%) responded with a rating of 7 or above, a notable proportion (28%) are less likely to recommend it—they indicated a rating of 6 or less on the scale. In line with this survey finding, some interviewees were very enthusiastic and said they had already recommended the programme to other schools, though others were concerned that the content was not directly linked to their maths teaching and that they could not see it benefiting the more concrete maths skills.

> *I would definitely recommend it, and then, in terms of other schools, I'd say just to be aware that if you're doing it around your other curriculum, you might be behind.* (SLT representative)

Figure 6: Likelihood to recommend the SPACE intervention to other schools

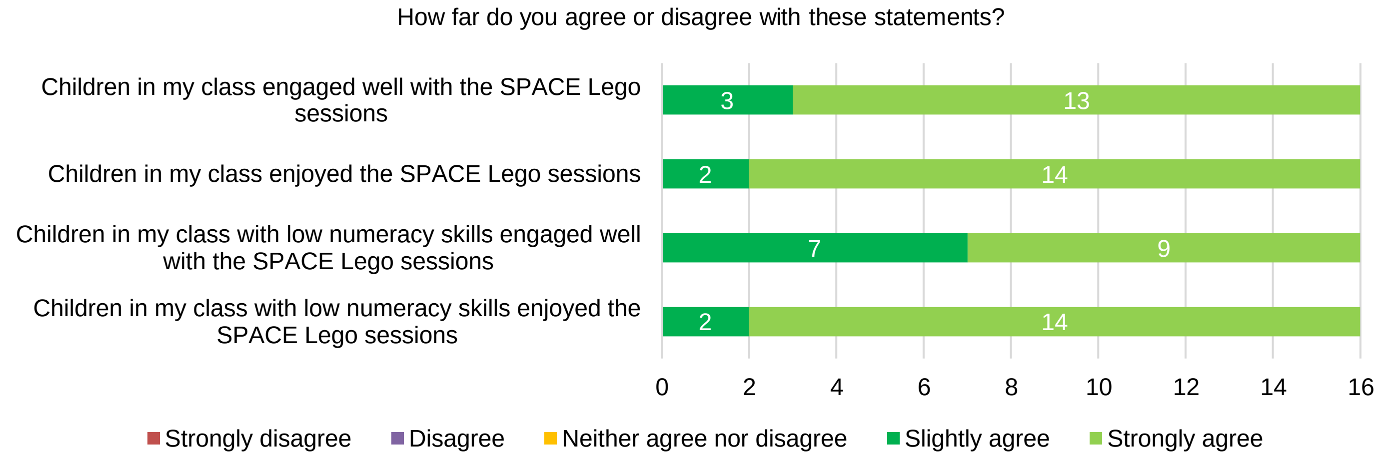


Research question 5: Do children, particularly those from disadvantaged backgrounds, engage well with SPACE?

Based on the survey, interviews, and observations, the overwhelming perception of children's engagement was that they showed a high level of engagement and enjoyment during the SPACE sessions. This finding was mostly consistent across children's demographics and characteristics, though there was some perceived variation, and some individual pupils struggled to engage.

Survey results at post-delivery (Figure 7) show that 100% of participating teaching staff agree or strongly agree that children, including those with low numeracy skills, engaged well with and enjoyed the SPACE sessions. There was no difference between perceived enjoyment of children overall and of children with low numeracy skills. However, there was a difference in the distribution of results for their engagement—participants were less likely to strongly agree that children with low numeracy skills engaged well with sessions.

Figure 7: Children's engagement and enjoyment of the SPACE programme



In interviews, teaching staff widely used words like ‘loved’, ‘excitement’, and ‘fun’ to describe how the majority of children engaged with the programme and reported that children in their class looked forward to the sessions, particularly in the first couple of weeks.

> *There’s only those small few that just have their struggles in general, that possibly it was a bit more of a challenge. I would say 98 per cent of the class really enjoyed it, got really excited when they saw the boxes coming in, and they knew we were doing it, would ask: ‘When are we doing the next session?’, and enjoyed being a little bit competitive with some of their friends.* (Teacher)

> *There wasn’t a time, I don’t think, when the engagement was different depending on ability or knowledge of language, anything like that.* (Teacher)

Children really enjoyed the SPACE sessions because they were so different from their usual classroom activities, and particularly because of the association between LEGO® equipment and play, as perceived by teaching staff and in the delivery observations. This enjoyment translated into high engagement with the activity.

> *Oh, they loved it. They absolutely loved it. I think it’s motivating, being LEGO®. I think if you ask them to do anything with LEGO they will do it just because it is exciting and it’s not something that we often do. So therefore, it feels like a treat. Their engagement levels are really high.* (Teacher)

The high levels of engagement translated into focused, calm sessions, in which children worked independently, again demonstrating high levels of engagement. Some teachers were surprised by this, as they perceived the LEGO® as a toy and expected children’s excitement to lead to disorganised and chaotic sessions.

> *It was actually usually a very calm session, that they were all really focused and really enjoying themselves doing it. So, I definitely think they definitely loved it and really enjoyed it.* (Teaching assistant)

Although not universal across classes, there was some indication that children with different characteristics engaged differently with the programme, particularly at the start. Girls, children who did not have LEGO® at home, children with lower abilities, and children with special educational needs and disabilities (SEND) were perceived by some teaching staff as not being as excited about the sessions as their peers, as reported in interviews and support sessions. However, these differences appear to have levelled out across the duration of the programme and were not perceived universally. Teachers attributed this difference to a lack of familiarity or confidence with the manipulatives for children who do not have access to them at home—who they saw as largely children from lower socio-economic households.

> *Most children as well that have LEGO® at home just took to it more. They’ve got that confidence, just pick up a LEGO set and be able to just dive into it because they’re familiar with it. Whereas those children probably who don’t have the opportunity to play with LEGO at home were just a little bit more reluctant.* (Teacher)

Teaching staff described differences in engagement for some children depending on their family's socio-economic background. However, we did not note any differences between schools with different percentages of children receiving FSM in interviewees' descriptions of engagement across schools, nor between the two geographical locations, one of which is an education improvement area.

Some of the variation in how interested girls were in the sessions was thought to be due to the weekly topics.

> *I think a lot of them felt very stereotypically boy-friendly. I think near the beginning some of my girls weren't as interested because stereotypically, the girls aren't the scientists and the pirates. But I don't think it affected the way they engaged with it.* (Teacher)

This view was not shared by all interviewees.

> *I don't think there was any real difference with gender. Generally speaking, the boys enjoyed it as much as the girls.* (Teacher)

Some children with behavioural or emotional regulation issues found it challenging to engage.

> *I think we did have a few behavioural issues; when they couldn't do it, they gave up. I had a few tears. I had a few strops out of the room. We had a few giving up completely*. (Teaching assistant)

However, teaching staff also reported being surprised about individual children who either loved the programme or struggled to engage with it. They reported that some higher-ability children struggled with not having verbal or written instructions to follow and got frustrated when they were not immediately good at the task.

> *Often your higher achievers are very quick with numbers but may not have good problem-solving skills when it comes to spatial awareness.* (SLT representative)

Conversely, some children otherwise identified as having lower abilities found that they could quickly grasp the concept—which teaching staff attributed to them being more comfortable with visual instructions. This, in turn, increased their confidence and self-esteem, with some children being seen to offer support to their peers for the first time. Interviewees also commented that this confidence transferred to other areas of their learning and behaviour. Similarly, in the interviews and support sessions, teachers reported that children with English as an Additional Language (EAL) enjoyed the visual instructions and revealed an aptitude that they had not previously been able to express.

> *Some of the ones that you'd say might find maths a bit more challenging have really excelled in LEGO® maths. I was surprised by some of it, and also shocked by other children that I thought might do better with it but didn't have that kind of spatial understanding.* (Teacher)

Overall, the implication is that, although there were different individual experiences reported, there is no clear, consistent evidence that children from disadvantaged backgrounds engaged significantly less well or better with SPACE.

Research question 1: Is SPACE feasible for implementation in maths lessons?

Overall feasibility

Feasibility describes the ease with which an intervention can be implemented in a given setting, including its suitability and fit (Proctor *et al.,* 2011). The feasibility of an intervention depends on many features, including the resources required, structural or organisational support, and competing priorities. In this section, we discuss the overall feasibility of SPACE and consider the necessary conditions or facilitators at the school and teacher level for it to be successfully delivered in schools.

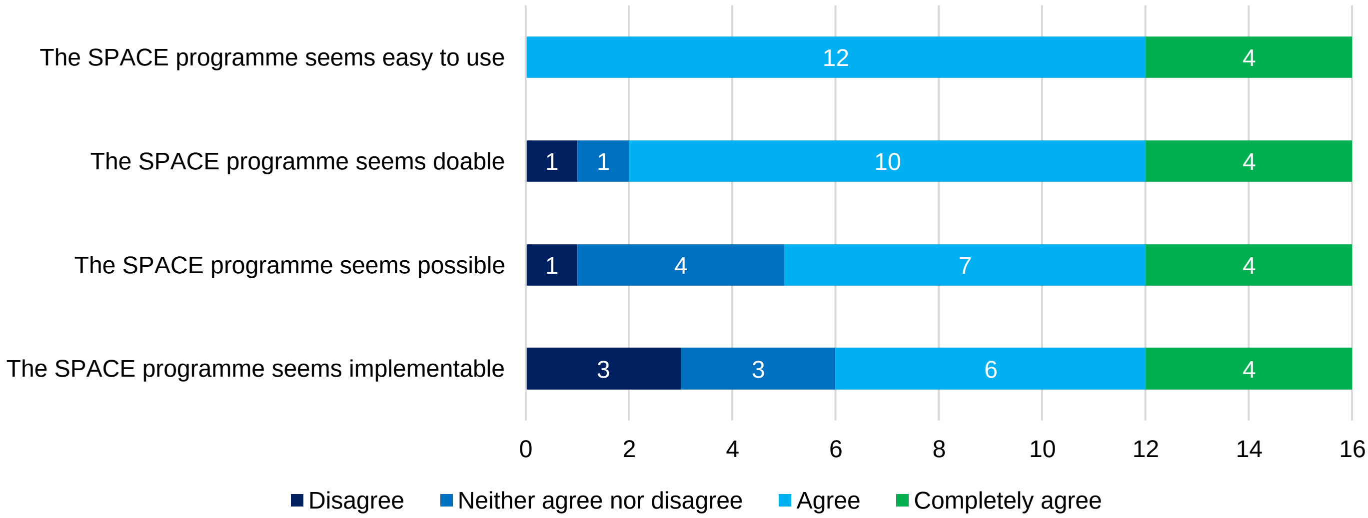
Implementation of the SPACE intervention was widely seen as feasible by teaching staff, with regard to the time and skills required to deliver the sessions, and the fidelity data shows that the programme was delivered as intended. However, several factors affecting the ease of implementation and significant issues regarding how the programme is embedded into maths lessons and the implications of changing how scheduled maths lesson time is used were identified.

The Weiner Feasibility of Implementation Measure was administered (Weiner *et al.*, 2017) to teaching staff in the Timepoint 2 staff survey to assess the programmes' ease, suitability, and fit. Participants were asked to indicate whether or not they agree with four statements:

- The SPACE programme seems implementable.
- The SPACE programme seems possible.
- The SPACE programme seems doable.
- The SPACE programme seems easy to use.

Results from the Feasibility of Implementation Measure indicate good levels of feasibility overall, with a feasibility score of 4.0/5 and 80% of responses agreeing or completely agreeing across the four statements. The statement with the highest rating (4.3/5) is 'easy to use', while the statement with the lowest rating (3.7/5) is 'implementable'. This difference likely reflects the dichotomy between the ease of implementing the sessions themselves and perceived challenges of implementing them within planned maths lesson time, which is discussed in detail in this section.

Figure 8: Feasibility of the SPACE programme (N=16)



Interviews and observations mirror these positive feasibility scores. Teaching staff and SLT representatives reflected that the programme was delivered easily without any issues or significant effort. Some felt that it had started well and become even easier and quicker across the term as the teaching staff and children got used to the routine of setting up and putting away the materials.

> *It went really smoothly. There have been no problems. It hasn't been a nightmare; it hasn't been difficult at all through the whole process.* (SLT representative)

In terms of delivery resources, interviewees noted that they were thorough and well organised, which facilitated delivery in the classroom. They reported that they had all the materials needed and it did not require significant set-up time, beyond initially organising the LEGO® boxes and booklets.

> *I think the organisation of the packs is what assisted with that […] the way it was organised, made it totally manageable and enjoyable.* (Teaching assistant)

> *We don't have to do much preparation for it. Just get the boxes out, give them their booklets, which is easy.* (Teaching assistant)

As described in the research question 2 section, the training and support were also seen as appropriate and contributed to the ease of delivery.

Interviewees described several factors that affected the ease of implementation of SPACE:

- **Space for storage.** Smaller schools found storing the materials and ensuring children did not handle them between sessions more difficult. One school reporting having to move their physical education equipment to a different part of the building to make space for it.
- **Classroom size.** The teaching staff felt it important to have enough space in the classroom for children to spread out and work independently. The LEGO® boxes require more table space than their standard classroom activity, so many utilised the carpet and other areas for the session. Some found it hard to keep children's boxes separate in sessions, while it was observed that several asked children to build inside the lid to save room.

  Obviously pieces would start going on the floor, or because there's like 30 in a small classroom, it's hard to try and keep the boxes separated from everyone else's. (Teacher)

  I had about half a dozen on the floor, as well as the tables, giving them a little bit more room…When you've got six at a table, it's a bit clustered, so actually, a little more space meant that they could concentrate better. (Teacher)
- **Presence of a teaching assistant or second adult in the room.** Teachers reported finding it more challenging to deliver the sessions without support, both in terms of set-up and supporting children during the delivery itself. They also felt that having more adult input improved outcomes for more children, as they could support them more explicitly with spatial language and problem-solving.

  Unfortunately, I was on my own on a couple of occasions. That was really tricky. (Teacher)

  Having an extra pair of hands is almost paramount. (Teacher)
- **SLT support**. Some teachers found that having interest from senior members of staff meant that they felt supported and able to invest time in the training and programme to deliver it as intended. One teacher shared:

  The headteacher is 100 per cent behind me. She promoted it. She was quite keen that it went ahead, and she saw the spark of, 'Ooh, LEGO®!' from me, and she said, 'You'll be the one then'. (Teacher)
- **Alignment with school strategy and culture.** Teaching staff in schools that identified a fit between the programme and their school's strategy to incorporate spatial learning and learning through play had a more positive view of the feasibility of delivery beyond the pilot evaluation.

  [SPACE] fits really well, and actually, we're trying to improve the importance of shape, space and measure and spatial reasoning within the school, so it's something that we know is a weakness. […] The idea of learning through play and all of that vocabulary and language fits in really well with our ethos and culture that we use here anyway. (SLT representative)
- **Time of the academic year**. Some teachers felt that the timing in the school year meant that they did not make the most of the programme. This was partly due to competing priorities in the first term of Year 2, when children were making the transition out of continuous provision, and partly due to the curricular coverage during that period, which had fewer obvious links with the content of SPACE sessions. However, others saw it as a useful activity to help children transition into more structured learning. Overall, implementation in the Spring Term might be more feasible.

  I think if we had done it alongside the time where we were doing position and direction, more things like that, I think it would have been more beneficial. I think where we were doing it alongside place value and addition, I don't think it really enhanced it, it was just the wrong time of year. (Teacher)

Adaptations to delivery

A few adaptations to the programme delivery were identified in the support sessions and interviews. In one school, teachers allowed free play once children had completed the models. The teacher reported that this approach incentivised children to rush through the models or to say they had completed them when they had not, so worked with the research assistant to prevent this. Another teacher asked children to redo their favourite model if they finished early. We also heard in a support session observation that a few teachers allowed pupils five minutes of free play before starting on the models to *'get it out of their system'* (Teacher), though this was not a common practice.

After the first few sessions, some teachers put the theme video on during set-up or the first two minutes of delivery, rather than requiring children to watch it as a class, because children found the repetitive content frustrating. This was considered acceptable by the delivery team.

One significant adaptation in some schools was taking a small group of lower-ability children out of the classroom to complete the programme with a teaching assistant. This was considered beneficial to children's engagement with the programme as it provided more space in the classroom while also allowing pupils with sensory issues or SEND to complete it in a less overstimulating space and with more targeted support. This adaptation was reported in three schools, though it is unclear how frequently this practice was used across schools, as it was identified qualitatively. There was some indication that children in the smaller groups did not work as independently as those in the whole-class setting.

Feasibility within maths lessons

The key challenge for schools was delivering SPACE within scheduled maths time. At the point of the half-day training, most schools had already timetabled their maths curriculum for the term and had not accounted for SPACE in their planning, meaning they had to drop or delay some intended maths content to make room to deliver the programme. This led some teaching staff to suggest that SPACE should be delivered as a small group intervention for children who needed more opportunities for working with manipulatives, or as a lunch club.

Teachers afforded more flexibility with regard to timetabling and lesson planning reported finding it easier to implement than teachers in schools with fixed schedules, such as multi-academy trusts, those following prescriptive curriculum like White Rose, or those with a parallel Year 2 class not participating in the programme. This particularly relates to the rate at which they were expected to cover their standard maths content.

> *It's the timetabling issues that were the hardest and making sure—because our curriculum is really tight, and quite formal in what has to be taught, there are no gaps in that. So, we had to be really creative with how we could fit this [SPACE] in.* (SLT representative)

Some teachers replaced whole maths lessons with SPACE. Others taught the planned maths content at the start of planned maths lessons but used time that would have been spent on exercises instead to deliver SPACE, meaning children did not get the reinforcement of applying the learning they usually would have had. Some found they were several weeks behind the planned content by Christmas and had to catch-up in additional maths sessions once SPACE delivery had ended. This was particularly a problem in schools with more rigid lesson plans, such as multi-academy trusts, and where another Year 2 class was not participating in SPACE and had moved faster through the planned teaching. Teachers commented that the maths content they were teaching 'did not stick' and children had often forgotten the material by the following lesson because they had not completed the usual related exercises.

> *Some of the things that they've learned haven't really stuck, because they've had like maths lesson, LEGO®, maths lesson, LEGO.* (Teacher)

Others felt that they could have fitted the programme into their teaching better with more time to plan.

> *The SPACE programme didn't fit with our unit solidly. It would have done. Probably we could've fitted it in somehow with time and planning.* (Teacher)

Some teaching staff interviewed were concerned that children's learning through SPACE was not directly linked to their concurrent maths teaching, such as addition and subtraction. Where teachers could not see a direct link between SPACE and the maths content they were teaching, or they could not see short-term positive outcomes, this timetabling issue was seen as more problematic. These teachers saw SPACE as 'replacing' maths or 'taking up maths time' rather than *being* maths, and thus potentially undermining children's overall maths attainment. One teacher also commented on a potential knock-on effect on their coverage of subjects such as history and art due to catching up on maths after the programme. Teachers suggested that the linkage with maths might be clearer if SPACE sessions included weekly learning objectives or vocabulary.

> *The negative potential is having some of those basic maths skills that they need because they've been focusing so much on spatial and problem-solving, what about the number aspect of maths?* (SLT representative)

> *We've been trying to do a lot of catch-up in other lessons when we should have been doing art.* (Teacher)

One teacher moved some of their curriculum coverage around to facilitate the link between SPACE and their maths content:

> *There were some aspects of addition and subtraction that I've just had to park and I'm going to come back to, when we've completed the shape unit.* (Teacher)

This was not a universal issue—some teachers did not see the change in the use of the scheduled maths time as problematic. For example, one teacher commented that their class would be missing an hour of their usual maths time per week in the following term due to swimming lessons and that different types of learning and experiences are important to children's overall development. Other interviewees also reported that they were not significantly behind on their usual content and could see the benefit of spending time on a different kind of learning activity. In addition, the survey results reported in the section on perceived impacts for children (reported in the subsection on research question 7 under the 'Evidence of promise' section below), which show that no teachers identified that maths ability had got worse due to participating in the intervention.

Teachers and teaching assistants frequently recommended adapting the programme from whole-class delivery to delivery with a smaller group or as a lunchtime club for children who would benefit most from additional opportunities to use manipulatives, such as those with lower ability or EAL. Some even stated that in order to do the full programme again, they would want these adjustments to do it outside of their scheduled maths class time. We discuss this issue further in the 'Conclusion' section below.

Overall, while the SPACE sessions themselves were deemed implementable in terms of the knowledge and times required for the activity, the evaluation has identified issues with the feasibility of delivering it within planned maths lessons. Efforts to help schools plan for and embed SPACE in their maths teaching could mitigate these issues and improve perceived feasibility overall.

Evidence of promise

This second section of the findings focuses on whether there was initial evidence that the programme could achieve its intended aims and outcomes, as detailed in the programme's Theory of Change (Appendix A). This pilot trial did not explore all the elements of the Theory of Change, notably the direct child-level outcomes, and identified some areas of change not described in the Theory of Change. For that purpose, the findings are structured around three areas: changes for teachers; perceived changes for children; and potential mechanisms of change. Child-level results measured pre- and post-intervention are discussed in the section on the appropriateness of the whole-class assessments.

The discussion in this section draws on the teaching staff survey. As noted earlier, the baseline survey was administered immediately after the SPACE training rather than at the beginning of the training sessions as intended. This is likely to have affected baseline responses. We hypothesise that it may have led to an overestimate of some aspects at baseline (e.g. understanding of the importance of spatial skills) and could have led to an underestimation of change over the course of delivery (e.g. confidence in teaching spatial skills).

Research question 6: Do the training and delivery of SPACE sessions lead to changes in teaching staff's understanding of the importance of spatial skills, use of spatial language, and confidence in teaching spatial skills?

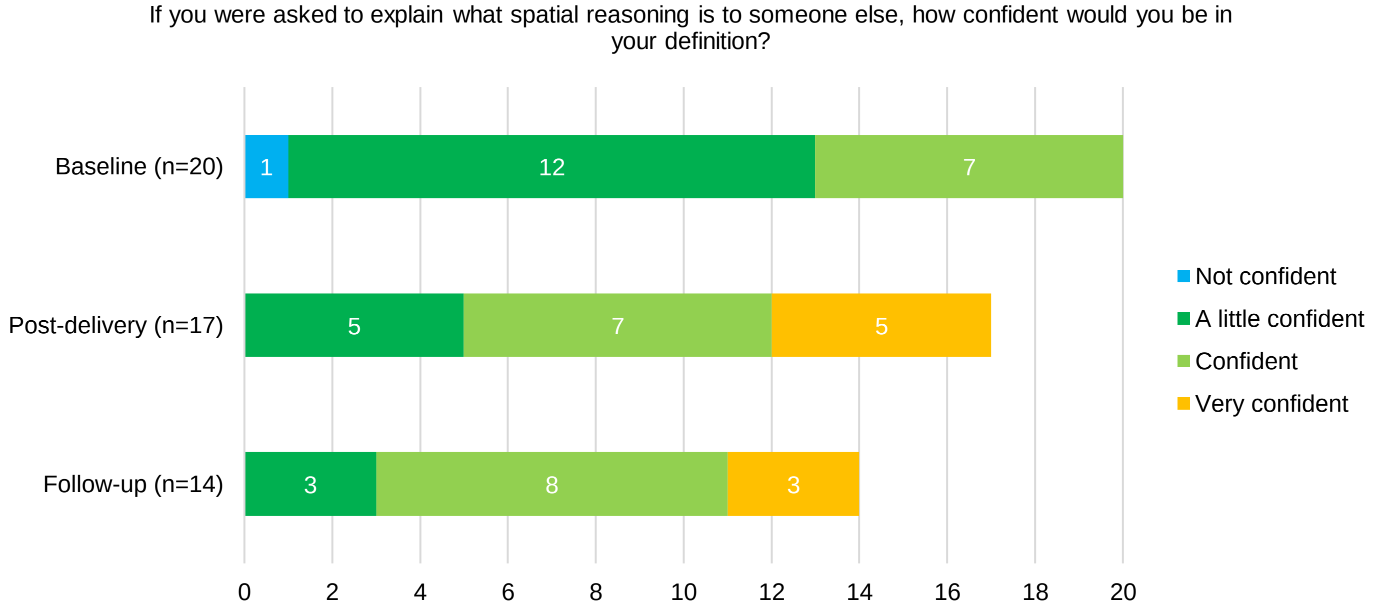
Data from the surveys, interviews, and delivery monitoring data suggest that the programme led to positive changes in teaching staff's self-reported understanding of the importance of spatial skills, their confidence in teaching these skills, and their use of spatial language. The evidence for each of these changes is reviewed under separate subsections below.

Changes in teachers' understanding of spatial skills

A key mechanism by which SPACE is intended to work is through teaching staff developing a deeper understanding of spatial skills and their importance for maths. Findings evidenced that participants indeed felt that training and implementing SPACE improved their understanding of spatial skills. This change was mainly assessed through three survey questions.

The survey asked participants how confident they would be in defining spatial reasoning if asked to explain it to someone else (Bates *et al.*, 2023). This question was asked at all three timepoints, and the results are shown in Figure 9. At all timepoints, the majority of participants reported that they were at least a little confident. Notably, confidence level increased between baseline (Timepoint 1) and post-delivery (Timepoint 2) despite the timing of the baseline assessment, so the actual increase may be greater. Some staff reported that they would be very confident after delivery, whereas none had reported this at baseline. The change in results between all timepoints showed improvements $F(2, 33) = 14.32$, $p<0.001$, $\eta^2 = 0.46$.[9] However, given the absence of a control group, this and other statistical results later in the report should only be interpreted as indicative and not to draw causal inferences.

Figure 9: Changes in confidence in defining spatial reasoning



The results of the interviews support the evidence from the survey—SLT representatives reflected that they had noticed a slightly better understanding of spatial skills among their teachers and that teachers' confidence in teaching these subjects had increased.

The surveys also asked participants how important they thought spatial thinking was for the development of children's mathematical skills. In all three surveys, teaching staff were asked to assign an importance value between 1 'not at all important' and 10 'extremely important' to spatial thinking. Figure 10 shows these results per survey.

Figure 10: Changes to the importance of spatial skills in children's development of mathematical skills

[9] For this and the other results, changes were derived from repeated measures ANOVA by transforming categorical (nominal) into ordinal (numerical) variables.

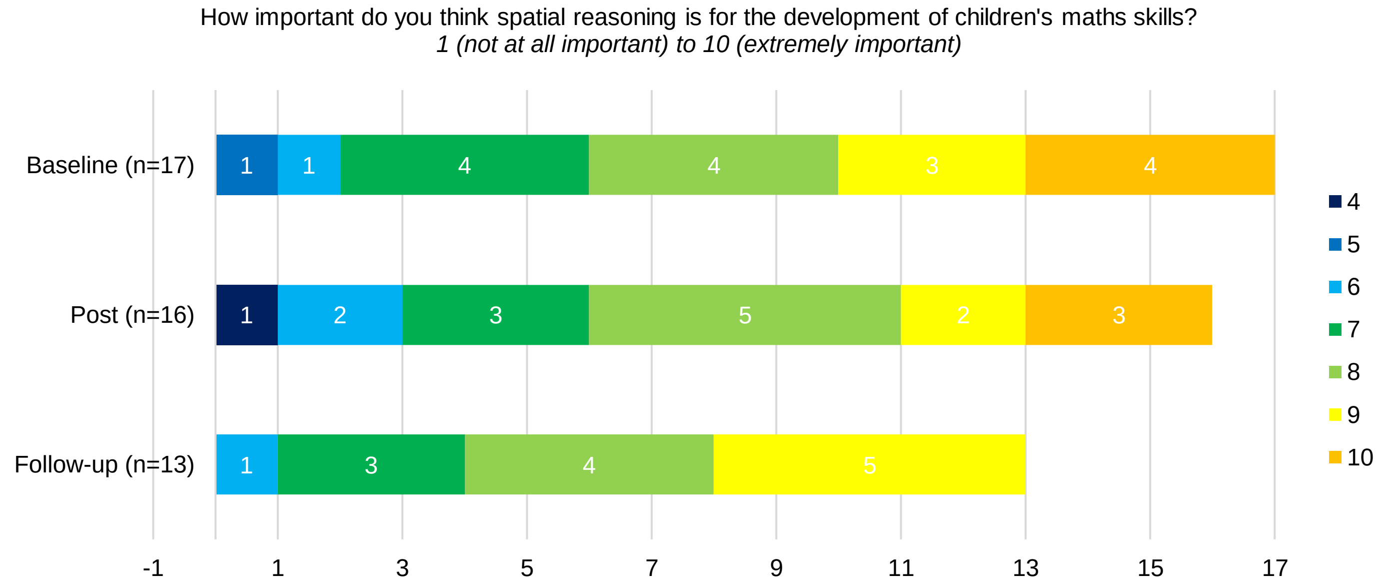
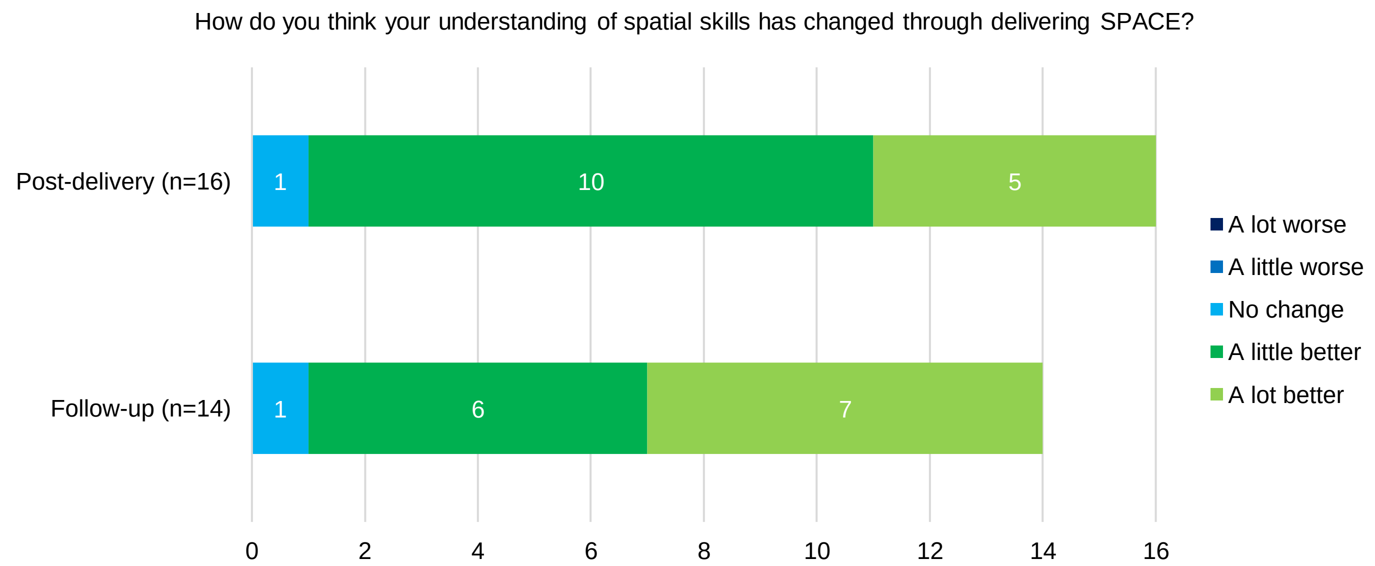


Figure 10 shows that most teaching staff considered spatial thinking to be of fairly high importance in the development of children's mathematical skills, across all three timepoints (averaging 8.1 for Timepoint 1, 7.8 for Timepoint 2, and 8.0 for Timepoint 3). Importance was rated at a minimum of 4 and a maximum of 10 out of 10 at all timepoints. There were no significant differences between timepoints ($F(2, 30) = 0.096$, $p = 0.908$), leading to the conclusion that participants generally considered that spatial thinking was very important for the development of children's mathematical skills, and that this perception remained stable throughout the study. However, we do not know what teachers believed before receiving the training.

To mitigate the effects of the administration error of the baseline survey, teaching staff were also asked about their own perceptions of change through delivering SPACE. Figure 11 shows the results for this question, which was asked after the delivery of SPACE (Timepoint 2) and after three months in the follow-up survey (Timepoint 3).

Figure 11: Self-perceived changes in the understanding of spatial skills



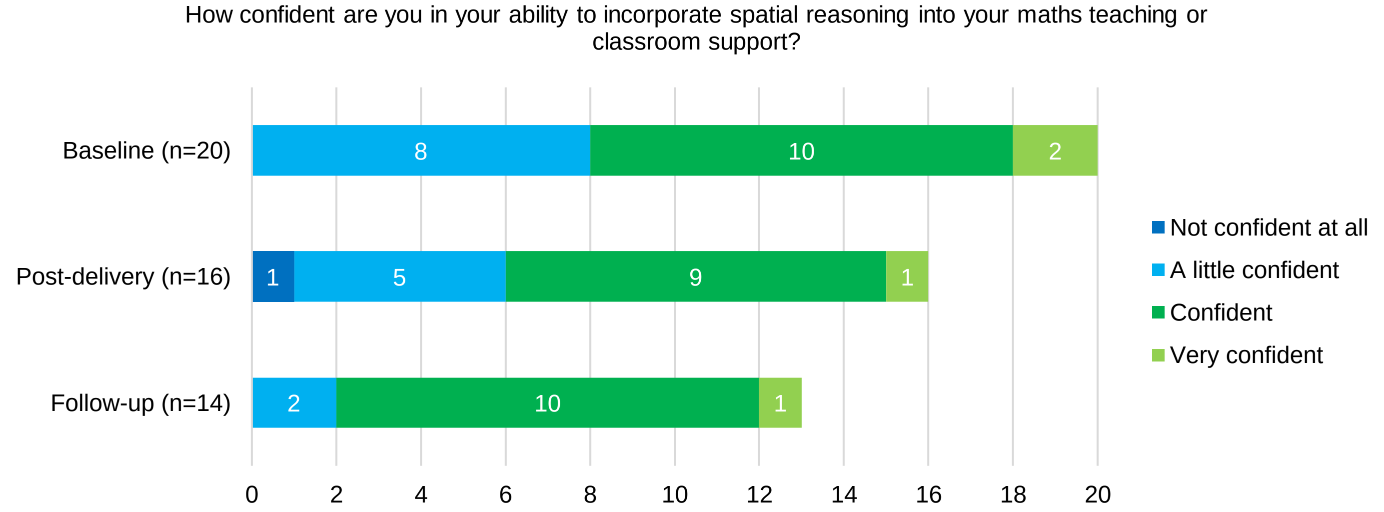
As shown in Figure 11, most participants felt that their understanding had got a little or a lot better (>90%), with only one respondent per timepoint reporting no change and none reporting that their understanding had got worse. Figure 11 also shows that teaching staff's self-reported improvement in understanding was sustained at the follow-up timepoint (Timepoint 3). A comparison of matched responses between post-delivery and follow-up showed that most participants provided the same answer at both timepoints, and two participants changed their answers from 'a little' to 'a lot' better.

Changes in teaching spatial skills

A second expected outcome of participation in SPACE for the teaching staff noted in the Theory of Change was increased confidence in delivering spatial activities and prompts. The results show that most of the teaching staff felt that their skills and confidence in incorporating spatial thinking had improved. A further change identified that was not described in the Theory of Change is that most participants reported that they had incorporated some spatial reasoning strategies or ideas into their teaching or classroom support following the implementation of the programme. These changes were mainly assessed through the analysis of six survey items, with findings detailed below.

Teaching staff were asked to rate their confidence in incorporating spatial reasoning into their maths teaching. Figure 12 presents the results for this item.

Figure 12: Changes in confidence in the ability to incorporate spatial reasoning into maths teaching or classroom support



The results in Figure 12 show that all participants felt at least a little confident across all timepoints, with the exception of one respondent in the post-delivery survey who reported feeling not confident at all. This participant did not complete the Timepoint 1 or Timepoint 3 surveys, so there is no further information to interpret this response. The results show a relatively stable distribution of confidence levels by time and statistical analysis across the three timepoints found no significant differences.[10]

To mitigate against spurious findings due to the baseline survey being administered immediately after training, participants were additionally asked to rate their perceptions of change between the baseline and post-delivery timepoints. These items looked at teaching staff's self-reported ability and confidence to incorporate spatial reasoning into their maths teaching or classroom support, elements not described in the initial Theory of Change. The results in Table 14 show that most participants at Timepoint 2 felt their ability or confidence had increased either a little or a lot. More importantly, for none of these items did participants report any negative changes.

Table 14: Perceived changes in ability and confidence to incorporate spatial reasoning attributed to delivering SPACE

Survey item	A lot worse	A little worse	No change	A little better	A lot better
How do you think your ability to incorporate spatial reasoning into your maths teaching or classroom support has improved through delivering SPACE?	–	–	5	7	4
How do you think your confidence in incorporating spatial reasoning into your maths teaching or classroom support has changed through delivering SPACE?	–	–	5	8	3

Supporting this finding, some interviewees also discussed an increased ability and confidence in incorporating spatial reasoning in their teaching. For example, some SLT representatives perceived, and others foresaw, increased confidence in teaching spatial reasoning:

[10] F (2, 33) = 0.04, p = 0.96.

…it will probably give her [teacher] more confidence in teaching the problem-solving element of maths. I would suspect it will give her more confidence and maybe encourage her to be more creative with some of those problem-solving questions. (SLT representative)

A fourth item in the post-delivery survey asked teaching staff about their intention to incorporate spatial thinking strategies or ideas into their maths teaching or classroom support after completing the SPACE programme. Extending the learning from SPACE to other areas of teaching was not included in their Theory of Change but would be a positive long-term outcome of the intervention. As shown in Figure 13, most teachers and teaching assistants responded that they intended to incorporate some strategies into their teaching practice or classroom support in the following term. Participants who indicated that they were planning to use spatial strategies were asked what strategies they planned to incorporate. In response, six participants indicated that they intended to incorporate the use of spatial language, three visualisations and two LEGO® as a pedagogical resource when teaching shapes and geometry. In line with this, interview findings revealed that some participants were interested in applying some of the strategies they used to support the delivery of SPACE when teaching other curriculum, such as using the prompt cards or working with LEGO. However, several interviewees commented that this future application would be dependent on the results of the evaluation or their access to LEGO resources.

In the follow-up survey (Timepoint 3), we sought to establish whether these intended strategies had been used and asked teaching staff what strategies or ideas about spatial thinking they had incorporated into their maths teaching or support since completing SPACE. In this case, 78.6% responded that they had used some strategies. No participants said they used a lot of strategies or ideas (see Figure 14). Participants who said they had incorporated spatial reasoning strategies into maths were also asked to specify, which strategies they had used. The most frequent answer related to the use of spatial vocabulary, and a minority also detailed the use of visualisations and manipulatives in demonstrations and constructions.

Figure 13: Intention to incorporate strategies for teaching spatial reasoning in the next term at post-delivery

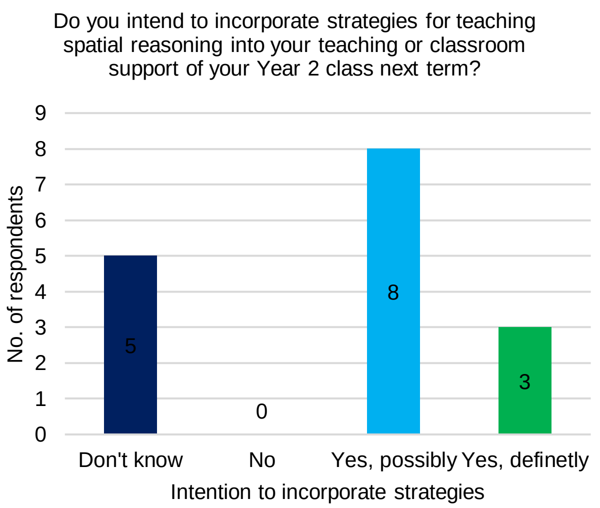
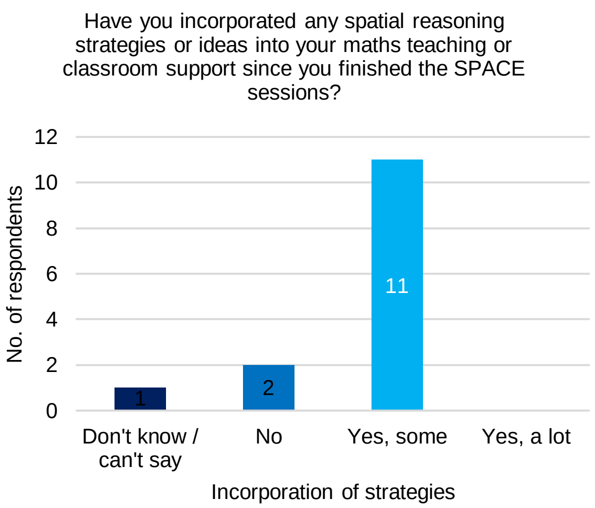


Figure 14: Incorporation of spatial reasoning strategies or ideas into maths at follow-up



Finally, a sixth item to explore whether SPACE influenced spatial reasoning teaching strategies was administered at baseline, post-delivery, and follow-up and explored how often teaching staff typically used eight activities that promote the development of spatial reasoning and maths skills (developed from Bates *et al.,* 2023 and Gilligan-Lee, Bradbury *et al.,* 2023). These activities were:

1. Using and explicitly teaching relational words, for example, ‘in’, ‘on’, ‘between’, ‘next to’, and ‘under’.

2. Building three-dimensional shapes with construction sets, for example, tiles, blocks, and LEGO®.
3. Using a number line with children.
4. Practising with turning and flipping shapes or fitting shapes together, for example, jigsaw puzzles.
5. Directing a toy or another child through a maze.
6. Comparing and describing the differences between two different pictograms or charts.
7. Measuring shapes and objects using a ruler.
8. Using drawing, counters, and/or arrays to solve multiplication and division questions.

Teaching staff were asked to indicate how frequently they used each activity in the classroom. No significant changes were found for most of these activities between timepoints, except for measuring shapes and objects using a ruler, where increased use was reported ($F(2, 33) = 3.81$, $p = 0.033$, $\eta^2 = .187$).

Changes in the use of spatial language

Finally, another expected outcome for teaching staff noted in the Theory of Change (Appendix A) was increased use of spatial language, discussion of spatial properties, and spatial relationships with students. Here, findings largely suggest that teachers did use more spatial language. They often expressed that they had become more aware of their choices about the specific language they used to explain concepts to their students.

> *I definitely thought more about the language that I was using when I was explaining to them how to make certain models, so it's definitely made me think more about the language choice that I make.* (Teacher)

In addition, as the following quote illustrates, for one teacher, the implementation of SPACE also had an impact on their choice of language outside maths.

> *I think it's just being more aware of the words that we're using around them all the time. So, even outside of maths lessons just when they're walking down the corridor using words like left and right, being overly conscious of using those types of words. It was quite interesting to see how many children don't know their lefts and rights. So, just trying to get those opportunities in as much as we can outside of maths, I think, as well.* (Teacher)

In the support sessions with the delivery team, teachers from some schools mentioned that they had noticed changes in their spatial language awareness because of the delivery of SPACE. Teachers from two schools reported that since the programme, they had used more spatial language, not only during SPACE and maths lessons, but also during the school day.

To conclude this section regarding teaching staff outcomes, we found evidence that the training and delivery of SPACE leads to positive changes in their confidence and understanding of teaching spatial reasoning and use of spatial language. Participants generally considered that spatial thinking was very important for the development of children's mathematical skills and this perception of high importance remained stable throughout the study. However, we did not find evidence that SPACE increases the perceived importance of spatial reasoning because the baseline survey was run directly after training on the importance of spatial reasoning.

Research question 7: What impacts do teaching staff perceive for children, particularly those from disadvantaged backgrounds?

Perceived impacts for children were explored through interviews and surveys with school staff. Assessing child-level impacts through teaching staff's perceptions was considered appropriate for this pilot study, as the focus is on understanding the feasibility and acceptability of the intervention. Direct measurements of child-level change and longer-term changes to maths skills were not included in the study design and would require an experimental design and validated measures, as proposed in later sections for the efficacy trial. Teacher perceptions are a useful tool to assess

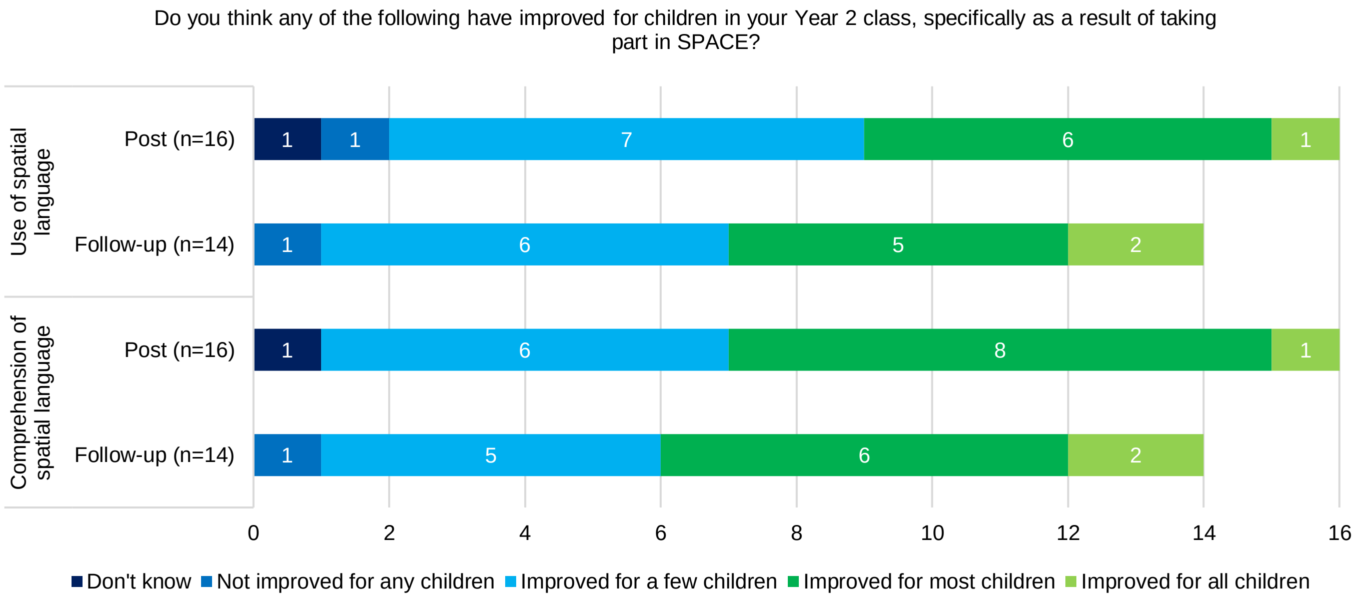
evidence of promise, as teachers are unlikely to sustain delivery of a programme they do not perceive as impactful. However, they can be susceptible to pre-existing biases and therefore, cannot be interpreted in isolation.
Interviewed teaching staff generally perceived positive improvements in their students, particularly in their ability to build LEGO® models in the SPACE sessions and their spatial language. To assess the more nuanced impacts of the delivery of SPACE on children as perceived by teachers, eight areas of potential change were explored. Some of these were directly related to the programme's Theory of Change: spatial language (production and comprehension); spatial reasoning; and overall changes in mathematical skills including problem-solving and geometry. In addition, we asked teaching staff about children's engagement in maths learning, and attention and concentration—outcomes not described in the initial Theory of Change.

This section presents findings regarding impacts for children as perceived by teaching staff. Results from the child assessments on problem-solving, geometry, spatial language, and mental rotation are presented in the subsection on research question 10 under the 'Readiness for an efficacy trial' section below).

Spatial language

Improved comprehension and production of spatial language were expected outcomes of SPACE's Theory of Change for pupils, deriving from their increased exposure to spatial language in the classroom, and the discussion of spatial properties and relationships from teaching staff. Teaching staff's view of children's spatial language skills was assessed through two survey items: children's use of spatial language; and their understanding of it. The results for both items, presented in Figure 15, show that most teaching staff perceived improvements in children's language and that these perceptions were maintained between the post-delivery and the follow-up survey. In the case of the spatial language use item, 88% noted an improvement at Timepoint 2 and 93% at Timepoint 3. At both timepoints, only one respondent reported no improvement for any child. For the item on spatial language comprehension, 94% of teaching staff who took part in the post-delivery survey and 93% of the participants in the follow-up survey reported perceived improvements in children.

Figure 15: Perceptions of improvement in children's understanding and use of spatial language



Findings from the interviews and monitoring notes generally support the survey data relating to language improvements. Interviewees frequently said that they had noticed children's improvement in their use and/or understanding of spatial language. The following quote exemplifies this idea:

> *I think their language improved. That was one of the very noticeable things. We're using words at any opportunity and it's quite nice when you hear someone you wouldn't expect, maybe a lower-ability child, to use even just the word horizontal. It's not a word that they would usually use or a word that they would have been exposed to otherwise without that teaching. (*Teacher)

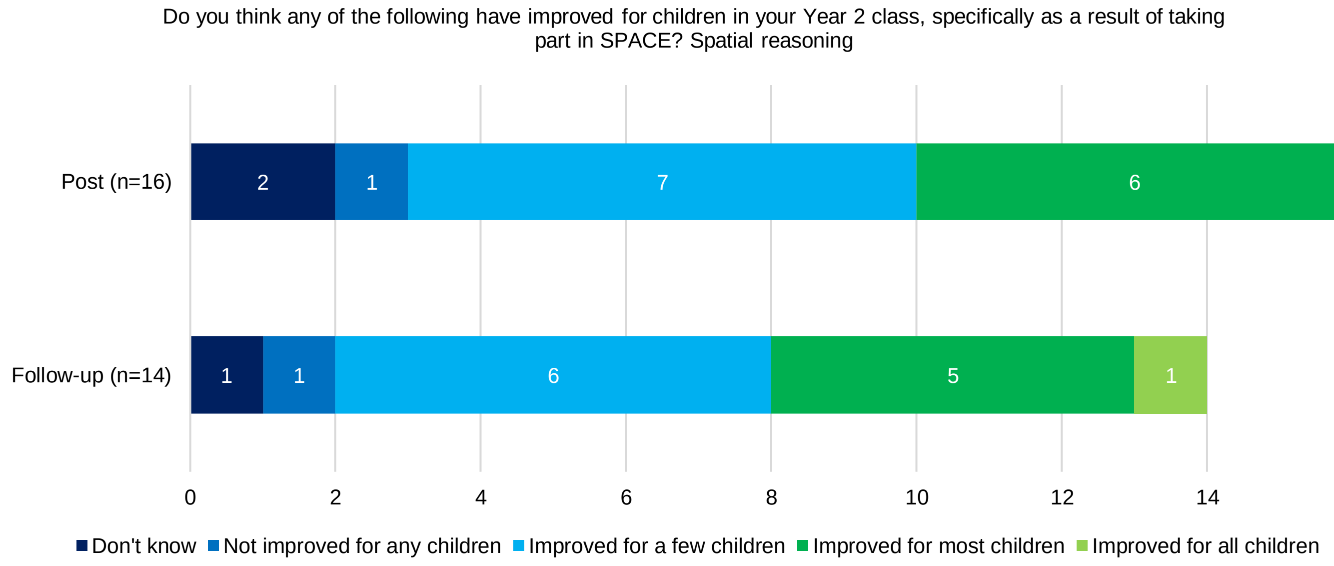
SLT representatives felt that one of the biggest changes in the children's vocabulary was in their use of mathematical concepts in non-maths contexts. Interviewees described children using 'richer language' compared to the previous Year 2 cohort, using spatial language when communicating with teaching staff, and in peer-to-peer communication.

However, other participants said that they were not entirely sure whether the children understood or used spatial language well, which may explain survey responses related to noticing improvements in only a few children. For example, some mentioned that children generally used this language with adults but not with each other, and others noted that children did not always use the words correctly.

Spatial reasoning

A second outcome for students in the Theory of Change was the improvement of spatial skills. These outcomes were expected to develop as students were exposed to spatial language during the sessions and used spatial reasoning skills to construct the LEGO® models. To assess changes in this regard, the pilot study collected information on teaching staff's perceptions of changes in pupils' spatial reasoning. Participants generally said they had seen improvements for at least some children. Specifically, the results of the post-delivery survey (Figure 16) showed that 81% of participants reported perceived improvements in children's spatial reasoning with a similar distribution at follow-up (86% reported improvements). At both timepoints only one respondent indicated that they did not think that any of the children had improved their spatial reasoning due to their participation in SPACE.

Figure 16: Perceived effects on spatial reasoning



The interviews support this finding of positive impacts on students' spatial reasoning. For example, several SLT representatives and teachers said that they had noticed impacts on the way students described building the models to their peers.

> *In terms of their progress in the programme, I definitely think that yes, their spatial reasoning has improved.* (Teacher)

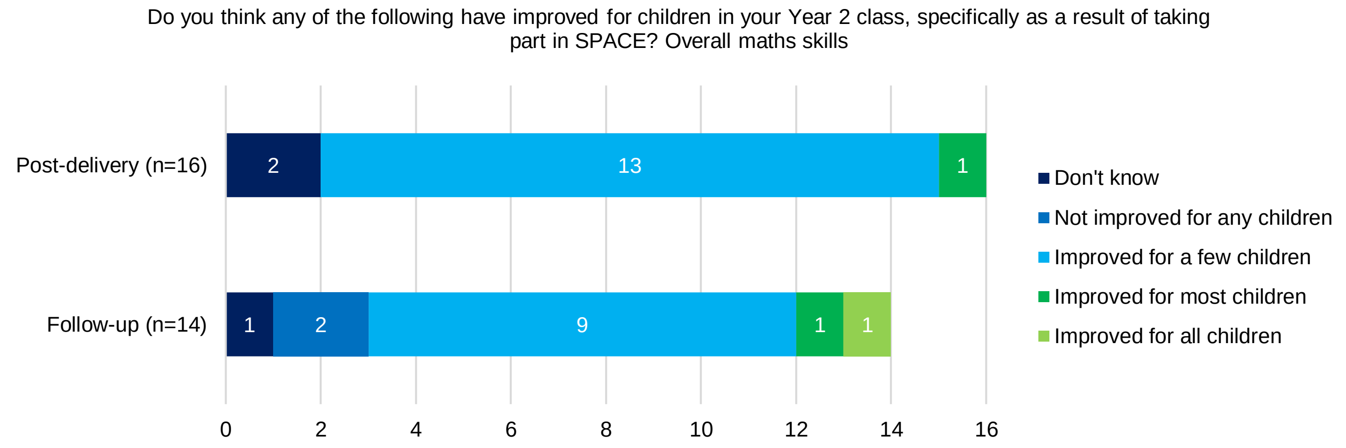
Likewise, information from support sessions backs this finding. For example, teaching staff talked about noticing changes in children's spatial awareness, for example, through their greater mastery of the concept of rotation.

Maths skills, problem-solving, and geometry

The SPACE Theory of Change also hypothesises that an improvement in spatial thinking and language skills, as a long-term outcome, leads to an improvement in maths outcomes, including in the areas of problem-solving and geometry. In this section, we explore these outcomes through teaching staff perceptions reported during and shortly after the intervention. The longer-term changes anticipated in the Theory of Change would require follow-up data collection at appropriate timepoints using objective measures not within the scope of this pilot study. In relation to improvements in overall maths, participants generally noted perceived improvements for a few students. However, there was more confidence that SPACE would lead to future improvements in spatial thinking and language.

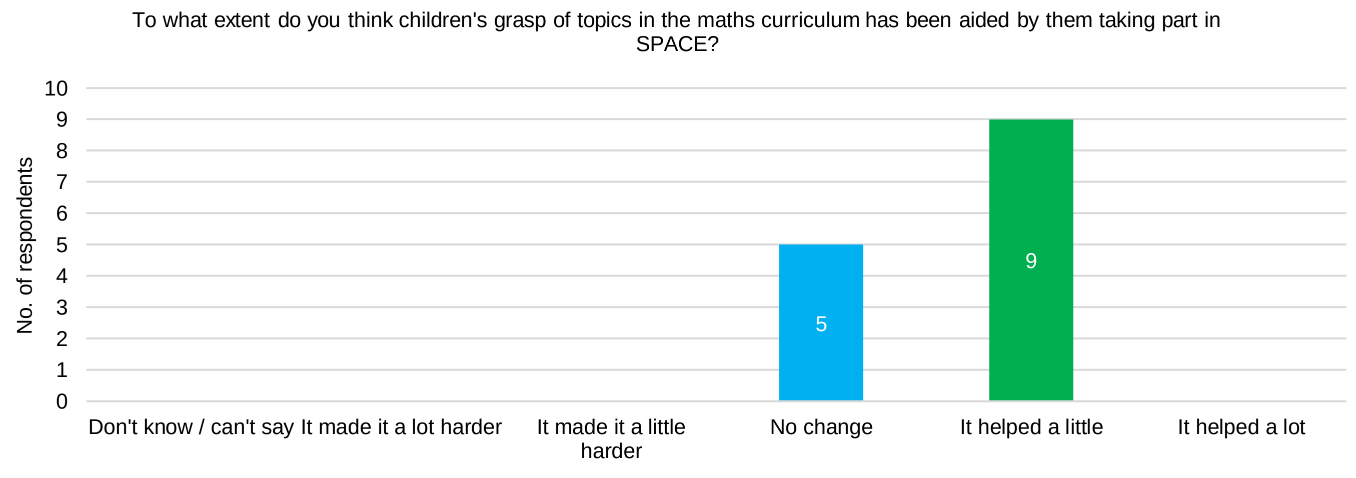
Perceived changes in children's overall mathematical skills were assessed in the teaching staff survey at post-delivery and follow-up timepoints. Figure 17 shows these findings. Most participants at post-delivery indicated that they perceived improvements in overall mathematical skills in at least a few children due to their participation in SPACE (81%). In the follow-up survey, this distribution changed in some ways, but most participants still indicated perceived improvements overall for at least a few.

Figure 17: Improvements in overall maths skills



The follow-up survey also asked about teaching staff's perspectives on children's grasp of the maths curriculum. Figure 18 shows that 64% of teaching staff believed that participating in SPACE helped their children 'a little' in understanding the maths curriculum.

Figure 18: Aid of SPACE in children's grasp of maths topics



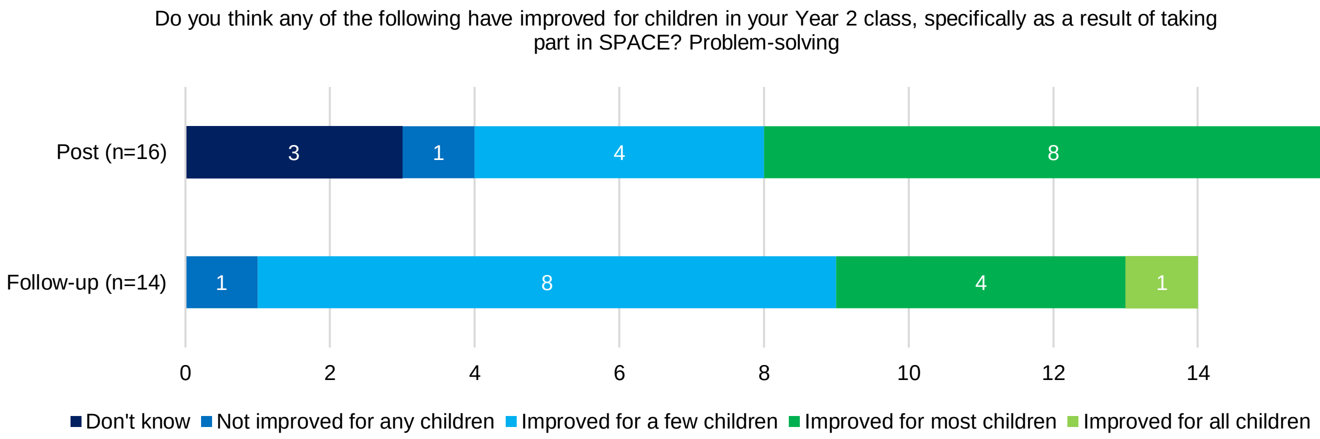
However, the qualitative interviews presented a more mixed picture. Most interviewees said they were not confident that SPACE had improved children's maths skills, particularly in the topics covered in that term: number; addition; and subtraction. Several SLT representatives and teachers said that they did not feel they could say whether there had been improvements because children were having maths lessons alongside SPACE, making it difficult to isolate the effects of SPACE. A small group of participants said that they were not convinced that there would be improvements in maths related to SPACE, because they did not perceive SPACE and maths as being obviously related. This latter finding may be linked to the themes identified in the 'Feasibility of implementation' section mentioned earlier, around teachers' scepticism of the link between SPACE and maths. Nonetheless, a few interviewees said that even though they had not yet seen changes in maths skills, these might become evident when they taught topics such as multiplication and shape later in the school year.

In the Timepoint 2 and Timepoint 3 surveys, participants were also asked whether participation in SPACE had made it harder for children to grasp topics in the maths curriculum and whether it had worsened children's overall maths skills.

In both the post-delivery and follow-up surveys, no participants reported any worsening and one reported that they did not know.

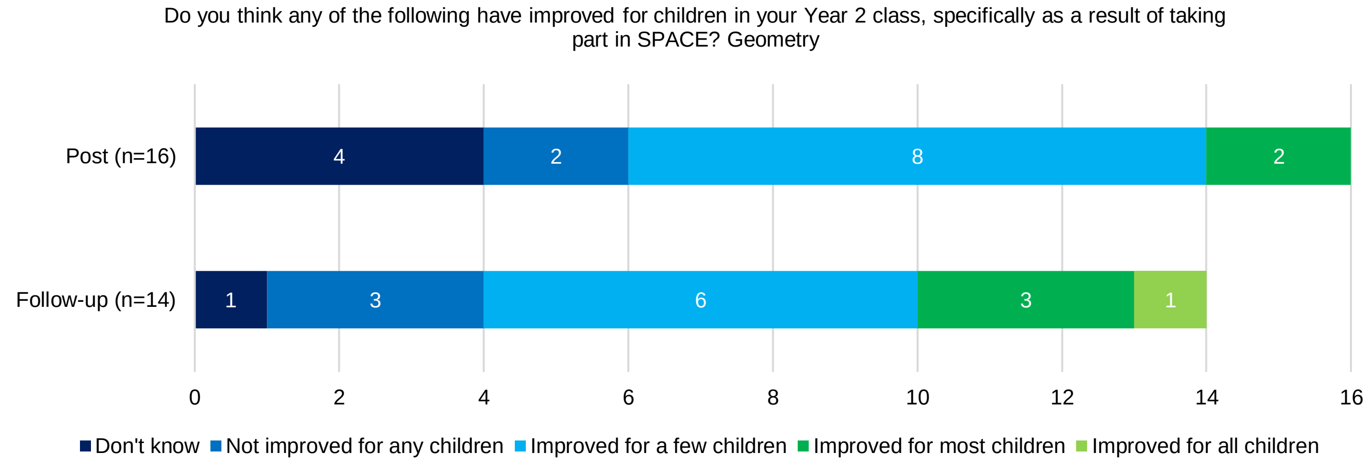
Turning to the specific points about impacts on problem-solving and geometry, Figure 19 shows that most participants perceived improvements in problem-solving for at least some students as a result of taking part in SPACE, both in the post-delivery survey (75%) and in the follow-up survey (93%). However, the percentage of participants who noticed changes for most students decreased between timepoints (from 50% to 29%) and the percentage of participants who noticed changes in a few children at follow-up increased, which could signal that these changes in some students were not sustained over time. In the interviews, a few teachers and teaching assistants noticed changes in their pupils in problem-solving. For instance, one teaching assistant noticed that the children showed more problem-solving resourcefulness when correcting their mistakes in positioning the LEGO® pieces. Similarly, in the support sessions, teachers mentioned that the children had improved their construction skills by taking time to observe the models and find solutions.

Figure 19: Perceived effect on problem-solving



Finally, concerning changes in geometry, the post-delivery survey results (Figure 20) showed that 63% of participants thought that some of their children had improved in geometry because of SPACE. However, 25% said they did not know if there was an impact. Follow-up results show a slightly different distribution of answers—71% of respondents noted that they had perceived improvements, while only 13% did not notice improvements for any children. These changes were statistically significant but with a small effect size, (t (12) =-1.17, p<0.001, d = -0.307).

Figure 20: Perceived effects in geometry



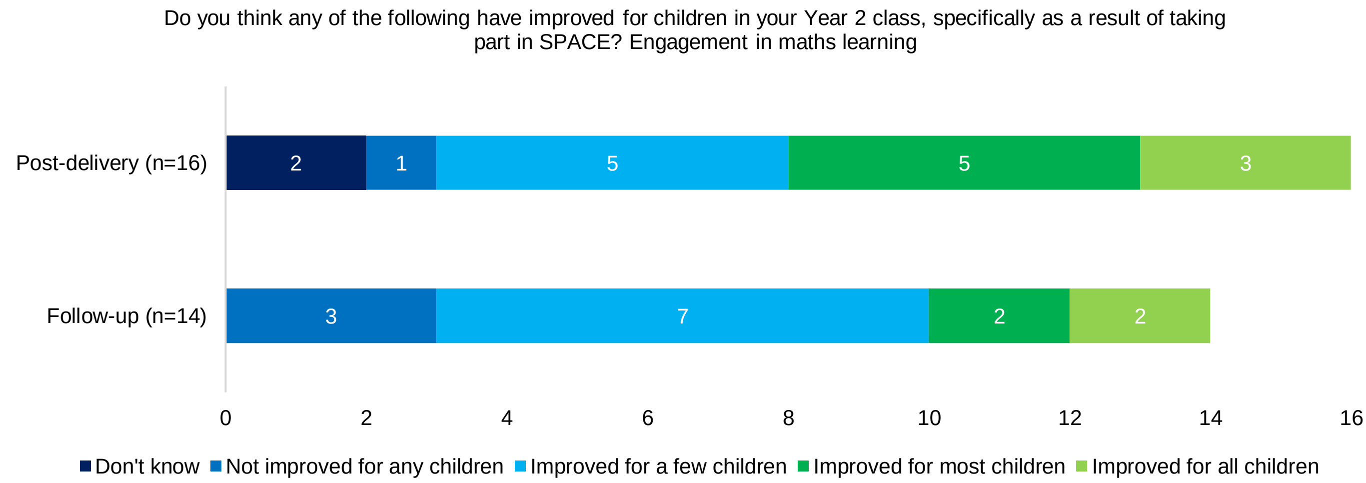
Qualitative evidence suggests that some participants expect more widespread improvements in geometry in the longer term. This aligns with the Theory of Change, which proposed change in maths as a long-term outcome. In interviews and support sessions, teachers mentioned that although it was too early to see any impact in geometry, they believed that when they covered this content in the Spring Term, students were more likely to find it easier than previous cohorts. Some interviewees specified they see the potential for improvement in topics such as symmetry and rotation.

Engagement in maths, and attention and concentration

Potential changes in students' engagement in maths, and attention and concentration were also explored—areas of interest not described in the initial Theory of Change. Survey findings in these areas indicate some changes for some students.

Regarding engagement in maths learning, Figure 21 shows that at post-delivery (Timepoint 2), 81% of participants in the teaching staff survey perceived improvements for at least some of their pupils and 50% for most or all children. However, the percentage identifying improvements in most or all children dropped to 29% in the follow-up survey, and the percentage of participants who saw no improvements for any children increased.

Figure 21: Impacts on engagement in maths



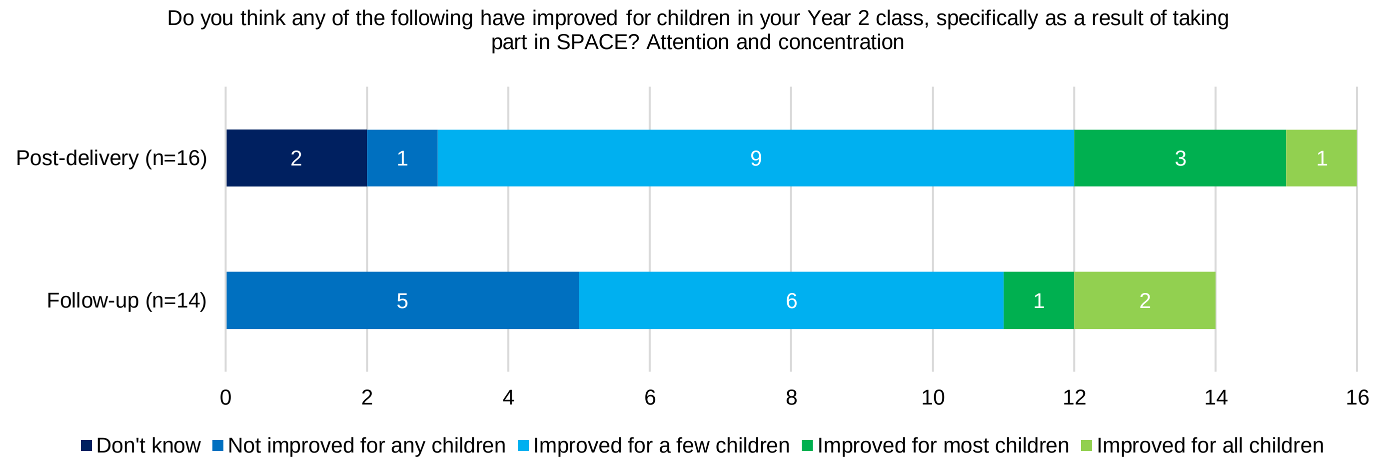
In interviews, one teacher commented that their students were more interested in being in the classroom because of SPACE.

The interview and monitoring data findings revealed a perception that pupils were more confident in themselves and their ability to perform well in SPACE. For example, in their final support session, one teacher shared that one of their low-achieving pupils had enjoyed every LEGO® session and had grown in confidence in maths in general over the six weeks. Interviewed teaching staff also repeatedly mentioned that they had noticed children's greater capacity for perseverance in model building, and a greater tolerance of frustration and resilience. The following quote illustrates this view:

> *I would say over the course of doing the sessions, it definitely impacts their resilience and perseverance to just doing a task, I think. Whereas a lot of them were getting quite upset and frustrated that they weren't completing them, by the end, we did see more of that determination and resilience that it's fine if they don't.* (Teacher)

Survey results also suggest that pupils' attention and concentration improved through participating in SPACE. Figure 22 shows that 81% of the participants in the post-delivery survey felt that at least some of their pupils had improved, though only 6% felt it had improved for all children. While some of this impact was sustained in the follow-up survey, the percentage who saw no improvement for any children increased from 6% to 36% between the two timepoints. The comparison between the two survey times showed a statistically significant difference between the survey administrations, but with a very small effect size (t (12) =0.321, p<0.001, d = 0.077).

Figure 22: Impacts in attention and concentration



Negative outcomes

Equivalent items were also included in the survey to explore whether participants had perceived any negative changes in children's spatial language, spatial reasoning, problem-solving, geometry, attention and concentration, and engagement in maths learning. Significantly, participants did not perceive any negative effects for any children in any of these areas at any timepoint. The interview findings supported these findings. No SLT representative, teacher, or teaching assistant reported any negative impact on pupils in any of the reported areas because of their participation in SPACE. Open-ended survey questions at post-delivery reflected thoughts in line with these findings. While some participants noted that a potential disadvantage of SPACE was that children had missed some of the maths teaching during the term, they did not consider that this was likely to have a negative effect on maths ability in the longer term.

Differential impacts on children

As reported so far, across the interviews, surveys, observations, and support session notes, there were consistently positive reports of improvements for the children who took part in the SPACE programme, but these were usually noted for some or most of the pupils (but not all of them). The qualitative evidence points to two possible dimensions that the teaching staff felt could explain variations in improvement between the children in their classrooms: level of ability; and prior familiarity with LEGO®.

First, several teachers and teaching assistants said that those who made the most significant improvements in their classrooms were the children who started with a lower level of attainment—in maths or overall. Moreover, some teachers reported that their students with lower ability found the programme more accessible and less anxiety-provoking than regular maths lessons because they saw it as play and were not disheartened by previous associations with finding maths challenging.

Some teachers also noted that some of their typically higher-achieving students were more challenged by SPACE than expected. Some teachers felt that children with lower abilities initially had an easier time with SPACE because they had more experience in using visual representations rather than written instructions. The following quote exemplifies this view:

> *…the higher abilities tend to struggle. They weren't maybe thinking and looking at it in the same way, and it was the children that really struggle with reading the ones that were being the most successful, because they weren't looking for the written word, they were following the visual representation on the page, and they felt really...It gave them a lot of confidence.* (Teacher)

Other interviewed teachers and teaching assistants felt that their higher-ability students had less room for improvement in SPACE than those with lower abilities.

There were different views about the impacts for children with SEND, which varied per case. Some teachers and teaching assistants mentioned some of these students really benefited from the programme, while others struggled to engage with it at all. As noted earlier, some children with more challenging emotional regulation problems, regardless of their ability level, had been less able to engage with the programme, and teaching staff thought they had not benefited from the programme as much as those who were able to engage throughout.

Previous familiarity with LEGO® was a recurring theme to explain how different children benefited from the programme. One teaching assistant felt that children who had LEGO in their homes had benefited less or little because they were already very strong in spatial skills. Children from more disadvantaged backgrounds without LEGO or similar toys were thought to be more likely to benefit from SPACE as they were not getting this kind of input at home. However, we did not see systematic differences in children's outcomes between schools with different levels of children receiving FSM, nor across the two delivery locations—one of which is an education improvement area.

Unanticipated changes in children following their participation in SPACE

Finally, the qualitative evidence also revealed changes in two other areas that were not pre-specified or identified in the Theory of Change. The most frequently mentioned outcome, identified in open-ended questions in surveys, interviews, and monitoring information, was the improvement in the children's fine motor skills. Some teachers noted improvements in children's handwriting, which they attributed to their greater dexterity in manipulating small objects and strengthening of the children's hands. The second unanticipated area of improvement reported was that children's independence and perseverance at completing a challenging task had increased, meaning they were more confident in completing tasks without support from an adult.

Research question 8: What are the perceived mechanisms and possible causal chains?

In this section, we discuss the perceived outcomes for children in the context of the SPACE Theory of Change (Appendix A) to provide further understanding of the perceived causal chains or mechanisms. The Theory of Change describes the overall aim of the SPACE programme: 'The programme aims to improve spatial thinking and spatial language skills using structured LEGO® play, leading to improved maths outcomes including arithmetic, geometry, and mathematics problem-solving' (Appendix A).

To summarise the Theory of Change, for teachers and teaching assistants, the training and experience of delivering the intervention in class are thought to improve their confidence, understanding, and perceived importance of spatial reasoning and lead to increased use of spatial language with their Year 2 pupils. For children, their participation in the SPACE LEGO® sessions is thought to increase their exposure to spatial language (through teachers' use) and to draw on their spatial thinking skills (e.g. visualisation) to build the models. These processes are expected to improve children's comprehension and production of spatial language and their spatial reasoning ability, including mental rotation, visuo-spatial working memory, embedding, and spatial-numerical relationships in the short-term. In the longer term, these outcomes are anticipated to lead to sustained improvements in children's spatial language and spatial reasoning and, subsequently, their mathematical ability, which were not assessed in this pilot evaluation.

Overall, the causal chains for achieving the short-term outcomes described in the Theory of Change and assessed in this evaluation are well supported by the evidence. We discuss these first for staff and then for children.

Mechanisms and possible causal chains for teaching staff

As discussed in research question 6 (see above in this section), there is supportive evidence for improved confidence in teaching spatial thinking—attributed by teaching staff to the regular repetition of the sessions.

> *If there's ever another programme or anything like that, I would feel more confident going into it and taking it on and understanding it. […] I think the consistency of it, the fact that you do it regularly on a regular basis, and there's not massive changes, so it gives you a chance to get to know the programme properly and build that confidence to then be able to deliver it.* (Teaching assistant)

There is evidence to support the proposition that the training would increase teaching staff's confidence in delivering the intervention (see research question 2 under the subsection 'Feasibility of implementation'). There is partial support for the training improving teachers' perceptions of importance of spatial thinking for maths. Due to the mistiming of the baseline survey, it is not possible to assess the influence of the training on this element—it was perceived as important at baseline and throughout. Qualitative insights reveal that while teachers valued learning about the rationale for how the programme could help maths, some of the training participants felt that the communication of the relationship between spatial thinking and the maths curriculum needed further development.
In addition, the analysis of research question 6 showed almost no variance between timepoints in terms of teachers' understanding. However, we are unable to determine whether this could be explained by the fact that the Timepoint 1 survey was administered after the training sessions, limiting the ability to assess changes resulting from the training or whether this indicates that the training did not lead to a deeper understanding of this issue among teachers.

Teachers' and teaching assistants' experience of training and then delivering the intervention was also expected to increase their use of spatial language in the classroom. Qualitative interviews with teaching staff, as well as observations of SPACE delivery, provided evidence of support for this mechanism (see research question 6, above in this section).

Teachers' increased confidence in delivering SPACE and their perceived importance of spatial thinking in maths were also anticipated to increase their use of spatial language, discussion of spatial properties, and spatial relationships with students, which was also supported by evidence from the qualitative interviews. Interviewees reported that their increased confidence in delivering SPACE encouraged them to use more spatial language with pupils. This was reported as the key new strategy used by teaching staff to support their pupils during lessons.

> *Not modelling it as much has been interesting. At first, I was kind of like just, 'I want to show you where that bit goes', or 'Let me show you how to do it first, and then you can copy me'. Doing that was challenging, but interesting because it's making sure I'm using the correct language and in a way that they're going to understand it, talking about the different places.* (Teacher)

Mechanisms and possible causal chains for pupils

The Theory of Change also describes potential mechanisms through which the programme can lead to short- and long-term outcomes for children.

The Theory of Change anticipates that children's exposure to spatial language and active engagement of their spatial thinking skills to build the models will lead to improved comprehension and production of spatial language and better spatial ability overall. There is supportive evidence for this. The teaching staff interviewed explicitly made connections between children's use of manipulatives in the SPACE sessions and their development of spatial thinking in other maths learning, particularly when engaging with visual processes like symmetry and copying patterns. They also noted that building models from the visual instructions supported spatial learning as a distinct mechanism from playing with the manipulatives, as children may already do at school or at home.

> *They [children] seemed to grasp it [shape and geometry unit] very quickly and they did very well. Whether that's a direct impact I don't know, because of the manipulation with the objects, and being comfortable using the resources.* (Teacher)

Interviewees noted that the increase in teachers' use of spatial language, even without explicit teaching of new vocabulary, could influence children's understanding and production of spatial language. For example, SLT representatives commented that they believed the impact of SPACE on pupils' language was due to the normalisation of using spatial language by teachers in sessions when they had to describe how to build models.

> *They [children] were talking about what they were doing, without knowing that they were talking mathematically, which was really good.* (SLT representative)

One teacher proposed that the association between teachers' and children's spatial language might be due to children reproducing teachers' conversations when they talk with other children, which would build up their spatial oracy.

> *I am having that talk with lots of different children as I'm moving around, and they quickly start emulating the conversations that I have.* (Teacher)

A teaching assistant felt that the prompts and questions that teaching staff posed to children throughout the sessions led children to think and problem-solve differently. For example, *'What do you think? How would you start? What would you put next and why? and Why didn't you put it there?'* (Teaching assistant). They also felt that requiring children to give full answers supported their development.
Others noted that spatial language production might take more time to develop in children than language comprehension, explaining that children were more likely to understand and use spatial language in conversation with teaching staff than to use it with their peers spontaneously.

> *They [children] can speak to an adult when an adult is prompting the language, but very rarely—even in any subject—do we hear them using the language to each other when an adult's not around. I think if we wanted more of the language, you would need more adult input around the room.* (Teacher)

Interviewees also identified other potential causal mechanisms and outcomes for positive change related to participation in the SPACE intervention. Engagement and confidence were seen as key to children's progress. The playful nature of learning with LEGO® was said to have facilitated their engagement with SPACE and helped them to benefit from it. In an observed support session, a teacher described in detail how the intervention had helped some pupils with lower maths ability to realise that maths is not 'only about numbers', which had improved their engagement in maths sessions and was expected to support their attainment overall.

Repetition of the same routine in sessions facilitated growth in children's confidence in model building. Some children with lower ability were unexpectedly good at model building, which gave them confidence that translated into other areas of learning. Finally, the programme was also seen to improve children's fine motor skills, which may in turn benefit progression in other areas, such as in maths and other subjects.

Overall, the evaluation provides support for most of the causal chains proposed in the SPACE Theory of Change for teachers and pupils, as perceived by teaching staff, in terms of the short-term outcomes. One causal chain had less supporting evidence in the link between children's use of spatial thinking skills to build LEGO® constructions and perceptions of children's improved maths skills. This could be explored further in an efficacy trial by improving the training materials to amplify that link and also by measuring child-level maths outcomes compared to a control group. Furthermore, the link between training and teacher outcomes could be understood more reliably by ensuring a true baseline of teachers' understanding is recorded. Due to the timeframe of the evaluation, it is difficult to assert the long-term, sustained effects of these outcomes, so this remains a matter for further study.

Readiness for an efficacy trial

In this section, we report on two issues relevant to readiness for an efficacy trial: i) whether SPACE can be delivered on a larger scale as an efficacy trial; and ii) whether the child assessments are appropriate for use in an efficacy trial.

Research question 9: How could SPACE be delivered on a larger scale?

We addressed this research question through considering the data and findings about fidelity, feasibility, and acceptability, and drawing on the group interviews with key members of the delivery team. The findings of this pilot evaluation and the scaling intentions of the delivery team have several implications for the level and type of programme development and changes to the delivery model required.

The scalability of a programme is dependent on several factors (e.g. see Kohl and Cooley, 2012; Milat *et al.,* 2020; VVOB, 2021). Key considerations are the nature of the problem, the extent to which the intervention addresses identified needs, alignment of the intervention with the relevant delivery setting and the wider policy context, the relative cost of the intervention, whether the intervention can be adapted to different delivery contexts and maintain its effectiveness, the ability of the intervention to reach the intended population and its acceptability to them, and the delivery model (training, materials, support, and monitoring).

The long-term goal of the developers is to get spatial reasoning prioritised and embedded in the EYFS statutory framework and primary school curriculum, through the adoption of teaching strategies that involve ‘spatialising’ the maths (and wider) curriculum. That is, to incorporate spatial reasoning into teaching methods within maths and other curriculum areas to improve children’s maths outcomes. If spatial reasoning was successfully embedded in this way, the SPACE intervention itself would not necessarily be delivered at national scale. Equipping teachers to teach spatial thinking would then become part of wider teacher training and professional development, and less dependent on specialised manualised programmes such as SPACE.

Influencing policy will require rigorous evidence regarding the impact of spatial thinking training in a whole-class setting on children’s maths attainment. Assuming evaluations demonstrate positive effects, the level of evidence required for this purpose may only entail an efficacy trial or may also need a full effectiveness trial after an efficacy trial. In either case, the SPACE intervention and delivery model should be available in a scalable format to support teachers’ acquisition of teaching strategies. Development work is therefore, required before the next stage of scale up and evaluation to test the feasibility of any adaptations made to improve scalability.

Both establishing an evidence base on the impact of integrating spatial thinking into the curriculum and delivering the intervention itself at scale raise a number of issues relating to delivery capacity and development of the programme model for scale:

- **Delivery team capacity**. The delivery team have the capacity and infrastructure to train teaching staff and support delivery to around 50 schools, without significant modifications to the current model, with optimisations to the delivery model in terms of recruitment, equipment distribution, training, materials, and support processes discussed in the ‘Conclusions’ section below. To deliver at scale beyond that, more substantial changes to the model (such as using a train-the-trainer model) and/or partnering with one or more organisations to support delivery would be needed.
- **Training model.** The delivery team has the capacity to train a sufficient number of schools for an efficacy trial using face-to-face training as delivered for the pilot study. Initial sample size estimates based on a small effect size (Cohen’s d = 0.2[11]) would require a trial of 150 schools, half of which would receive the intervention. However, larger scale delivery would require moving to a train-the-trainer model and/or online delivery. It is not clear that the delivery team would have the capacity to support a train-the-trainer model at large scale, and this might involve working with a partner organisation that has the required capacity and experience. In the pilot study, teachers widely welcomed the face-to-face training model, which implied that a train-the-trainer model might be preferrable to remote delivery.

  The optimal training model requires more exploration by the delivery team. We recommend that variants of the selected training model are tested at the next stage of evaluation (e.g. efficacy) to identify the most appropriate one.
- **Demand and recruitment.** Addressing disparity in maths attainment was identified as a strategic priority among the schools participating in this evaluation. The delivery team have capacity and networks to recruit schools at larger scale, with the support of the EEF. The recruitment process could be streamlined by focusing more on the successful recruitment pathways and less on personalised emails.
- **Integration into maths teaching.** Although the programme was overall very feasible for delivery in itself, a key challenge was integrating it into maths teaching so that it is seen by teaching staff as a way of teaching maths rather than viewed as ‘additional to’ planned maths teaching. This would require schools to be recruited earlier, with specific support given to building SPACE into planned teaching in advance of the start of the school year. The evaluation identified the need for the programme content to be more closely aligned with the current maths curriculum, and this would need to be highlighted and supported in the training.
- **Disadvantaged pupils.** Although perceptions were mixed, the pilot trial did not raise significant concerns about the feasibility of the programme for disadvantaged pupils, and there was a perception that these pupils might have most to gain from SPACE. However, there would be value in including more guidance about how to support children who initially find it difficult to engage with SPACE.
- **Equipment.** The provision and distribution of the LEGO® equipment is likely to become a bottleneck in terms of large-scale delivery of SPACE. The materials are expensive to obtain, and it is unlikely that schools would be able to fund this themselves. The logistics of assembling the boxes for each

[11] Based on a power calculation (https://powerupr.shinyapps.io/index) using the following standard parameters to estimate sample size based on effect size d = 0.2: two-level cluster randomised controlled trial, power = 0.8, alpha = 0.05, 1:1 random allocation, average 25 students per school, prop. variance in outcome explained by level 1 = 0.5, prop. variance in outcome explained by level 2 = 0.2, prop. variance in outcome between level 2 units = 0.2.

child are also challenging at scale. These challenges could be overcome at efficacy trial scale with sufficient funding from the EEF (possibly with a funding partner, see 'Conclusions' section below). However, this becomes a significant challenge to scaling the programme beyond efficacy trial. We have identified two options for addressing this, both of which are being considered by the delivery team. The first is to form a partnership, with the LEGO Foundation (the principal investigator has existing links with LEGO) or with a commercial supplier of LEGO or similar physical manipulatives through, which the equipment would be made available. Capacity would also need to be developed for assembling boxes at scale. The second option would be to consider developing the programme with digital rather than physical equipment. The implications of this on the efficacy of the intervention would need to be carefully explored. The previous BLOCS study found an increased effect for the digital version compared with the physical version of the intervention, however the BLOCS study was impacted by factors related to the Covid-19 pandemic. A recent meta-analysis of 29 spatial training studies clearly demonstrates the strong benefits of spatial training with physical manipulatives over digital interventions (Hawes *et al.,* 2022).

- **Support for delivery quality.** The current model of support and monitoring is resource-intensive for the delivery team and a less intensive model will be needed for delivery at scale. Although the support sessions were well received and beneficial, the evidence from teachers suggests they could be delivered less frequently and online without reducing their benefit. There would also be value in exploring whether teachers trained in and using SPACE might provide support to local school clusters as 'SPACE champions', also providing the core delivery team with feedback to identify implementation issues and schools that might need more support. The feasibility and effectiveness of this approach should be tested at efficacy trial, including whether other schools would be willing to take on this role and its consequences for the quality of delivery.
- **Monitoring.** Monitoring information was collected in face-to-face support sessions and via email in the pilot delivery, though some data regarding delivery and assessment sessions was missing. A more streamlined system involving digital data collection would be required for the programme to be delivered at scale, with sufficient delivery team monitoring support to minimise missing data.
- **Fidelity.** The high levels of fidelity suggest that it would be possible to replicate SPACE at scale with high quality. This pilot did not identify any concerns regarding the ability of schools with high levels of pupils receiving FSM to deliver the programme as intended.
- **Delivery team vision for scale.** Having a clear vision and the motivation to scale an intervention are key determinants of scaling success. The SPACE delivery team is actively pursuing scale for the core elements of the SPACE intervention. A documented plan, including a timeline and the required actions for the preferred scaling pathway would help the team to translate this motivation into successful scaling.
- **National context.** The delivery team is already engaged in work to promote the importance of spatial thinking to policymakers. This work should continue to ensure the alignment of the programme with the national policy context and the priorities of schools to encourage sustained take up of the programme at scale.

Our recommendation is that SPACE is taken forward to an efficacy trial with modifications to the programme and delivery models that make it suitable for scale. There are two options regarding the extent of modifications to be made. The first is to test an initial set of modifications, which would be deliverable at efficacy trial scale but not beyond. This would involve adaptations to address the concerns regarding how the programme is integrated into maths teaching and streamlining the supporting and monitoring processes as well as the ongoing work to promote spatial reasoning with policymakers. The second option is to test a fully scalable model in an efficacy trial. This would involve these adaptations plus more significant investment in developing the programme for scale, including developing a less expensive delivery modality or securing partnerships that provide the required logistical support and developing a train-the-trainer model and network of 'champions' to support delivery.

Research question 10: Are the proposed whole-class assessment measures appropriate for use in an efficacy trial?

The child assessments were administered pre- and post-intervention to establish feasibility, focusing on the appropriateness of the measures for this age group and setting, and to test their psychometric properties. The assessments also provide an insight into the potential promise of the intervention regarding children's outcomes. However, there was no control group in the pilot evaluation to provide a counterfactual so it is impossible to know if changes would have occurred through usual maths teaching. In this section, we report on the assessments based on the feasibility of administering the assessments in the classroom and the appropriateness of the measures for detecting change for this group, followed by a summary of the pre- and post-changes in assessment results.

The three assessments examined pre- to post-intervention differences in spatial language, mental rotation, and maths. The spatial language measure comprised two subscales: spatial language production; and spatial language comprehension. The maths measure also comprised two subscales: geometry; and problem-solving. For completeness, psychometric properties of subscales as well as the total scores are described below.

To administer the child assessments, teaching staff were provided with materials: i) assessment manual and scripts for teachers, which provide all the information required to administer the child assessments as intended; and ii) assessment booklets for children to complete.

Feasibility

The feasibility of administering the whole-class child assessments was examined based on the duration of the assessments, the proportion of missing data, and feedback from teachers. The evaluation also collected qualitative information through interviews, observations, and monitoring reports.

Data on the duration of assessments was obtained from 13 of the 15 participating schools, with a mean duration of 90 minutes (range 70–100 minutes). This is substantially longer than intended (45 minutes). Furthermore, teachers reported that completing the assessments with children who were not present for the whole-class assessment sessions and supporting lower-ability children with the assessments required additional staff time.

The completion as part of the trial was substantially longer than experienced by the delivery team when piloting the assessments. Piloting was conducted by the delivery team postdoctoral researcher with smaller groups of children (spatial language and mental rotation) and by two teachers at whole-class level (maths measure). The time it took to explain the questions and support children to complete them in this setting is likely to have been less than when conducted by teachers who were unfamiliar with the content in a whole-class context. This differential should be considered when developing new assessments in this way.

Most schools chose to administer the assessments across two or three separate sessions, which added some set-up time but was preferred due to the attention span required for children to focus across the tests.

This length was challenging for teachers and children. This teacher perceived the language assessment as taking an hour to do, though the data shows the longest was 45 minutes, demonstrating how much time they were perceived to take:

> *They were tricky. They were also long. The spatial language one was quite long. It took about an hour to do. Whereas at this time of the year, most of my tests—the longest one of our tests is 25 minutes, so an hour is quite a lot of time for five, six, and seven-year-olds to sit for to do it.* (Teacher)

There was minimal missing data on child assessments. At baseline, there was available data for 400–408 (there was a small degree of variance in how many children completed each of the three assessments) out of an anticipated total of 415 pupils (96%–98% complete). At endline, there was available data for 397–402 pupils (96%–97% complete). The high completion rates are likely due to the assessment being delivered to the whole class. Teaching staff were also asked to complete the assessments with any children missing from class when assessments were completed, provided they had the staff capacity to do so.

Feedback from teachers indicated that the training to deliver the child assessments was sufficient: 13 out of 14 teachers agreed (N=8) or strongly agreed (N=5) that the training was sufficient to deliver the assessments. The high level of complete data also indicates that teachers were able to administer the assessments as intended. Teachers found the manuals easy to follow:

> *They were easy enough to follow. Yes. I mean, you just follow the script, don't you? It's the same as any other testing that we do*. (Teaching assistant)

A recurring theme in feedback was that Year 2 children were not familiar with this type of assessment and, therefore, did not know what to expect or how to behave. A particular problem that arose was that children were not used to having to sit and concentrate on one task for so long at a time, so they were easily distracted, except in two schools where children had experience with SATs. Teachers were also not used to presenting children with content that they had not been specifically taught and felt it unfair (and uninformative) to do so.

Pupils with lower ability or with SEND had greater difficulty in accessing these assessments, and this was a particular issue in a school with a high proportion of children on FSM:

> *I think if it was a type of school where most of the children weren't deprived or of a lower ability, it would work fine, but they're not. I would probably say out of my 30 children possibly about 8 of them would cope with them independently, the rest couldn't, just due to the fact of things like they can't read.* (Teacher)

Some teachers interviewed said that administering the assessments became a little easier in the second assessment session because children had matured a little and remembered the previous experience and knew what to expect. The second application was marginally quicker than the first, and fewer children became frustrated.

Teachers also provided specific feedback on the appropriateness of the assessments for children at that stage in Year 2. The most criticised child outcome measure was the maths assessment. On average, the monitoring reports indicate that this instrument took 41 minutes to administer, with a range of 30 to 50 minutes—compared to an intended 25 minutes. Teachers found this the most challenging of the three child assessments because most of the items focused on curricular content that children had not covered, including multiplication, fractions, and graphs. This also raises ethical issues due to the potential for causing distress to children.

> *It was clear whoever had made them had made them for like an end-of-Year 2 child, not an Autumn Term Year 2 child, and they learn a lot in that gap […] Like even my most able children were not going to be able to tell me three-quarters of 16, and it wasn't because the LEGO® programme wasn't successful; it was just because we haven't taught fractions yet. I felt that you wouldn't get the data that you wanted to get because we hadn't taught the knowledge for them to be able to use their spatial awareness to help them solve the problem if that makes sense.* (Teacher)

In response to this concern, the delivery team reflected that they needed to be clearer to teachers that assessments used in research need to be designed to capture change over time, and, therefore, differ from educational assessments where you want all children to do as well as possible. They also discussed that if they only included assessment content that had been covered in the curriculum, there was a risk that the measurement would become about children's understanding of learned procedures rather than overall maths ability.

Teachers widely said in the monitoring meetings that the maths assessment had been challenging for children—some of whom were frustrated to the point of tears or who had to leave the classroom. This was exacerbated by some formatting issues, for example, not clearly indicating where pupils should write their answers.

The spatial language assessment was more acceptable to the participants. This instrument took an average of 30 minutes to complete (compared with an intended 10 minutes), with a range of 15 to 45 minutes. Teachers reported that the test was easy to administer as it came with a clear, step-by-step script, and children found it easy to understand, especially the comprehension section. However, several teachers also noted that the production section was more challenging for the children because many of them did not yet know how to write or had handwriting that was too large for the space provided. In several schools, teaching staff scribed answers for these children or went through the booklets to indicate children's intended answers. Some teachers noted that this section was particularly difficult for pupils with English as a second language.

The mental rotation task was also fairly well received by the participants. This instrument took an average of 21 minutes to complete, with a range between 10 and 30 minutes (compared with an intended 10 minutes). Teachers and teaching assistants found the instrument easy for children to understand, and smooth and quick to administer. The most difficult aspect of administration was that, in some cases, children did not understand that they could not turn the booklet. Some interviewees also felt that the booklet had too many items on the same page, which could be a bit overwhelming for their children, and needed more consistency between the formatting of the example items and the test items.

Findings from the pre- and post-intervention child assessments administered for this pilot evaluation show that the assessments as they are currently designed are not appropriate for delivery in this context. The duration, content, and delivery method of the assessments are the most significant barriers. Significant development work and validation are required to design assessments that are an appropriate duration and level of difficulty for this age group.

Descriptive statistics

The team also assessed the psychometric properties of the assessments to ensure their reliability and appropriateness for detecting change. Psychometric properties help us understand whether the tests are measuring what they are supposed to measure. A well-constructed outcome measure should be able to detect differences between pupils and changes in their performance over time. Descriptive statistics of the whole-class child assessments are shown in Table 15, and results are visually represented in histograms in Figure 23.

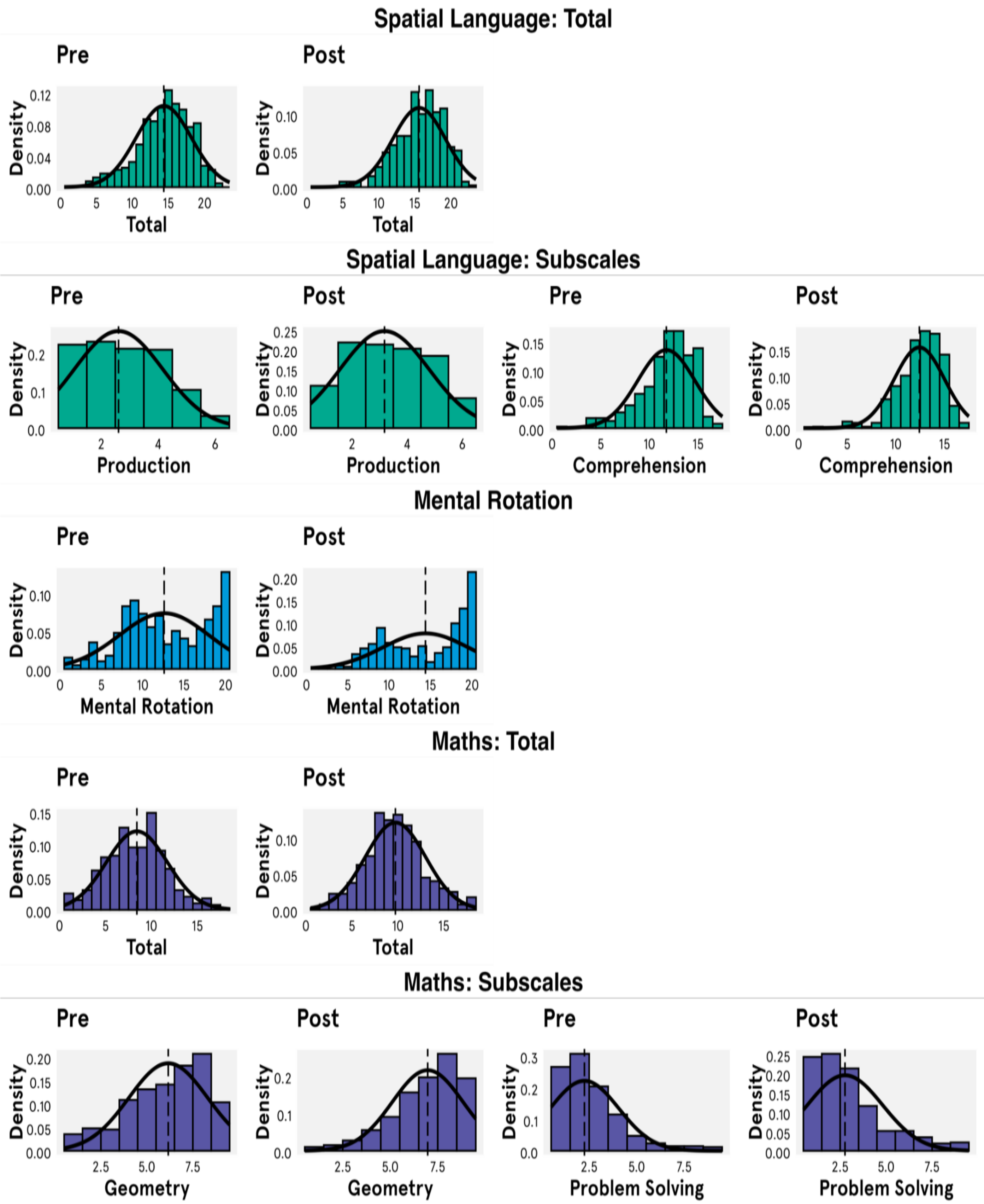
The total scores for the spatial language assessment ranged from 0–23. While the Shapiro-Wilk test (a statistical test of normality[12]) indicated that the data were not normally distributed, visual inspection of the histograms indicated minimal deviation from normality, with only a slight negative skew (this was also observed in Q-Q plots, see Appendix M). This means that the distribution was fairly normal, with a few lower scores skewing the distribution. Further analysis of the subscales (looking at comprehension and production separately) indicated greater variance and more normal distribution within the comprehension subscale, and less variance and a less normal distribution in the production subscale. This is related to the smaller number of items on the spatial language production measure (meaning children could only score between 0 and 6 marks). The comprehension subscale showed more indication of a negative skew. At the item level (looking at individual items), there were three items in the comprehension measure at baseline that scored at ceiling (meaning most pupils got them correct) and one item scored at floor (meaning most pupils got it wrong). Note, however, that the subscales were designed to be treated as one measure and so the distribution of the full measure is key here.

Mental rotation assessment had scores ranging from 0–20. The Shapiro-Wilk test indicated that the data were not normally distributed. Visual inspection of the histograms indicated a bimodal distribution of scores, with two peaks (Q-Q plot showed minimal deviation from normality, with some evidence of a ceiling effect, see Appendix M). One group of students scored around 10 out of 20 (indicating that they were performing at chance level), while the second group scored around 20 out of 20 (indicating perfect performance). This bimodal distribution of scores might indicate that this test measures the acquisition of a specific skill that students either have (and score 100%) or have not acquired (and perform at chance), because of the two-option, multiple-choice format of the assessment. Item-level analyses indicated no individual items scoring at ceiling or floor.

For the maths assessment, total scores ranged from 0–18. The Shapiro-Wilk test again indicated a non-normal distribution, but visual inspection of the histograms showed minimal deviation from normality, with minor positive skew (a few higher scores are affecting the distribution of results; a Q-Q plot showed almost no deviation from normality, see Appendix M). Examining the distributions of scores on the problem-solving and geometry subscales at both pre- and post-assessment, suggests a potential ceiling effect on the geometry subscale (many pupils scored highly), and potential floor effect on the problem-solving subscale (many pupils got low scores). Note, however, that the subscales were designed to be treated as one measure and so the distribution of the full measure is key here. Item-level analyses indicated no individual items were consistently answered at ceiling or floor.

[12] The normal distribution describes a symmetrical plot of data around its mean value, where the width of the curve is defined by the SD. When graphed it appears as a 'bell curve'.

Figure 23: Histogram density plots of whole-class spatial language, mental rotation, and maths assessments



Note: Dotted vertical lines indicate the mean, solid black lines indicate the overall distribution.

Table 15: Descriptive statistics of whole-class assessment measures

Measure	Timepoint	N	N missing	% missing	Mean	SD	SE	Median	Min	Max	Range	Skew	Kurtosis	Shapiro-Wilk	P-value
Spatial language total	Pre	408	7	5.15	14.35	3.86	0.19	15	0	22	22	-0.77	0.81	0.96	.000
	Post	403	12	5.21	15.59	3.67	0.18	16	0	23	23	-1.09	2.35	0.93	.000
Spatial language: Comprehension	Pre	408	7	5.15	11.74	2.95	0.15	12	0	17	17	-1.24	2.12	0.91	.000
	Post	403	12	5.21	12.41	2.59	0.13	13	0	17	17	-1.63	4.91	0.88	.000
Spatial language: Production	Pre	408	7	5.15	2.61	1.55	0.08	3	0	6	6	0.17	-0.83	0.94	.000
	Post	403	12	5.21	3.18	1.62	0.08	3	0	6	6	-0.09	-0.83	0.95	.000
Mental rotation	Pre	404	11	3.96	12.59	5.41	0.27	12	0	20	20	-0.24	-0.87	0.94	.000
	Post	408	7	3.92	14.39	5.15	0.26	16	0	20	20	-0.51	-0.91	0.88	.000
Maths total	Pre	400	15	5.25	8.41	3.32	0.17	8.5	0	18	18	0.03	-0.03	0.99	.035
	Post	403	12	5.21	9.73	3.28	0.16	10	0	18	18	-0.01	0.33	0.99	.005
Maths: Geometry	Pre	400	15	5.25	6.16	2.14	0.11	6.5	0	9	9	-0.72	-0.17	0.93	.000
	Post	403	12	5.21	6.97	1.86	0.09	7.5	0	9	9	-1.27	1.74	0.88	.000
Maths: Problem-solving	Pre	400	15	5.25	2.25	1.81	0.09	2	0	9	9	1.21	1.91	0.89	.000
	Post	403	12	5.21	2.76	2.04	0.10	2	0	9	9	1.07	0.84	0.90	.000

Max, maximum; Min, minimum; SE, standard error.

Statistical analyses

Repeated-measures one-way ANOVA was used to examine the main effect of time. This is a statistical test to compare the results at baseline (pre-intervention) and endline (post-intervention) in the three main outcome measures. This analysis was not done on the individual subscales because the subscales were not designed for this purpose. The results of these analyses, as shown in Table 16, show statistically significant increases in scores in all three measures between baseline and endline. However, the effect sizes were small, meaning that while there was consistent improvement in scores at endline, the real-terms change was only modest. This means that a positive difference was observed in the three outcomes of interest after the SPACE intervention compared to before it.

Table 16: Repeated-measures ANOVA results for pre- to post-assessment spatial measures

Measure	F	DF	P-value	Partial η^2	CI (low)	CI (high)
Spatial language	22.14	1, 809	<0.001	0.03	0.01	0.05
Mental rotation	23.60	1, 810	<0.001	0.03	0.01	0.05
Maths	32.02	1, 801	<0.001	0.04	0.02	0.07

CI, confidence interval; DF, degrees of freedom.

Additional exploratory analyses investigated whether there were differences in baseline and post-intervention outcome scores depending on the school attended and individual pupil characteristics. At the school level, individual differences across schools were explored (N=15 schools), as well as the percentage of pupils eligible for FSM at each school. Schools were categorised as a binary variable, comparing schools where fewer than 22% of pupils were eligible for FSM to schools were 22% of more of pupils were eligible. At the pupil level, the effect of age (in months; September 2016 to August 2017, coded numerically as 1–12) was explored, as well as intervention 'dosage' (the number of SPACE sessions each pupil attended ranging from 0–12; post-intervention scores only).

We explored whether there were any differences in spatial abilities prior to the SPACE intervention depending on the school attended, FSM level, and pupil age. A one-way ANOVA was run, entering each predictor into a separate model. We ran nine different models, with each of the three predictors (school, FSM level, and pupil age) for each of the three outcome measures. We provide results with and without correction for multiple comparisons. The results indicated significant effects of school for all three outcomes measures, with medium to large effect sizes (Table 17; all of which survived correction for multiple comparisons). This means that the school a pupil attended affects their scores on the spatial language, mental rotation, and maths assessments before the intervention.

Table 17: Results of baseline one-way ANOVAs exploring the effects of school- and participant-level variables on baseline assessment measures

Assessment measures	F	DF	P-value	Partial η^2	CI (low)	CI (high)
School						
Spatial language	4.06	14	<0.001**	0.13	0.04	0.16
Mental rotation	5.73	14	<0.001**	0.17	0.08	0.22
Maths	4.24	14	<0.001**	0.12	0.05	0.17
FSM level						
Spatial language	4.54	1	0.035*	0.04	0.00	0.14
Mental rotation	0.07	1	0.785	0.00	0.00	0.04
Maths	4.62	1	0.034*	0.04	0.00	0.15
Pupil age						
Spatial language	9.51	1	0.002**	0.02	0.00	0.06
Mental rotation	6.51	1	0.011*	0.02	0.00	0.05
Maths	10.58	1	0.001**	0.03	0.00	0.07

** Denotes tests that are statistically significant (alpha = 0.05).*

*** Denotes tests that remain statistically significant following multiple-comparisons correction using the Bonferroni method (alpha = 0.005).*

CI, confidence interval; DF, degrees of freedom.

There were also significant effects of school FSM level on results for the spatial language and maths measures, with small effect sizes, however, these effects did not survive corrections for multiple comparisons. There was no significant effect of FSM level on mental rotation scores. These findings indicate some evidence suggesting that the percentage of pupils eligible for FSM at a school affects spatial language and maths scores at baseline; schools in the higher FSM category performed worse than those in the lower FSM category. However, these effects would need to be confirmed in a larger more highly powered study.

Finally, there were also statistically significant effects of pupil age on all outcome measures, with small effect sizes (although the effect on mental rotation did not survive correction for multiple comparisons). This provides some evidence that older pupils tended to perform better, particularly on spatial language and maths measures.

Additional analysis was conducted using mixed-ANOVAs (m-ANOVAs) to assess whether FSM level, pupil age, and attendance (between-subjects factors) influenced how children's scores changed over time (within-subjects factor). FSM level (categorical), pupil age (continuous), and attendance (continuous) were included as additional factors.[13] We also looked at whether any of these variables interacted with time, meaning whether pupils from different groups improved differently over time, for example, if children who attended more sessions improved in their results more than those who attended fewer sessions.

Results demonstrated that there were no statistically significant interaction effects. This means that there was insufficient evidence to conclude that the level of improvement in spatial language, mental rotation, or maths differed depending on the FSM level of the school attended, pupil age, or attendance (Table 18). However, the main effect of time remained statistically significant in each of these models, indicating that the effect of time was still present even after controlling for these covariates and potential interaction effects.

Table 18: Results of m-ANOVAs exploring interaction effects of FSM category and pupil age with time

Assessment measures	F	DF	P-value	Partial η^2	CI (low)	CI (high)
Spatial language						
Time	5.87	1	0.016	0.03	0.00	0.09
FSM	10.72	1	0.001	0.06	0.01	0.13
Age (months)	6.67	1	0.011	0.03	0.00	0.10
Dosage (attendance)	4.05	1	0.046	0.02	0.00	0.08
Time*FSM	0.71	1	0.402	0.00	0.00	0.04
Time*age	0.13	1	0.724	0.00	0.00	0.03
Time*dosage (attendance)	0.00	1	0.999	0.00	0.00	0.00
Residuals		184				
Mental rotation						
Time	4.68	1	0.032	0.02	0.00	0.08
FSM	0.03	1	0.856	0.00	0.00	0.02
Age (months)	0.82	1	0.365	0.00	0.00	0.04
Dosage (attendance)	7.06	1	0.009	0.04	0.00	0.10
Time*FSM	0.28	1	0.598	0.00	0.00	0.03
Time*age	0.50	1	0.482	0.00	0.00	0.04
Time*dosage (attendance)	1.85	1	0.175	0.01	0.00	0.06
Residuals		188				
Maths						
Time	5.78	1	0.017	0.03	0.00	0.10
FSM	5.12	1	0.025	0.03	0.00	0.09
Age (months)	0.03	1	0.872	0.00	0.00	0.02
Dosage (attendance)	1.00	1	0.319	0.01	0.00	0.05
Time*FSM	0.02	1	0.875	0.00	0.00	0.02
Time*age	0.63	1	0.428	0.00	0.00	0.04
Time*dosage (attendance)	0.29	1	0.594	0.00	0.00	0.03
Residuals		174				

CI, confidence interval; DF, degrees of freedom.

[13] School was not included in this model as FSM category and schools were 'aliased coefficients', meaning that because schools were nested within groups of FSM, including both in the model was unable to explain any additional variance.

Conclusions

SPACE was piloted in 15 primary schools in Portsmouth and Surrey to improve Year 2 children's spatial reasoning and maths skills. The programme is designed to be delivered within the scheduled maths lesson time, twice a week for six weeks, and involves children building LEGO® models from pictorial instructions, with support from teachers and teaching assistants using spatial prompts. This pilot evaluation focused on ten research questions covering three key areas of interest: the feasibility of delivering SPACE; evidence of promise that the programme was leading to the intended changes; and readiness of the delivery model to be scaled up for further evaluation. A summary of findings from these research questions are summarised in Table 19 and explored in more detail here in the 'Interpretation' section below.

Table 19: Summary of pilot findings

Research question	Finding
Feasibility of implementation	
1. Is SPACE feasible for an implementation in maths lessons?	The findings indicate that further development is needed to ensure that the delivery format and timing of school recruitment, planning, and training are integrated into the school processes for sustainable feasibility. While SPACE was considered feasible by the teaching staff involved, in terms of ease of delivery, there were concerns expressed by school staff that SPACE took time from their usual maths curriculum. Options for adaptations to integrate SPACE into planned maths content were discussed.
2. Are the training and support provided sufficient for teachers and teaching assistants to implement SPACE?	The training, resources, and support were well received by teaching staff and were found to be sufficient to deliver the programme. Training and support sessions were well attended. Minor amendments to the materials and support offer have been suggested to improve their appropriateness for scale up.
3. Was SPACE implemented as intended?	The composite fidelity score shows that SPACE was implemented as intended across all schools. The core elements (attendance and dosage) were delivered with high fidelity and the indicators of quality were delivered with medium to high fidelity. Minor adaptations to the delivery model were noted, which were not thought to have affected the quality of delivery overall. Key facilitators include staff capacity and SLT support.
4. Is SPACE acceptable to teaching staff?	The programme was perceived as enjoyable and easy to deliver by teaching staff, indicating good acceptability. Teaching staff saw value in learning through play and expressed interest in continuing the use of the intervention or similar activities using physical manipulatives. However, acceptability was moderated by: i) their perception of the impacts for children; and ii) the extent to which they could see the association between SPACE content and the wider maths curriculum, two considerations that may be linked.
5. Do children, particularly those from disadvantaged backgrounds, engage well with SPACE?	Children engaged extremely well with the SPACE programme, including children from disadvantaged backgrounds. Children were perceived to find the activity motivating and to focus for the duration of each session. There was some variation in engagement, both positive and negative, with the programme noted for children with SEND, EAL, or behavioural issues compared with other children, but these differences were not uniform and potential additional benefits were also noted for some children.
Evidence of promise	
6. Do the training and delivery of SPACE sessions lead to changes in teaching staff's understanding of the importance of spatial skills, use of spatial language, and confidence in teaching spatial skills?	Survey and interview findings suggest that SPACE led to self-reported changes in teaching staff's confidence and understanding of teaching spatial reasoning and use of spatial language, in line with the Theory of Change. There is some evidence that these changes may be sustained three months after the programme delivery. Participants generally considered that spatial thinking was very important for the development of children's mathematical skills at all timepoints and found the training useful in explaining why the programme was thought to be beneficial.
7. What impacts do teaching staff perceive for children, particularly those from disadvantaged backgrounds (e.g. spatial language, spatial skills, and maths skills and attainment)?	Teaching staff perceive positive impacts for children in terms of spatial language and spatial skills, but there was more mixed perceived impact on maths skills, at this early stage. The programme was also seen to have a positive impact on fine motor skills, resilience, or perseverance and confidence both within and beyond maths lessons—areas not described in the Theory of Change. It was also seen to support the transition from continuous provision into a more structured learning environment of Year 2. Positive change was noted for children from all backgrounds, particularly those considered to have fewer opportunities to engage with physical manipulatives in the home.
8. What are the perceived mechanisms and possible causal chains?	A variety of causal mechanisms for change in teacher and children's outcomes were identified through this pilot evaluation. Teaching staff's increased awareness of the importance of spatial skills and effect of their language was thought to increase their use of spatial prompts to explain

Research question	Finding
	tasks and concepts, rather than physically modelling them. For children, the repetition of the task was thought to increase confidence and engagement in maths and their exposure to spatial language led to them using it with teachers and peers.
Readiness for an efficacy trial	
9. How could SPACE be delivered on a larger scale?	Two potential viable pathways to scale were identified, assuming future evaluations continue to identify positive evidence of impact that warrants continued delivery and scale up: i) using evidence from an efficacy trial for political advocacy with the aim of embedding spatial reasoning into the national curriculum; and ii) adapting SPACE to be suitable for scaled-up delivery by the delivery team working with a partner. Necessary amendments to the delivery model and scale-up activity per pathway have been discussed.
10. Are the proposed whole-class assessment measures appropriate for use in an efficacy trial?	The data show that the whole-class assessment measures are not appropriate for use in their current form. Although the results indicate that they can detect change over the intervention period, the duration and content make them unacceptable to teachers and children.

Success indicators

Success indicators were developed prior to the delivery of the programme to pre-empt key features of the programme delivery that would be important when considering progressing this intervention along the evidence pipeline to a larger trial. Each of these indicators was assessed as summarised in Table 20.

Table 20: Success indicators

Pilot criteria	Success indicator	Data sources	Findings
Feasibility of implementation	Teaching staff consider the intervention implementable (with minor amendments).	Teaching staff surveys, teaching staff interviews, and observations.	The intervention was implementable but requires some adaptation to integrate it into planned maths content.
	Teaching staff consider the intervention acceptable (with minor amendments).	Teaching staff surveys, teaching staff interviews, and observations.	Clearly demonstrated with minor amendments noted.
	Schools are able to deliver the intended intervention dosage within the defined period.	Programme monitoring data, and teaching staff interviews.	Clearly demonstrated.
	Schools are able to deliver the intervention with medium to high fidelity.	Teaching staff surveys, teaching staff interviews, programme monitoring data, and observations.	Clearly demonstrated.
Evidence of promise	Findings indicate that SPACE has a positive influence on teacher knowledge, understanding, and/or confidence in spatial skills.	Teaching staff surveys and teaching staff interviews.	Data support some positive outcomes in teacher understanding and confidence in spatial skills.
	Indication of SPACE leading to improvements in children's spatial reasoning and maths skills.	Teaching staff surveys, teaching staff interviews, and child outcome measures.	Data support positive outcomes in children's spatial language and spatial skills, but mixed findings in relation to the long-term outcome, maths skills.
	There are no indications that pupils receiving FSM benefit less from SPACE, or that there are barriers to some pupils benefiting from SPACE.	Teaching staff and SLT interviews.	There are mixed views about, which children benefit more or less from the SPACE programme, but there is not consistent evidence that disadvantaged pupils benefit less.
Readiness for an efficacy trial (at scale)	There are viable strategies to collect sufficient data to monitor compliance and fidelity.	Programme monitoring data and delivery team focus groups.	Viable strategies are in place for efficacy trial at scale.

Pilot criteria	Success indicator	Data sources	Findings
	SPACE can be scaled for an efficacy trial (with minor amendments).	Observations, teaching staff, and SLT interviews.	Adaptations to training and planning would be needed for integration into maths teaching.
	There is a viable vision for delivering SPACE at scale.	Delivery team focus groups and SLT interviews.	Two viable pathways to scale have been identified. Further work is required to identify, which of these is most appropriate and to adapt the delivery model accordingly.

Formative findings

The intervention was considered easy to deliver and was found enjoyable by children and teaching staff. However, amendments to the intervention are recommended to ensure it is successfully planned for and embedded into maths lessons.

Integrating SPACE with maths teaching

Timing of delivery

For this evaluation, schools were recruited in the Summer Term for an immediate start to the programme in September 2023. While this ensured that SPACE could be delivered within the Autumn Term, it presented challenges to the programme delivery. One issue was staff turnover: in some schools, the person responsible for signing up for the programme had left the role by September 2023. This meant that the Year 2 teacher taking part was not necessarily invested or interested in the programme. This also meant that they had not seen all the information and consent materials prior to starting the term.

The pilot study identified that it is critical for any whole-class intervention to be built into the lesson planning well in advance of the intervention starting. Therefore, schools need to be recruited in sufficient time for class teachers to embed SPACE into their maths lesson planning so that it does not replace planned lessons. For interventions starting in the Autumn Term, this means providing clear information and instructions prior to the summer holidays. This can be challenging due to competing priorities in the Summer Term such as assessments, school staff turnover, and holidays. If delivery continues in the Autumn Term, this would require the delivery term to engage with this in detail with schools during the Summer Term, and perhaps also to move the SPACE training to that term. While schools were informed of the need to conduct SPACE as part of the maths sessions prior to the summer holidays via the MoU, information sheets, and webinar, more explicit initial communication about the programme in advance of the training could mitigate these timetabling issues. Delivery in the Spring Term could mitigate these issues.

Across the study, there were suggestions that the Autumn Term is not appropriate for the delivery of a new intervention with this age group. None of the participating schools covered geometry as part of the maths curriculum in the Autumn Term. While this is not considered an issue for the programme itself, it contributed to teachers' perception that SPACE was taking time from planned maths teaching. Some interviewees also felt it would have been more effective if run in parallel with teaching on 'Space and Shape'—typically in the Spring Term of Year 2. While delivery in the spring may mitigate this issue, the delivery team emphasise that SPACE is relevant to all parts of the curriculum and should be deliverable at any stage in the year. This should be made clear in any updates to the training and support materials.

We recommend that the timing of the programme is accounted for in future delivery timelines, and mitigations are put in place to mitigate these issues if delivery proceeds in the Autumn Term.

Delivery model

A key challenge identified in this study was that teaching staff did not all clearly see the link between SPACE training and the maths curriculum content. This may be due to problems clearly conveying the links between spatial thinking and maths during training and delivery. We recommend that this issue is approached in two ways: i) the training and support sessions should make clear the ways in which the programme supports children's mathematical development and provide more tangible and nuanced links to the curriculum than are already provided; and ii) the delivery model should be adapted to include a weekly learning objective or set of vocabulary that directly link to the curriculum. The delivery team will also consider other methods of embedding SPACE into the maths curriculum, such as providing teachers with explicit guidance during the planning and set-up phase as well as at the end of the programme. It is thought that these changes to the intervention will both increase the acceptability of the programme to teaching staff and increase the efficacy of the programme in leading to positive outcomes for children's spatial and mathematical aptitude.

Several teachers and teaching assistants suggested delivering SPACE as a small group or lunch club activity. Although changing the format of the sessions in this way would, on the surface, appear to address concerns regarding the programme replacing maths, there are several factors to consider. The BLOCS intervention, which was delivered by the University of Surrey team prior to this evaluation, was delivered by a classroom teacher or a delivery team research assistant with small groups during lunchtime. They perceive several issues with delivery in this format, discussed in the delivery team focus group:

- In BLOCS, children often did not attend the sessions because they wanted their lunch break and saw SPACE as optional. This is likely to be seasonal, with children less inclined to stay inside during the warmer months.
- An additional session outside of maths time would mean children were getting more maths time overall, so changes to their attainment could not be attributed to the programme. An active control group would be required to demonstrate efficacy, which can be difficult to scale up.
- Delivery outside the classroom requires additional staff time or an external deliverer, placing additional burden on schools.
- The activity would not be associated with maths by children or teachers, thus, removing some of the proposed mechanisms of change.
- Not embedding it in maths time may restrict the extent to which teaching staff develop their awareness of spatial reasoning and use of spatial language, inhibiting the causal chains to positive outcomes that result from teaching staff developing new spatial reasoning teaching strategies.
- The selection of children for smaller group delivery may mean some children who could benefit from the programme are not given the opportunity to participate. Interviewees' comments on those unexpectedly struggling or benefit, reinforce this concern.

Overall, the delivery team believes that whole-class delivery in maths time provides the most potential for positive outcomes, as it enables equal opportunity for all children to benefit and facilitates the 'spatialisation' of the curriculum through the teaching staff's active involvement in delivery. Other adjustments to the programme content and delivery format such as the term in which its delivered and supplementing the existing activity with links to the Year 2 curriculum may mitigate these concerns while ensuring the programme format is ready for delivery on a larger scale.

Refining delivery materials

Overall, this study found that the delivery materials were sufficient for delivery as intended, but that improvements would support the quality of delivery and ensure that the programme is integrated better into classroom teaching. Minor amendments to the delivery materials have been suggested for ease of use and quality of delivery. Improvements to the presentation of the model booklets and weekly theme videos have been advised to ensure they are accessible and engaging for children. This includes clearer page numbers to help children and teachers navigate the booklets, easier-to-read contrasts in the pictorial diagrams, and more varied and interactive weekly theme videos to keep children engaged with the topics. These are all straight forward feasible changes. A review of the format of the prompt cards is needed, to allow teaching staff to access their content while engaging with children. Suggestions included classroom posters with key terms and ideas or a key ring of helpful vocabulary and ideas that teaching staff can have on a lanyard. These formats have been delivered by the delivery team for other projects with success. No major amendments are required to the current delivery materials.

Pupil assessments

The pilot evaluation found that the current assessment tools are limited in their appropriateness for measuring the impact of the SPACE intervention. The piloting of these assessments provided initial insight into the potential promise of SPACE in improving maths, mental rotation, and spatial language, showed they are at an appropriate level for detecting change, and they were easy to deliver in a whole-class setting (with minor adjustments).

However, the duration and content of the assessments presents significant issues. The tests took significantly longer than intended and used limited maths teaching time in some cases, though they could be administered at any time. The content was problematic due to children being unfamiliar with it at this stage in Year 2. Furthermore, there were concerns about how well children who require additional support could engage. Therefore, significant changes to the pupil assessments will be required prior to an efficacy trial to ensure the duration of the sessions is not a burden on schools, they are suitable for the attention span of this age group, and they do not cause distress to children.

Full validation will also be required to ensure they are measuring the pre-defined constructs. Improvements to the assessment booklets will also be necessary, including clearer page numbers, fewer items per page, consistent graphics in the mental rotation task, and clearer answer booklets.

Prior to an efficacy trial, suitable pupil assessments should be developed, which do not place too much burden on schools for whole-class administration. Any piloting should reflect the delivery context to provide accurate estimates of duration. While there needs to be a balance of difficulty, to ensure that there is sufficient room for improvement to detect chance, assessment content should not be discouraging or distressing to children. Consideration should also be paid to the cultural and linguistic accessibility to ensure that no children are excluded from participation, especially when delivering in schools with high numbers of children receiving FSM and/or children with EAL.

Interpretation

This pilot evaluation has found that SPACE is a well-packaged and well-liked programme, which shows promise in improving children's spatial reasoning abilities. SPACE is straightforward to deliver as intended for whole-class delivery in Year 2, as reflected in the feasibility data and high levels of fidelity across schools. Children and teaching staff enjoyed the programme and children from all backgrounds engaged with it well. There were some indications that different groups may engage with the programme differently, such as EAL (positive effects) and ability (mixed effects), though this was not seen uniformly across schools and is not thought to have significantly hindered participation, apart from in the case of a few individual children with significant learning or behavioural issues. However, there are issues to resolve around how the intervention is incorporated into maths lesson planning, which have consequences for the acceptability and feasibility of the programme in the long-term and at scale.

The training and support provided were sufficient for high-quality delivery as intended and could even be reduced to align with capacity needs without reducing their effectiveness. Only minor adaptations were made to the delivery model, which are not thought to compromise the integrity of the programme.

Evidence of promise supporting the Theory of Change is provided by the survey and interview findings, as well as indicative results from the pre- and post-assessments. Results of this evaluation regarding teacher- and child-level outcomes show promising evidence and support several elements of the Theory of Change. Notably, there is strong evidence that the programme improved perceived improvement in teaching staff's understanding of spatial reasoning and confidence in using spatial reasoning teaching strategies in the classroom and increased awareness and use of spatial language with children. Some areas of change were identified that were not described in the initial Theory of Change. SPACE appears to have led to more sustained changes in teaching strategies to incorporate spatial skills more generally in the classroom. The extent to which this increase in confidence translates to teaching staff's perceptions of importance of spatial thinking for children's maths skills is unclear due to the small survey sample size and an inflated baseline (as described below, teachers were asked to rate the perceived importance shortly after receiving training on why spatial reasoning is important).

The evidence for improved outcomes in children in this pilot study is based on staff perceptions of children's skills and unvalidated pre- and post-measures. Positive impacts on children's spatial language were perceived by teaching staff, as well as some unexpected areas such as confidence and fine motor skills. The programme was thought to increase children's use of spatial language and spatial skills, and there are indications that some children from disadvantaged backgrounds, who do not otherwise have access to physical manipulative toys at home, may particularly benefit from it. While the child assessments and teaching staff surveys provide further indicative evidence for SPACE leading to positive changes in maths skills, the qualitative findings suggest that there are mixed views from teaching staff on how they are linked. We cannot rule out that teacher perceptions of impact for children are influenced by their perceptions of the link between spatial reasoning and maths. Unanticipated positive changes were also noted in fine motor ability, perseverance in a task, and confidence.

Further programme development and evaluation is required to more fully understand whether the mechanisms of change described in the Theory of Change are being achieved, particularly around teachers' understanding of the link between spatial skills and maths, and the longevity of any potential impacts. Perceivable improvements in children's outcomes were identified as important to teaching staff regarding the acceptability and sustainability of the programme.

The generalisability of this pilot evaluation is limited by the sample of schools. The sample represented a variety of schools in terms of size, type, Ofsted rating, and percentage of FSM eligibility. However, by design, the sample of schools were restricted to two contrasting geographies, which may limit the relevance of their experience nationally. Furthermore, the participating schools were perceived to be generally highly motivated, and many noted that spatial skills, learning through play, or other related constructs were included in their school strategies. An efficacy trial would need to recruit and retain a larger number of schools, which is likely to include schools that are less motivated to participate and whose school strategies may be less aligned with the programme.

The timing of the baseline survey will have affected responses. The survey was designed to be delivered before the training but was distributed post-training due to an administrative error. This may mean that baseline results overestimate participants' perceived importance and understanding of spatial reasoning at baseline, leading to an underestimate of change at Timepoint 2. This error was partially mitigated by including questions on self-perceived change at Timepoint 2.

While adjustments are recommended to ensure the programme is well aligned and integrated with normal school curriculum, streamline training and support, improve the quality of materials, and make the delivery model more scalable, the core activity of SPACE—training spatial skills using physical manipulatives and pictorial instructions—should be retained.

Overall, we recommend that SPACE is taken forward to efficacy trial, with prior development of the training and support model and content, as described in research question 9 (under 'Findings' section, 'Readiness for an efficacy trial' subsection), to ensure that is it better embedded into maths lessons and appropriate for delivery at scale without diminishing quality.

Future research and publications

The next stage of research regarding the SPACE programme would be to proceed to an efficacy trial with a model that has been adapted to be appropriate for the preferred route to scale. Assuming the programme proceeds as a whole-class intervention, this would be run as a randomised controlled trial design with a business as usual control condition. As this is a whole-class intervention, randomisation at the class level is recommended, though the school level should be considered to mitigate concerns around contamination, as the evaluation identified challenges in schools with multi-form entry where only one class was participating. It will be important to monitor whether Year 2 classes in the control condition progress more quickly through the maths curriculum and the implications of this.

As the outcome measures used in this pilot study were not found to be appropriate for use in this context, a revised set of assessments will need to be selected or designed and validated for use in an efficacy trial. Standardised measures of spatial reasoning of the appropriate duration are not currently available, but the feasibility of using standardised tests already in use, for example, at the end of Year 2, should be explored in order to minimise the testing burden on schools and children. This will provide this additional benefit of providing evidence about sustained changes in children's outcomes, as well as far transfer (i.e. the programme benefiting children's attainment more widely) in addition to near transfer (improving at the intervention task itself) and medium transfer (improving in spatial reasoning skills).

It is also important that a future trial assess differential experiences and outcomes for children from disadvantaged backgrounds.

As changes in teachers' knowledge and behaviour appear to be crucial to the Theory of Change, these should also be assessed more robustly in an efficacy trial to confirm them as causal mechanisms.

A more systematic and easy-to-use process of collecting monitoring information from schools will be required for a larger trial. Monitoring information is crucial to understanding compliance, quality, and fidelity particularly on a larger scale, where there is less personal contact between the delivery teams and schools. A simple to use online form or weekly delivery record should be designed. A similar form has previously been used by the delivery team and is considered very feasible.

A review of the SPACE Theory of Change will benefit future evaluation. This pilot evaluation identified possible variations and additions to the mechanisms of change and short-term outcomes described in the current version, particularly relating to children. Some elements to consider include:

- an additional mechanism of change regarding the programme increasing children's confidence and engagement in maths lessons in general by showing them that maths is not only about numeracy;
- a mechanism of change regarding children's motivation to participate due to the association of LEGO® with play;
- unanticipated outcomes, including improved fine motor skills, resilience or perseverance, and confidence; and
- potential unintended adverse consequences in the form of less progress in the maths curriculum should also be noted in the Theory of Change.

The delivery team plans to publish three papers based on their analysis of the maths assessment data. Paper one will explore SPACE intervention effects, investigating effects of gender and school-based socio-economic status. Paper two will explore associations between spatial abilities and maths in six-year-olds. Paper three will explore the effects of SPACE intervention in comparison to control data generated by the delivery team outside the context of this feasibility study. The evaluation team is also planning a paper based on the feasibility study.

References

Allen, L. and Dowker, A. (2022) 'Spatial Working Memory Counts: Evidence for a Specific Association Between Visuo-Spatial Working Memory and Arithmetic in Children'. *International Electronic Journal of Elementary Education,* 14: 3, 199–211.

Ayala, M.C., Strasser, K., Susperreguy, M.I. and Castillo, K. (2024) 'Socioeconomic Gaps in Specific Mathematical Skills at Different Ages in Primary School'. *Journal of Educational Psychology,* 116: 5, 762–784.

Bates, K.E., Williams, A.Y., Gilligan-Lee, K.A., Gripton, C., Lancaster, A., Williams, H., Borthwick, A., Gifford, S. and Farran, E.K. (2023) 'Practitioners' Perspectives on Spatial Reasoning in Educational Practice From Birth to 7 Years'. *British Journal of Educational Psychology,* 93: 2, 571–590.

Bower, C., Odean, R., Verdine, B.N., Medford, J.R., Marzouk, M., Golinkoff, R.M. and Hirsh-Pasek, K. (2020) 'Associations of 3-Year-Olds' Block-Building Complexity With Later Spatial and Mathematical Skills'. *Journal of Cognition and Development,* 21: 3, 383–405.

Bower, C., Zimmermann, L., Verdine, B., Toub, T.S., Islam, S., Foster, L. and Golinkoff, R.M. (2020) 'Piecing Together the Role of a Spatial Assembly Intervention in Preschoolers' Spatial and Mathematics Learning: Influences of Gesture, Spatial Language, and Socioeconomic Status'. *Developmental Psychology*, 56: 4, 686–698.

Brainin, E., Shamir, A. and Eden, S. (2022) 'Promoting Spatial Language and Ability Among SLD Children: Can Robot Programming Make a Difference?' *Journal of Educational Computing Research,* 60: 7, 1742–1762.

Chu, F.W., vanMarle, K. and Geary, D.C. (2016) 'Predicting Children's Reading and Mathematics Achievement From Early Quantitative Knowledge and Domain-General Cognitive Abilities'. *Frontiers in Psychology*, 7: 775.

Cornu, V., Schiltz, C., Pazouki, T. and Martin, R. (2017) 'Training Early Visuo-Spatial Abilities: A Controlled Classroom-Based Intervention Study'. *Applied Developmental Science,* 23: 1, 1–21.

Cornu, V., Schilz, C., Martin, R. and Hornung, C. (2018) 'Visuo-Spatial Abilities Are Key for Young Children's Verbal Number Skills'. *Journal of Experimental Psychology,* 166: 604–620.

Davis-Kean, P., Domina, T., Kuhfeld, M., Ellis, A. and Gershoff, E.T. (2022) 'It Matters How You Start: Early Numeracy Mastery Predicts High School Math Course-Taking and College Attendance'. *Infant and Child Development,* 31: 2, e2281.

Data Protection Act 2018, c. 12. Available at: https://www.legislation.gov.uk/ukpga/2018/12/contents/enacted (accessed 1 August 2024).

Department for Education. (2020) *'Mathematics Guidance: Key Stages 1 and 2. Non-Statutory Guidance for the National Curriculum in England*. London: Department for Education. Available at: https://assets.publishing.service.gov.uk/media/60095297e90e073ed0cf386c/Maths_guidance_introduction.pdf (accessed 1 August 2024).

Department for Education. (2024). '*Early Years Foundation Stage Statutory Framework: For Group and School-Based Providers*. *Setting the Standards for Learning, Development and Care for Children from Birth to Five'.* London: Department for Education. Available at: https://assets.publishing.service.gov.uk/media/65aa5e42ed27ca001327b2c7/EYFS_statutory_framework_for_group_and_school_based_providers.pdf (accessed 1 August 2024).

Dowker, A. (2005). Early identification and intervention for students with mathematics difficulties. Journal of learning disabilities, 38(4), 324-332.

Dowker, A. (2017) 'Interventions for Primary School Children with Difficulties in Mathematics'. In: J. Sarama, D. Clements, C. Germeroth and C.l Day-Hess (eds.) *'Advances in Child Development and Behavior, Volume 53: The Development of Early Mathematics Education'.* New York, NY: Elsevier, pp. 255–287.

Farran, E.K. and Atkinson, L. (2016) 'The Development of Spatial Category Representations From 4 to 7 Years'*. British Journal of Developmental Psychology,* 34: 4, 555–568.

Farran, E.K., McCarthy, S., Gilligan-Lee, K.A., Bates, K.E. and Gripton, C. (2024) 'Translating Research to Practice: Practitioner Use of the Spatial Reasoning Toolkit'. *Gifted Child Today,* 47: 3, 202–215.

Frick, A. (2019) 'Spatial Transformation Abilities and Their Relation to Later Mathematics Performance'. *Psychological Research,* 83: 7, 1465–1484.

Gale, N.K., Heath, G., Cameron, E., Rashid, S. and Redwood, S. (2013) 'Using the Framework Method for the Analysis of Qualitative Data in Multi-Disciplinary Health Research'. *BMC Medical Research Methodology*, 13: 1, 1–8.

Georges, C., Cornu, V. and Schiltz, C. (2023) 'The Importance of Spatial Language for Early Numerical Development in Preschool: Going Beyond Verbal Number Skills'. *PLoS One,* 18: 9, e0292291.

Gibbons, A. (2020) '*DfE Concedes 'Importance' of Space and Shape in EYFS'*. Tes Magazine. Available at: https://www.tes.com/magazine/archive/dfe-concedes-importance-space-and-shape-eyfs (accessed 20 April 2023).

Gifford, S. (2019). '*The Case for Space in the Early Years*'. British Society for Research into Learning Mathematics. Available at: https://bsrlm.org.uk/the-case-for-space-in-the-early-years/ (accessed 20 April 2023).

Gilligan, K.A, Hodgkiss, A, Thomas, M.S, Farran, E.K (2019) 'The Developmental Relations Between Spatial Cognition and Mathematics in Primary School Children'. *Developmental Science,* 22: 4, e12786.

Gilligan-Lee, K., Hawes, Z., Williams, A.Y., Farran, E.K. and Mix, K.S. (2023) 'Hands-On: Investigating the Role of Physical Manipulatives in Spatial Training'. *Child Development,* 94: 5, 1205–1221.

Gilligan-Lee, K.A., Bradbury, A., Bradley, C., Farran, E.K., Van Herwegen, J., Wyse, D. and Outhwaite, L.A. (2023) 'Spatial Thinking in Practice: A Snapshot of Teacher's Spatial Activity Use in the Early Years' Classroom'. *Mind, Brain, and Education,* 17: 2, 107–116.

Gilligan-Lee, K.A., Hawes, Z.C., and Mix, K.S. (2022) 'Spatial Thinking as the Missing Piece in Mathematics Curricula. *NPJ Science of Learning*, 7: 1, 10.

Gilligan-Lee, K.A., Hodgkiss, A., Thomas, M.S., Patel, P.K. and Farran, E.K. (2021) 'Age-Based Differences in Spatial Language Skills From 6 to 10 Years: Relations With Spatial and Mathematics Skills'. *Learning and Instruction*, 73, 101417.

Goldsmith, L. J. (2021). Using framework analysis in applied qualitative research. Qualitative report, 26(6), 2061-2076.

Gross, J., Hudson, C. and Price, D. (2009) *'The Long-Term Costs of Numeracy Difficulties'.* London: Every Child a Chance Trust. Available at: https://www.numicon.co.nz/uploads/66441/files/Numicon_research_ECC_paper.pdf#:~:text=The%20research%20reviewed%20here%20shows%20that%20numeracy%20difficulties,risk%20of%20involvement%20with%20the%20criminal%20justice%20system (accessed 1 August 2024).

Hawes, Z. and Ansari, D. (2020) 'What Explains the Relationship Between Spatial and Mathematical Skills? A Review of Evidence From Brain and Behaviour'. *Psychonomic Bulletin & Review,* 27: 3, 465–482.

Hawes, Z.C.K., Gilligan-Lee, K.A. and Mix, K.S. (2022) 'Effects of Spatial Training on Mathematics Performance: A Meta-Analysis'. *Developmental Psychology*, 58: 1, 112–137.

Janus, M. and Duku, E. (2007) 'The School Entry Gap: Socioeconomic, Family, and Health Factors Associated With Children's School Readiness to Learn'. *Early Education and Development,* 18: 3, 375–403.

Kohl R. and Cooley L. (2012) *'Scaling Up—From Vision to Large-Scale Change: A Management Framework for Practitioners. Second Edition'.* Washington, DC: Management Systems International 2012, pp. 1–61.

Lindner, N., Moeller, K., Dresen, V., Pixner, S. and Lonnemann, J. (2022) 'Children's Spatial Language Skills Predict Their Verbal Number Skills: A Longitudinal Study'*. PLoS One,* 17: 10, e0277026.

Lowrie, T., Logan, T. and Ramful, A. (2017) 'Visuospatial Training Improves Elementary Students' Mathematics Performance'. *British Journal of Educational Psychology*, 87: 2, 170–186.

Milat, A., Lee, K., Conte, K., Grunseit, A., Wolfenden, L., van Nassau, F., Orr, N., Sreeram, P. and Bauman, A. (2020) 'Intervention Scalability Assessment Tool: A Decision Support Tool for Health Policy Makers and Implementers'. *Health Research Policy and Systems*, 18: 1, 1–17.

Mix, K.S., Levine, S.C., Cheng, Y.-L., Stockton, J.D. and Bower, C. (2021) 'Effects of Spatial Training on Mathematics in First and Sixth Grade Children'. *Journal of Educational Psychology,* 113: 2, 304–314.

Mix, K.S., Levine, S.C., Cheng, Y.-L., Young, C., Hambrick, D.Z., Ping, R. and Konstantopoulos, S. (2016) 'Separate But Correlated: The Latent Structure of Space and Mathematics Across Development'. *Journal of Experimental Psychology. General,* 145: 9, 1206–1227.

Moss, J., Bruce, C.D., Caswell, B., Flynn, T. and Hawes, Z. (2016). '*Taking Shape: Activities to Develop Geometric and Spatial Thinking'.* Newmarket, ON: Pearson Canada Inc., pp. 13–16.

Pro Bono Economics. (2014) *'Cost of Outcomes Associated With Low Levels of Adult Numeracy in the UK'.* London: Pro Bono Economics.

Proctor, E., Silmere, H., Raghavan, R., Hovmand, P., Aarons, G., Bunger, A., Griffey, R. and Hensley, M. (2011) 'Outcomes for Implementation Research: Conceptual Distinctions, Measurement Challenges, and Research Agenda'. *Administration and Policy in Mental Health*, 38: 2, 65–76.

Pruden, S.M., Levine, S.C. and Huttenlocher, J. (2011) 'Children's Spatial Thinking: Does Talk About the Spatial World Matter?' *Developmental Science,* 14: 6, 1417–1430.

Ritchie, S.J. and Bates, T.C. (2013) 'Enduring Links From Childhood Mathematics and Reading Achievement to Adult Socioeconomic Status'. *Psychological Science*, 24: 7, 1301–1308.

Rojo, M., Gersib, J., Powell, S.R., Shen, Z.A., King, S.G., Akther, S.S., Arsenault, T.L., Bos, S.E., Lariviere, D.O. and Lin, X. (2024) 'A Meta-Analysis of Mathematics Interventions: Examining the Impacts of Intervention Characteristics'. *Educational Psychology Review,* 36: 9.

Schmitt, S.A., Korucu, I., Napoli, A.R., Bryant, L.M. and Purpura, D.J. (2018) 'Using Block Play to Enhance Preschool Children's Mathematics and Executive Functioning: A Randomized Controlled Trial'. *Early Childhood Research Quarterly,* 44, 181–191.

Spencer, L., Ritchie, J., Ormston, R., O'Connor, W. and Barnard, M. (2013) 'Analysis: Principles and Processes'. In: J. Ritchie, J. Lewis, C. McNaughton Nicholls, and R. Ormston (eds.) '*Qualitative Research Practice'*. London: SAGE Publications Ltd, pp. 269–294.

Svane R.P., Willemsen M.M., Bleses, D., Krøjgaard, P., Verner, M. and Nielsen, H.S. (2023) 'A Systematic Literature Review of Math Interventions Across Educational Settings From Early Childhood Education to High School'. *Frontiers in Education*, 8: 1229849.

Uttal, D.H., Meadow, N.G., Tipton, E., Hand, L.L., Alden, A.R., Warren, C. and Newcombe, N.S. (2013) 'The Malleability of Spatial Skills: A Meta-Analysis of Training Studies'. *Psychological Bulletin,* 139: 2, 352–402.

Verdine, B.N., Golinkoff, R.M., Hirsh-Pasek, K., Newcombe, N.S., Filipowicz, A.T. and Chang, A. (2014) 'Deconstructing Building Blocks: Preschoolers' Spatial Assembly Performance Relates to Early Mathematics Skills'. *Child Development,* 85: 3, 1062–1076.

VVOB (2021). '*User Guide to the Education Scalability Checklist'.* Brussels: VVOB. Available at: vvob_esc_def_interactif.pdf (accessed 1 August 2024).

Wechsler, D. (2001) '*Wechsler Individual Achievement Test--Second Edition (WIAT-II)'.* [Database record]. APA PsycTests. https://doi.org/10.1037/t15173-000

Weiner, B.J., Lewis, C.C., Stanick, C., Powell, B.J., Dorsey, C.N., Clary, A.S., Boynton, M.H. and Halko, H. (2017) 'Psychometric Assessment of Three Newly Developed Implementation Outcome Measures'. *Implementation Science,* 12: 108.

Yang, W., Liu, H., Chen, N., Xu, P. and Lin X. (2020) 'Is Early Spatial Skills Training Effective? A Meta-Analysis'. *Frontiers in Psychology,* 11: 1938.

Zhang, X. and Lin, D. (2015) 'Pathways to Arithmetic: The Role of Visual-Spatial and Language Skills in Written Arithmetic, Arithmetic Word Problems, and Nonsymbolic Arithmetic'. *Contemporary Educational Psychology,* 41, 188–197.

Appendices

Please find the Appendices in separate document on the webpage.

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

https://educationendowmentfoundation.org.uk

@EducEndowFoundn

Facebook.com/EducEndowFoundn